THE APPLICATION OF ADVANCED METROLOGY TECHNIQUES TO BALLISTICS AND TOOLMARK IDENTIFICATION

KATIE ANNE ADDINALL

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Huddersfield (in collaboration with Forensic Pathways Ltd)

Submission date September 2017

Abstract

Since the early 1900s, the field of ballistic toolmark evidence has been developing as the instrumentation and computational power available has advanced. However, the use of these advanced techniques has not yet been validated for use in criminal proceedings. This has resulted in ballistic toolmark evidence being presented using the same techniques that have been employed for decades, unable to utilise more advanced techniques that have currently not been deemed acceptable in courts worldwide.

Ballistic toolmark evidence currently relies on the use of comparison microscopy. Magnified optical 2D (greyscale) images of two separate surfaces are viewed side by side to ascertain the degree of similarity using visual comparison. Only highly trained expert examiners are able to carry out this comparison, and as such it is a time expensive method. The technique is built on subjective methodology, there are no mathematical outputs as the results are based on the opinion and experience of the expert examiner.

With the advance of computational power and measurement techniques, it is now possible to create a digital system for the measurement and correlation of ballistic toolmark evidence. Evidence can be acquired as high density topographical datasets, and these datasets can be correlated against one another using mathematical algorithms, resulting in a comparison result based on a mathematical score or percentage match. Consequently, using these techniques could result in a more time efficient, repeatable and accurate system without problems of subjectivity or user bias.

The novel contribution in this thesis has been shifting ballistic toolmark evidence and correlation from subjective and 2D qualitative methods to the use of the most advanced topographical areal datasets and mathematical correlation. A direct comparison of the efficacy of 2D digital and areal based systems was achieved, this showed that when the correct data is processed via the advanced system, there is a significant increase in the efficiency of hit list information.

Novel contributions to these findings also include optimising the pre-processing of areal datasets, for both cartridge and bullets, and the effect of cartridge case materials on the overall topography of the toolmark.

It was found that using current methods of data pre-processing resulted in a less efficient correlation system. For both bullets and cartridge cases however, using the developed pre-processing methods detailed in this thesis resulted in a more efficient method of correlation. Bullet correlation was also achieved using a full areal dataset of the toolmark, and such a method has not been published previously.

Material analysis was attained across various cartridge manufacturers, which was then compared to the overall topography of the toolmark. It was found that differences in material composition would lead to differences in the topography of the toolmark. This is the first instance of such findings being published.

Finally, a direct comparison of two separate advanced measurement systems was obtained, using the Alicona G4 focus variation instrument and the Alias ballistic imaging system. Using the same pre-processing methods for all datasets acquired, it was found that the quality of the dataset is significantly affected by the measurement method. Extraneous data such as optical spiking and data dropout was found to affect the efficacy of evidence correlation. This thesis presents the use of advanced methods for ballistic toolmark evidence, while considering issues with data fidelity, substrate material differences and pre-processing techniques.

Contents

List of Figures

List of Tables

Dedications and Acknowledgements

Firstly, a huge thanks to my supervisors Professor Liam Blunt, Dr Paul Bills and Dame Professor Jane Jiang. Each of you have helped gain fantastic opportunities for myself and have been there to help shape what this work has become. I truly appreciate each opportunity that was given to me and thank you for your support.

To Forensic Pathways, I thank you for the opportunity to work with your collection of evidence, which this thesis would have impossible without. I also thank you for the advice and the opportunity to work with you throughout my PhD research.

To Chris, Andy and Karl. Your continued enthusiasm and ability to pick me up will never be forgotten. I feel that I have learnt so much from each of you, not only in your engineering expertise but in getting through the ups and downs. With you all by my side we had some exciting adventures! And for every member of the Centre of Precision Technologies, you made me feel like I was a part of the family, even without an engineering background.

To my family, thank you for dragging me out every now and again, and helping me through it all. I'm sure that there were times in which I was on everyone's last nerve, but the seaside trips and archery days out helped me keep everything in perspective. Finally, to my husband Marc. You saw the highest highs and the lowest lows and continued to steadfastly believe in me. I can't imagine what this journey would have been like without you beside me cheering me on. I cannot fully describe how much that meant to me.

The author gratefully acknowledges the UK's Engineering and Physical Sciences Research Council (EPSRC) funding of the EPSRC Centre for Innovative Manufacturing in Advanced Metrology (Grant Ref: EP/I033424/1).

List of abbreviations

$ACCF_{max}$: Areal Cross-Correlation Function maximum

Alias: Advanced ballistics identification system

ATF: Bureau of Alcohol, Tobacco, Firearms and Explosives

CCF: Cross-correlation Function

CMS: Consecutive Matching Striae

Ds: Scale difference function

ENFSI: European Network of Forensic Science Institutes

EU: European Union

FBI: Federal Bureau of Investigation

FTI: Forensic Technology Incorporated

FV: Focus Variation

FWHM: Full Width Half Maximum

IBIN: Integrated Ballistics Information Network

ISO: International Standards Organisation

NIBIN: Nation Integrated Ballistic Information Network

NBIC: National Ballistic Imaging Comparison

NIST: National Institute for Standards and Technology

OCF: Open Case File

PCA: Principle Component Analysis

pOCT: parallel Optical Coherence Tomography

SLD: Super Luminescent Diode

SRM: Standard Reference Material

UK: United Kingdom

USA: United States of America

CHAPTER ONE

INTRODUCTION

In this chapter forensic science, the creation of ballistic toolmarks and the comparison of toolmarks is introduced.

It is vital to ballistic toolmark identification to understand the creation of the toolmark, and the different scales of information present within the toolmark. Identification relies upon the comparison of individual characteristics alone, and should characteristics be misclassified, the risk of false identification is increased.

Modern measurement instruments can acquire the height of each measured data point, thus increasing surface information compared to 2D imaging techniques. As such, treatment of datasets must be considered, and is introduced in this chapter.

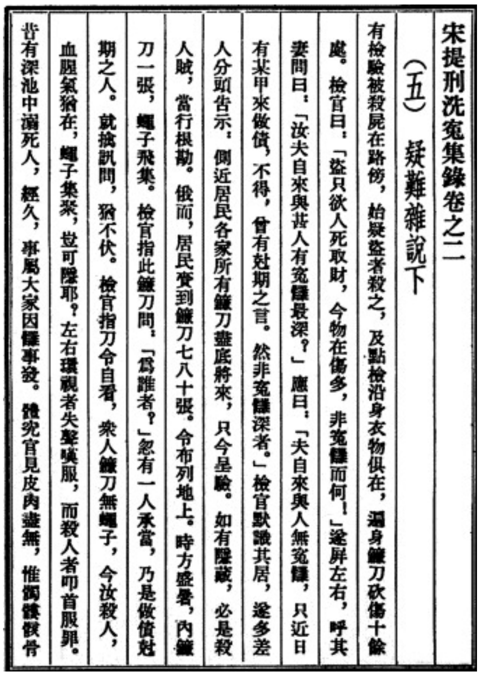
1 Introduction

1.1 Forensic science

"The application of scientific methods and techniques to matters under investigation by a court of law" (Waite, 2012)

Forensic science has been long established in many different evidence fields, evolving over time as understanding and instrumentation has become more advanced.

The first documented example of forensic science dates to the 13th century, in which a death investigator in China used comparative studies to determine the tool used in a murder. Sung Tzu created a sample set of toolmarks in bone using weapons available in the village. Comparative studies to toolmarks found on the victim's bones proved the weapon used in the murder was a sickle. Knowing an object with traces of blood on the surface would attract flies, he gathered all the sickles belonging to the villagers. One of the sickles attracted flies while the others did not, and the owner of the sickle confessed to the murder (Figure 1-1) (Benecke, 2001).



宋提刑洗冤集錄卷之二

（五）疑難雜說下

有檢驗被殺屍在路傍，始疑盜者殺之，及點檢沿身衣物俱在，遍身鐮刀砍傷十餘處。檢官曰：「盜只欲人死取財，今物在傷多，非冤讎而何！」遂屏左右，呼其妻問曰：「汝夫自來與甚人有冤讎最深？」應曰：「夫自來與人無冤讎，只近日有某甲來做債，不得，曾有尅期之言。然非冤讎深者。」檢官默識其居，遂多差人分頭告示：側近居民各家所有鐮刀盡底將來，只今呈驗。如有隱藏，必是殺人賊，當行根勘。俄而，居民賫到鐮刀七八十張。令布列地上。時方盛暑，內鐮刀一張，蠅子飛集。檢官指此鐮刀問：「為誰者？」忽有一人承當，乃是做債尅期之人。就擒訊問，猶不伏。檢官指刀令自看，衆人鐮刀無蠅子，今汝殺人，血腥氣猶在，蠅子集聚，豈可隱耶？左右環視者失聲嘆服，而殺人者叩首服罪。

昔有深池中溺死人，經久，事屬大家因讎事發。體究官見皮肉盡無，惟留髑髏骨

Figure 1-1: Excerpt of Sung Tzu's book detailing the murder investigation.

Since then the methodology involved in forensic investigation has changed little. Crime scenes are searched for evidence, and hypotheses are put forward regarding what circumstances may have resulted in the evidence being present at a crime scene. The conditions are recreated in the lab, interchanging variables as described by the hypotheses.

The results lead to a forensic scientist being able to assign the probability of evidence given each variable. These probabilities are then presented in court to convey to the jury the most probable scenario of the crime scene.

While evidence is probability based, this is not to say that it cannot convey a very strong likelihood in favour of a particular hypothesis. For example: DNA evidence, known as the "gold standard", will give a likelihood of over 99.999% that two matching DNA profiles came from the same person (Lucy, 2005). While DNA evidence currently conveys a high strength of evidence, this is not the case in other forensic disciplines. Ballistic toolmark evidence does not yet command the same legal status. To be able to build a statistical reference based on ballistic toolmark evidence, a meaningful standard operating procedure must first exist, which must encompass various factors on the variation of toolmarks. Currently, no such procedure for the analysis of ballistic toolmarks exists (Bunch & Wevers, Application of Likelihood Ratios for Firearm and Toolmark Analysis, 2013; Spiegelman & Tobin, 2013).

1.2 An introduction to ballistic toolmark evidence

Toolmarks come under the forensic discipline known as impression evidence. When a harder object comes into contact with a softer material, the contact results in the plastic deformation of the softer surface. This plastic deformation results in the permanent impression of the tools surface in the softer material (Gambino, et al., 2011).

When firing a gun, the trigger is depressed which forces the firing pin into the primer (assuming centre fire ammunition) of the cartridge, causing a permanent firing pin impression in the soft primer cap. The primer compound within the primer cap, usually a mixture of lead styphnate, antimony sulphide and barium nitrate is ignited due to the percussive force of the firing pin hitting the primer cap. (Rendle, 2005). The accelerant within the primer cap is ignited, and this ignition in turn ignites the main propellant encased in the cartridge (Figure 1-2). The deflagration of the main propellant pushes the bullet forward through the rifling of the barrel before exiting the firearm. As the barrel is slightly smaller than the circumference of the bullet, the relatively softer material of the bullet is compressed and engraved, and the machining marks of the rifling are imparted onto the surface of the bullet (Figure 1-2) (Xie, Xiao, Blunt, Zeng, & Jiang, 2009).

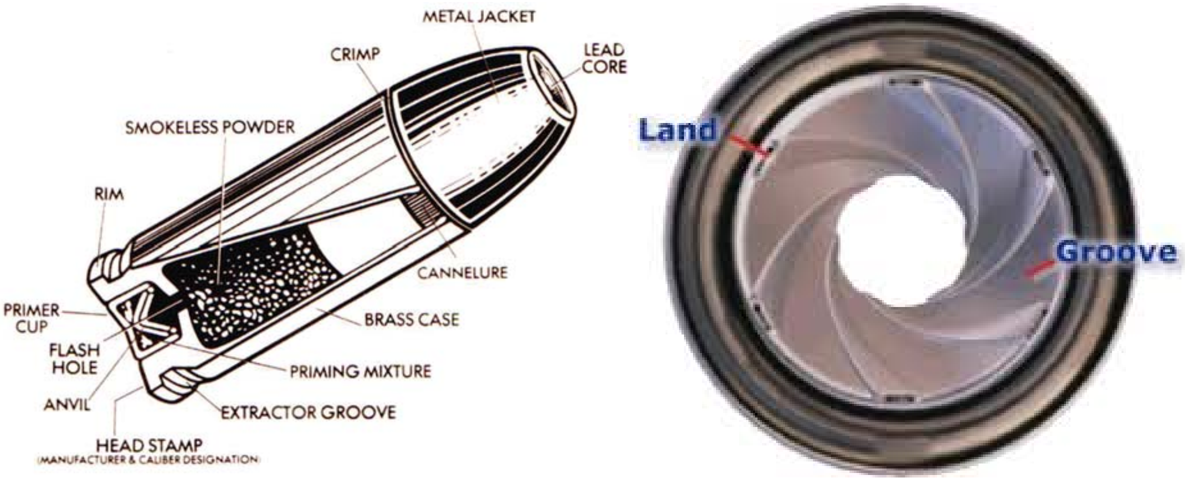


Figure 1-2: Cross sectional diagram of a bullet (left) and rifling of the barrel (right) (Parker, 2012)

The force exerted from the expanding gases of the primer and main propellant force the cartridge backwards, contacting the breechface of the firearm and engaging the ejector mechanism. This contact with the breechface may result in breechface machining marks imparted into the base of the softer cartridge case, shown in Figure 1-3 (Cork, Rolph, & Meieran, Ballistic Imaging, 2008):

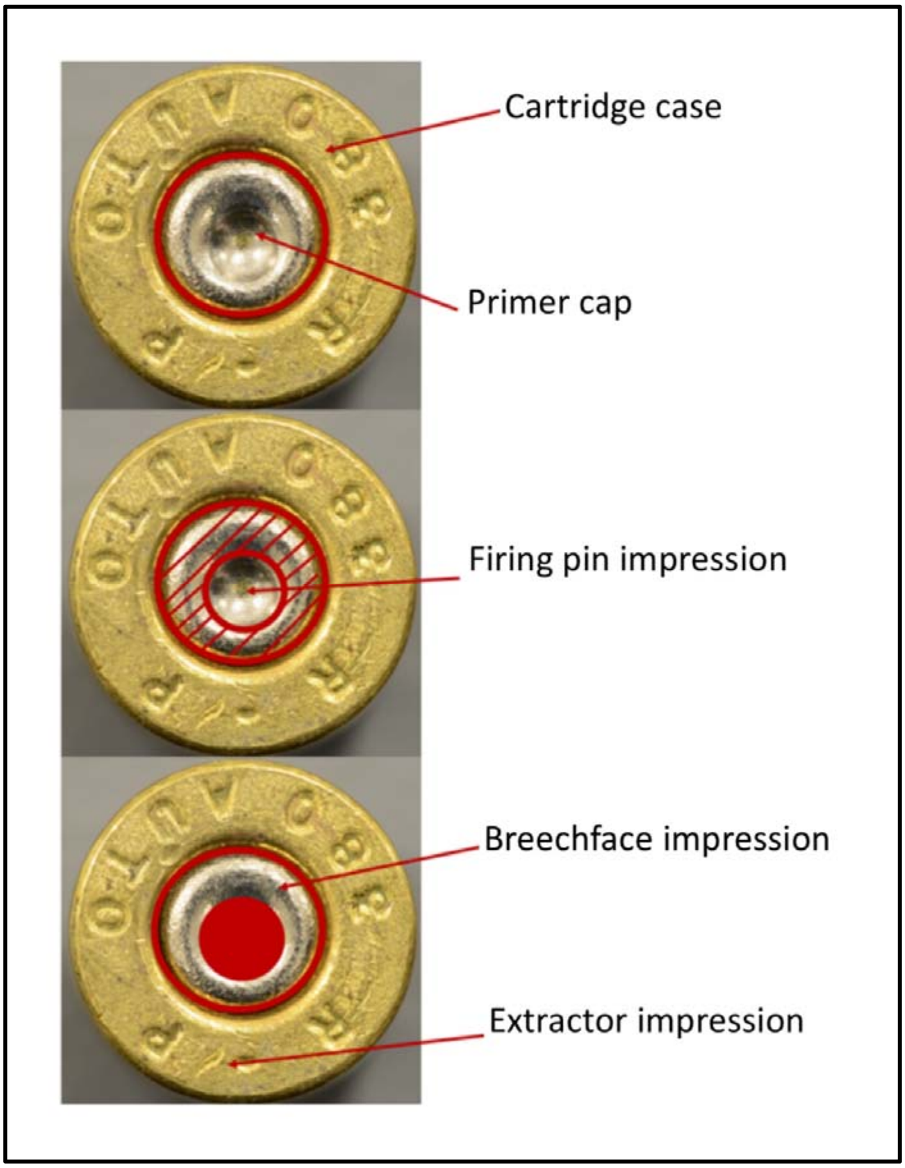


Figure 1-3: Picture of toolmarks imparted in cartridge case surface

The principles of identifying ballistic toolmark evidence began with Balthazard in 1912. He found that the cutters used to rifle the inside of barrels would not leave the exact same markings in its successive excursions through barrel blanks. After extending this preliminary study Balthazard also stated that the following toolmarks would be unique to a particular firearm:

- Striations imparted from rifling marks to the bullet surface
- Firing pin impression on the base of cartridge cases
- Breech face marks imparted onto the base of cartridge cases
- Extractor and ejector marks (Heard, 2013).

The rifling techniques used can be separated into two groups: those that remove material and those that displace material. Removal techniques include hook cutting (Figure 1-4), scrape cutting and broaching. Displacement rifling techniques include mandrel rifling, button rifling, electrochemical and hammering (Vickery, 2015).

Figure 1-4: Picture of a hook cutter (Rifling Manufacturing , 2017)

Removing metal from the barrel results in characteristic longitudinal cutting marks being imparted from the rifling tool being used. This will result in inverse toolmarks on bullets appearing as striation within the engraved area, see Figure 1-5. When metal is displaced to create rifling, the surface remains smooth, and any marks that are imparted will occur as concentric circles. As these marks are perpendicular to the bullets travel, toolmarks imparted onto the bullets surface will appear as individual marks. A combination of both methods may be used in each barrel. (Bonfanti & De Kinder, 1999). The Presence of toolmarks imparted onto a bullet can vary. Toolmark generation relies on the jacket material of the bullet, and the 'tightness of fit' within the barrel. Softer material such as lead will deform more than copper, and therefore toolmarks imparted in these materials will vary. It is also known that there is a level of variability in the Groove Engraved Areas

(GEAs) to bullets due to contact not always being present between a bullet and a GEA, dependant on the size and position of the bullet (Bolton-King, et al., 2010).

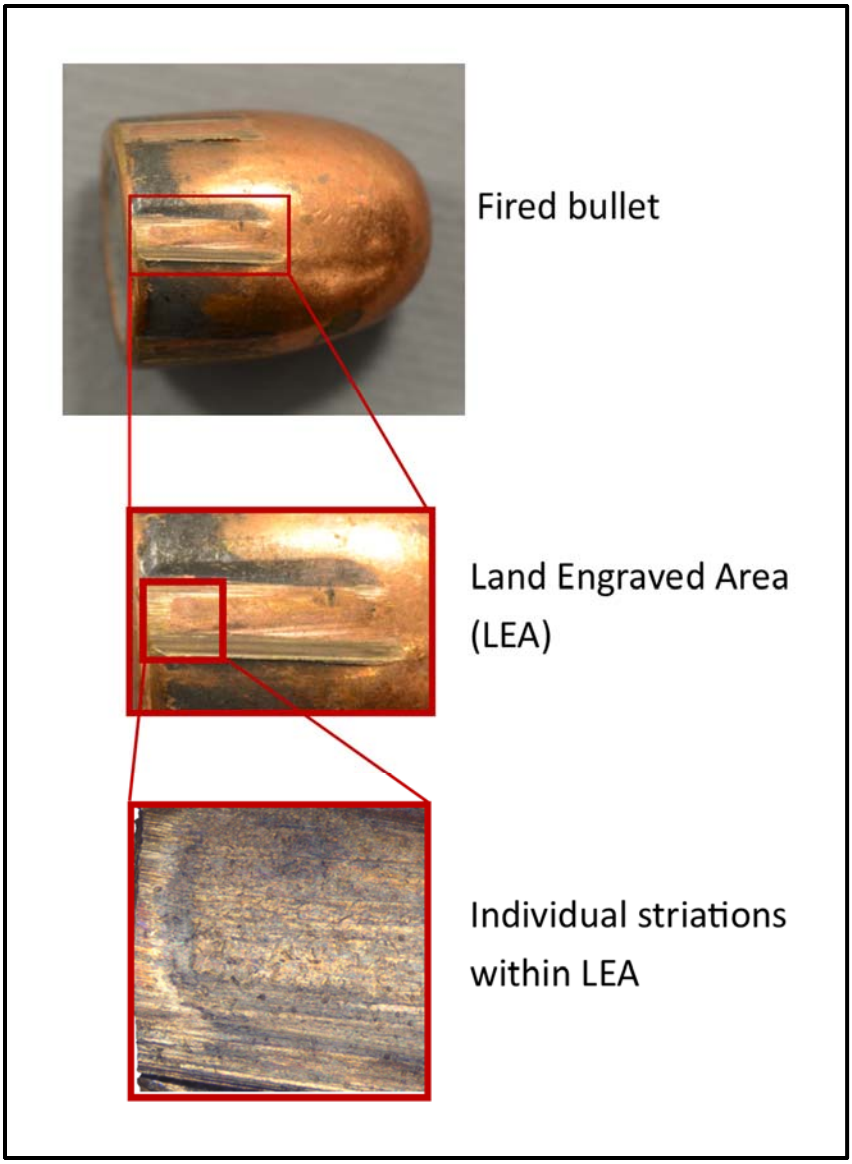


Figure 1-5: Picture of longitudinal toolmarks imparted on bullets.

Figure 1-4 (page 18) shows a hook cutter which is used to remove metal from the barrel in several passes per rifling groove. In each pass the cutter is rotated, creating a twist in the rifling. These rifling twists impart spin to the bullet (Figure 1-6), giving it a straighter line of trajectory and making the firearm more accurate (Beesly, 1961).
As a rifling button is passed through the barrel, it is able to shape all the rifling grooves through the barrel in one pass. As this is a material displacement technique, the button pushes material around the barrel to create the rifling. This displacement causes concentric circles in the barrel (Figure 1-6 (Rifling Manufacturing , 2017)).

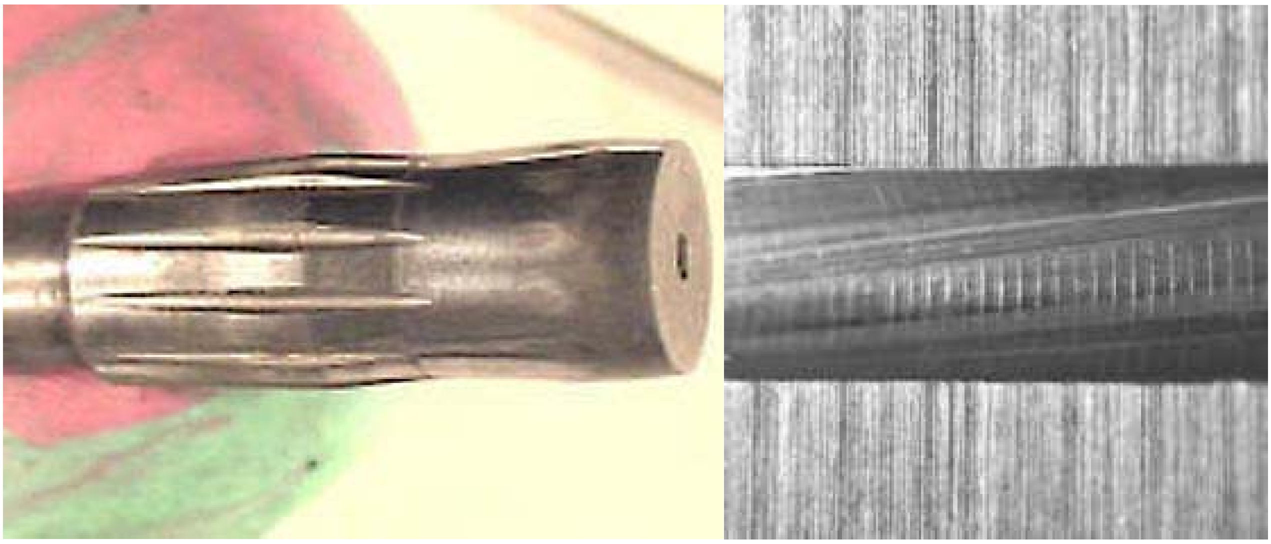


Figure 1-6: Picture of a button rifling tool and the concentric circular marks left in the barrel

The toolmarks imparted onto the surface of a cartridge can be subdivided into characteristics based on the individual characteristics. The toolmarks are classed as follows:

1.2.1 Class characteristics

Class characteristics relate to larger scale topography of the tool, giving information on the type of tool used, the overall dimensions and is characteristic of the make and manufacturer. In the case of ballistic toolmarks, class characteristics will ascertain the brand of firearm, thus narrowing a comparative search to a smaller group of manufacturers (Sarıbey & Hannam, Comparison of the Class and Individual Characteristics of Turkish 7.65 mm Browning/.32 Automatic Caliber Self-Loading Pistols with Consecutive Serial Numbers, 2013). Class characteristics from the barrel may include such information as the number of rifling grooves, the width of the rifling and the twist angle and direction of the rifling. Cartridge cases will be imprinted with class characteristics such as the overall dimensions of the firing pin, breechface and extractor mechanism, which can be seen in Figure 1-3 (page 17).

1.2.2 Subclass characteristics

Subclass characteristics are defined by the Association of Firearm and Toolmark Examiners (AFTE) as a group of surface features which are more restricted than class characteristics, in that they will apply to a small group of tools.
Subclass characteristics emerge as a result of the manufacturing processes. For example, a cutting tool may have become worn in a specific way over time, to the point of being discernible from other similar tools. Should this cutting tool be used to manufacture a small

number of firearms before being changed, each of the firearms will contain the same subclass characteristics (Nichols, 2007).

1.2.3 Individual characteristics

Individual characteristics are small- usually micron scale- marks within a toolmark that are unique to that tool. Therefore, the toolmark imparted by this tool will contain information that relates only to that one tool.

Individual characteristics occur during the manufacturing process, and are due to how the tool has been used, manufactured and stored. Individual characteristics can be formed in some of the following situations:

- Tool cutting edge manufacture
- Characteristic wear of the tool from use
- Storage of the tool in suboptimal conditions, leading to rust and corrosion of the tool surface
- Cleaning of the tool using abrasive materials, leaving marks on the surface
- Misuse of the tool causing damage (Sarıbey, Hannam, & Tarımcı, 2009).

In terms of ballistic toolmark evidence pertaining to the cartridges, this may lead to individual corrosion on the firing pin/ breechface of the firearm. Toolmarks imparted by this firearm will transfer individual characteristics into impressions on the cartridge surface.

1.2.4 Knowing the difference

It is important for a forensic examiner to understand the difference between each class of characteristic. It is not unknown for subclass characteristics to be misinterpreted and assumed to be individual characteristics (Nichols, 2007; Tobin & Blau, 2013). Should this happen, it is possible to match the toolmarks to the incorrect firearm, known as a 'false positive', as subclass characteristics could appear the same from a group of several firearms. Figure 1-7 and Figure 1-8 illustrate the different types of characteristics.

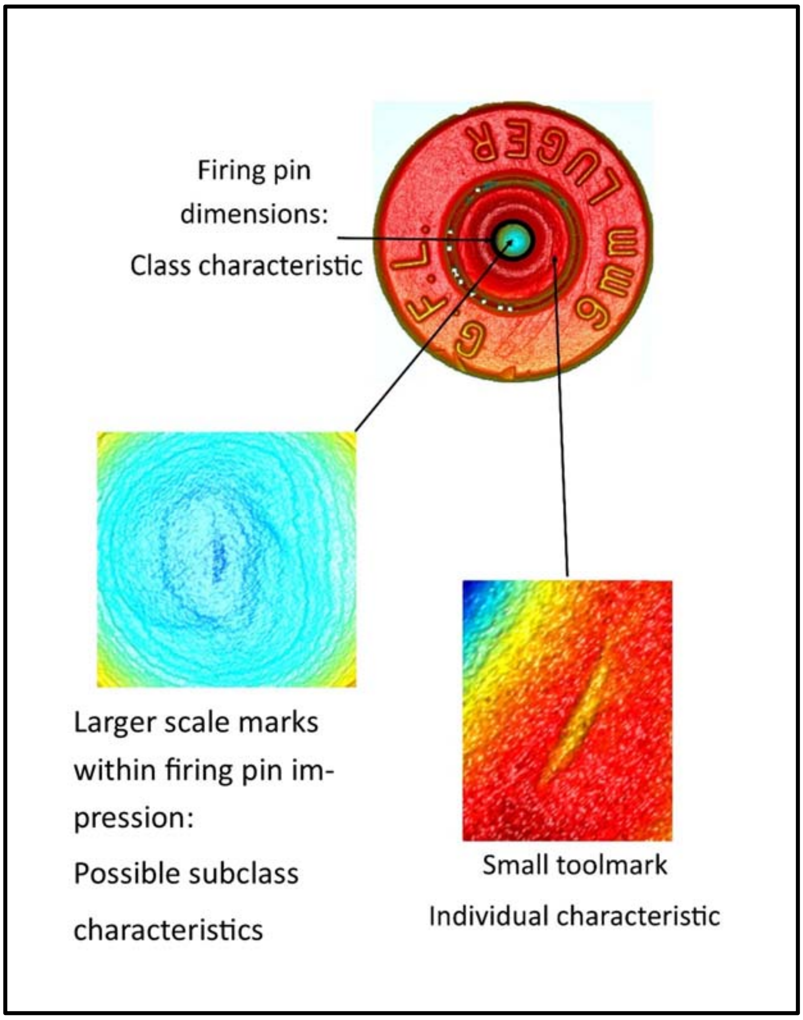


Figure 1-7: Toolmark characteristics found on cartridge cases

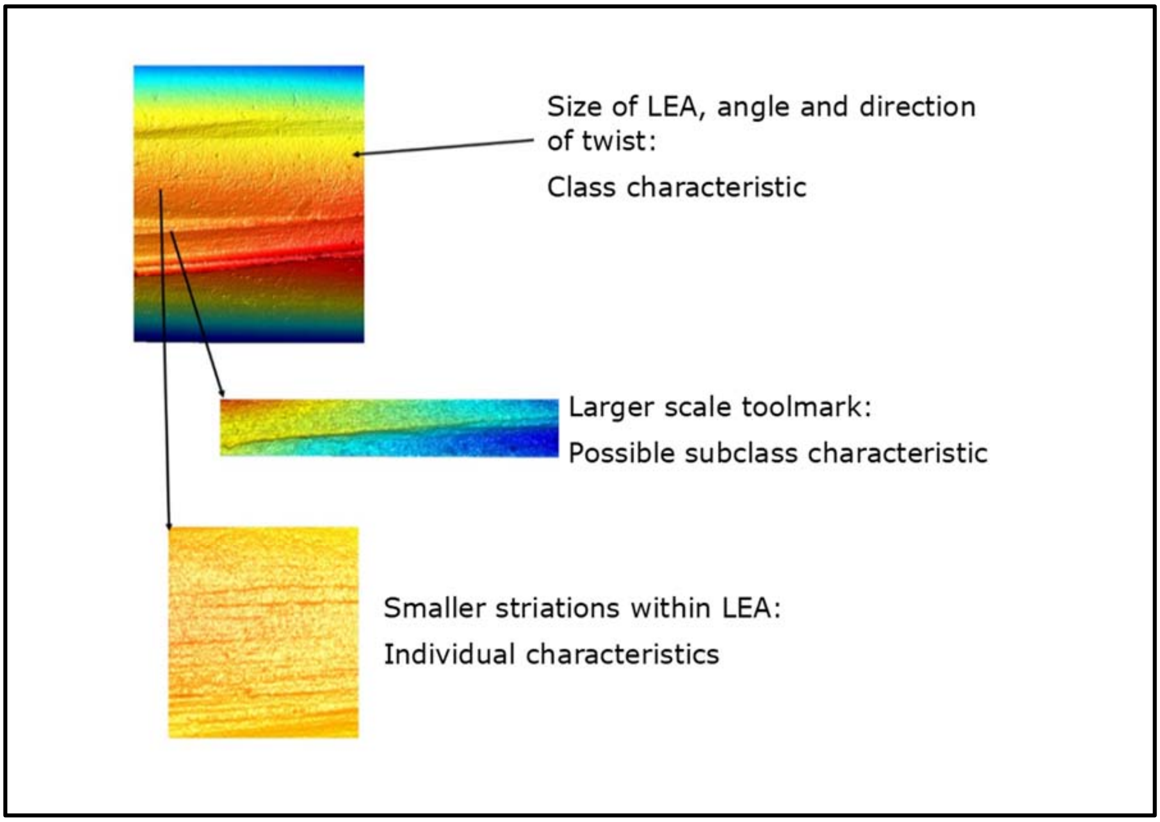


Figure 1-8: Toolmark characteristics found on bullets

1.3 Criteria for matching toolmarks

Past methodology is based upon the ability of an expert examiner to decide whether or not two toolmarks could have originated from the same tool, or the same firearm. Conclusions derived from the comparison must be 'beyond reasonable doubt'. This statement can be considered subjective, as it is not based on a quantitative result acquired using repeatable methodology, but is based on an opinion on whether the forensic examiner deems there to be sufficient agreement between the two toolmarks. Therefore, it is suggested that forensic examiners move away from this term for more standardised terms, which will be described in Chapter Two.

The Association of Firearm and Toolmark Examiners (AFTE) published a theory of identification for ballistic toolmark examiners, in which it is declared that the following statements must be true to be able to consider two toolmarks as a match (AFTE Criteria for Identification Commitee, 1992):

- There must be a significant amount of duplication between the two marks
- The agreement between these two marks is significant when it has exceeded the amount of agreement that can be found in two toolmarks that were not made by the same tool.

Should these criteria be fulfilled, then it can be said that it would be a 'practical impossibility' that two different tools were used, therefore the toolmarks can be considered a match (AFTE Criteria for Identification Commitee, 1992). While the term 'practical impossibility' was used in the AFTE theory of identification published in 1992, this term is now outdated and is being replaced by the verbal association of likelihood ratios (European Network of Forensic Science Institutes, 2010). The AFTE theory aims to introduce repeatability into the comparison of ballistic toolmark evidence, the methodology still relies on subjective techniques. Deciding whether toolmarks present have succeeded the level of agreement that would be found in non-matching toolmarks would be based purely on the experience and knowledge of the examiner. As such experience will vary significantly between examiners, the method can no longer be considered objective (Tobin & Blau, 2013).

1.4 Shifting from 2D to areal

2D or greyscale measurement methods result in an image acquisition in which no topography information of the surface is recorded. Information recorded in 2D imaging will include the x and y position of each pixel and the colour or greyscale value.

In areal measurement, such as confocal microscopy, interferometric techniques and focus variation (described in Chapter 2.5), height information is recorded for the measured surface. This results in data on the x, y and z position of each measured point on the surface. Therefore, the main difference in 2D and areal measurement is the acquisition of height data from the surface. The topography of the surface can be inferred in greyscale imaging through variation in lighting, however this is affected by variations in environmental lighting. With a shift from 2D to areal, there will be an increase in surface information, the majority of which will not relate to individual characteristics within a toolmark. Therefore, for areal datasets to be used for efficient correlation, pre-processing methods must be utilised to separate spatial frequencies within datasets. These methods will be further discussed in Chapters 2,4 and 5.

1.5 Aims and objectives

The overall aim of this thesis is to demonstrate the ability to shift ballistic toolmark measurement and correlation from 2D digital acquisition with visual comparison to advanced measurement techniques coupled with objective mathematical correlation.

Objectives:

- Demonstrate the advantages of areal measurement systems compared to 2D digital systems
- Effectively apply mathematical correlation algorithms to acquired datasets
- Gain insight into differences in quality of dataset using various optical measurement techniques
- Compare correlation efficacy in datasets acquired using different areal systems
- Complete a direct comparison of correlation results gained using 2D digital and areal based measurement systems.
- Distinguish the differences in toolmark topography when impressed into different substrate materials.
- Compare correlation results based on material composition of primer cap/bullet.

CHAPTER TWO

LITERATURE SURVEY

The purpose of this chapter is to demonstrate to the reader the need for ballistic toolmark identification and an objective approach to the identification process.

Firstly, a discussion of current firearms legislation, along with steps taken to prevent the presence of illicit firearms within the UK. This leads on to a discussion on past approaches to the measurement and identification of ballistic toolmark evidence.

Leading on from past identification methods, this chapter will discuss current methods, were a shift from greyscale to areal measurement can be seen. With a shift in measurement process comes a need for pre-processing of data, and correlation algorithms to ensure objectivity in identification, of which various methods used in current research methods will be highlighted. Methods used in this study will be created using 9mm Luger bullets and cartridge cases contained within the Odyssey collection, and the associated EU project will be described.

Finally, aims and objectives of studies within the thesis will be introduced along with novel contributions to research.

2 Literature Survey

2.1 Introduction

The following section details firearms related crime with emphasis on the UK. This includes the legislation put in place to decrease the level of availability of firearms to the general public, persons most likely to be involved in firearms crime and the routes in which firearms become available in illicit domains.

2.1.1 Firearm legislation

According to the Firearms Act (Home Office, 1968) a firearm can be defined as a lethal barrelled weapon from which a missile can be discharged. Included within the legislation are any component parts to the firearm, and any accessories which were designed to reduce noise or flash during firing. Within section 5 of the act, the following firearms are prohibited:

- Self-loading and automatic firearms
- Pump action firearms
- Barrels less than 30cm in length
- Firearms that have been disguised
- Firearms and ammunition capable of firing noxious substances and explosives

Given the prohibitions listed above, it is illegal to own the majority of pistols (in cases of persons not involved with law enforcement and military), unless a special licence has been obtained. Licences are only allowed under specific criteria, for example the legal culling of animals or as part of a museum collection and are therefore difficult for most people to obtain. Exceptions are made when a firearm is classed as obsolete, meaning ammunition is not commercially available for the firearm anymore. In this case, it is legal to own an obsolete firearm with no license, and most obsolete firearms are owned by collectors.

UK firearms laws actively change to reduce the risk of firearms related incidents. One of the most noticeable law changes came as a direct result of the Dunblane massacre, in which 17 children and 1 teacher were murdered by Thomas Hamilton. Hamilton was in possession of four legally owned handguns at the time of the mass murder. This prompted the then government to remove the ability to possess handguns legally for the vast majority of people (McGuire, 1996). However, with trafficking channels able to import firearms into the illicit domain, as described in the following section, this has not completely reduced handgun possession within the UK.

2.1.2 Firearm related crime

Crime statistics have shown that in 2005 there were 12,337 homicides and 69,825 non-fatal injuries caused by firearms in the USA. During 2005/2006, there were 49 homicides and 4,955 non-fatal injuries caused by firearms in the UK. While firearm related crime remains a small percentage of crime committed in the UK, accounting for 0.3% overall in 2008/2009, there is a definite increase in the number of crimes involving firearms in the UK. Between the mid-1990s and mid-2000s, it has been reported that the number of firearms related crimes doubled (Caddick & Porter, 2012; Davies, Wells, Squires, T.J, & Lecky, 2012). In the UK, as of March 2016 there were 153,404 firearm certificates and 567,015 shotgun licenses on issue (Turner, 2017).

As can be seen in official Home Office crime statistics, while firearm related crime did decrease from 2008 to 2013, since 2013 to March 2016 it has been increasing, and all years saw more firearm related crime than statistics from 2005, see Figure 2-1 (Allen & Dempsey, 2016):

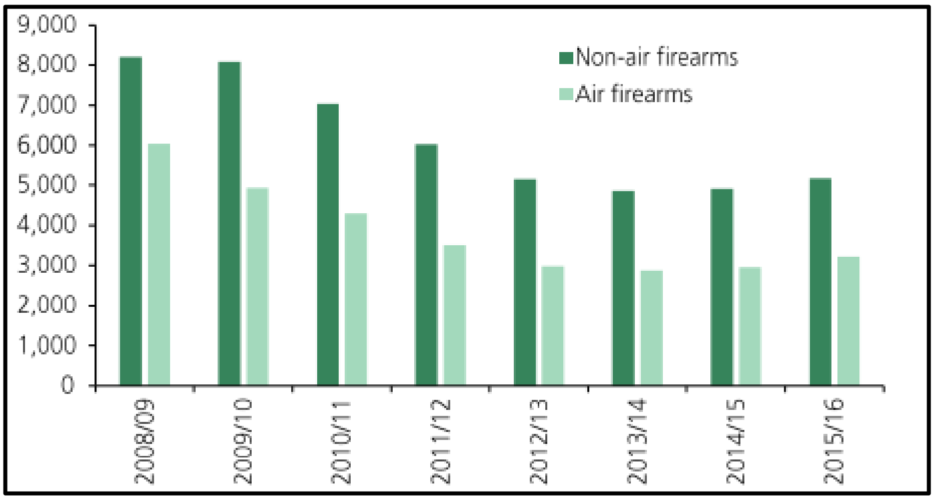


Figure 2-1: Official firearm crime statistics of England and Wales from 2008 to 2016

The increase in UK gun crime has purportedly been caused by an increase in street gang culture, similar to street gangs operating in the USA. Street gangs are known to operate using extreme violence and illegal gun ownership. However, not all research agrees this is the case. The term "street gang" is not fully accepted within criminology research, as it is believed to be too generalised a term to properly explain the increase in firearm crime. Instead, it has been proposed to examine instances of firearm related crime dependant on the lifestyle of the perpetrator. The younger perpetrator, more likely to be involved in gang like activity, has been found more likely to be in possession of less lethal firearms such as air pistols. Professional criminals, included those less likely to be operating in "gangs" such as armed robbers and those involved in drugs offenses, are more likely to be linked to firearm incidents in which a lethal firearm was used (Caddick & Porter, 2012; Hallswoth & Silverstone, 2009).

In a research study, 1,570 arrestees were interviewed with regards to their personal experience of illegal firearms. It was found that 20% of the above sample admitted to

having been in possession of an illegal firearm within their lifetime. The most common firearm within this sample was a handgun, and the most common reasoning behind the possession being "protection" (Bennett & Holloway, 2004).
Trafficking of firearms represents a large amount of the firearms that are found within the illicit domain. Most firearms found to be illegally possessed were in the first instance manufactured legally, to enter illegal possession after leaving the manufacturers (Spapens, 2007).[1]

2.1.3 Steps in the reduction of illicit firearms

Police intelligence services across the EU are constantly tracing the movement of illicit firearms and are able to collect information transnationally to limit the number of firearms being illegally imported into the UK. In two separate examples, three members of the Real Irish Republican Army (RIRA) were found guilty of the attempted importation of weapons, explosives and detonators into the UK. Members of the Lithuanian security services were posing as arms dealers, and were approached by RIRA members trying to buy the firearms (O'Neill & Hamilton, 2011; BBC, 2010) .
In the UK Operation Trident was set up in 2000 to respond to firearm related crime between young black men. Since its inception, the operation has increased to include all firearm crime. The operation now runs with four teams, with a combined police staff of 350, and 85 support staff including crime analysts. It has been found through such police investigation that 60% of all firearms offences occur in London, Manchester and Liverpool (Roberts & Innes, 2009).
In an effort to gain intelligence into the illegal possession of firearms, the Forensic Science Service (FSS) created an intelligence database, known as the National Firearms Forensic Intelligence Database (NFFID), to obtain intelligence information in firearms related crimes. From September 2003 to September 2008, 8,887 guns were submitted to the FFS (Hannam, 2010).
After the breakdown of the FSS, all firearm intelligence databases are now linked through the National Ballistic Intelligence Service (NABIS), created in 2008. NABIS runs from four hubs, located in: Birmingham, London, Manchester and Strathclyde. Through these four

[1] The author would like to note at this point, that while not directly related to the thesis, there is still an availability of reactivated and obsolete firearms within the illicit domain. Firearms must be deactivated or proofed by a proof house within Europe, and deactivation occurs so that firearm collectors may legally own the firearm. Obsolete caliber firearms may be owned in their original state as it is no longer possible to procure ammunition for such firearms. However, theft and fraud may result in firearms being reactivated with replacement barrels and firing pins to be then used illegally, and obsolete ammunition may be illegally manufactured. The most recent case of this occurring was in the case of firearms dealer Mr. Paul Edmunds, who was responsible for the manufacture of ammunition of obsolete firearms found in 90 separate crime scenes. Edmunds appeared in court in February 2017 and was sentenced to 30 years in prison, thus demonstrating a current availability of such firearms (McGuire, 1996; Spapens, 2007; Crowson, 2017).

hubs it is possible for the NABIS system to share and evaluate all ballistic evidence and information throughout the UK. The system relies on optical microscopy, i.e. 2D images, and comparators used by the Integrated Ballistic Identification System (IBIS), which is manufactured by Forensic Technology Inc. (FTI) (NABIS, 2017). However, NABIS have now introduced areal measurement systems using confocal microscopy in partnership with FTI. The NABIS system operates with each lab maintaining their own open case file (OCF). An open case file acts as a library for evidence (cartridge cases and bullets). NABIS have reported that within their OCFs, 200 hits a year occur, of which the majority are assumed to be cold hits in which the evidence has matched to evidence in other case files, but the firearm has not been identified.

The most recent example of evidence presented by NABIS being used in a court of law involves an antique firearms dealer using his knowledge to create ammunition for obsolete firearms procured legally from the USA. Using the NABIS system, it had been found that there had been an increase in obsolete handguns being recovered by police, starting in 2009. As ammunition for an obsolete firearm must be home-made, expert examiners found upon further investigation that where ammunition for these obsolete firearms had been discovered, the manufacturing processes for the ammunition were very similar, indicating that they were originating from the same workshop. In total, evidence from 90 separate crime scenes from across the UK was found to be linked, which resulted in the firearms dealer Mr Paul Edmunds appearing in court in February 2017 and ultimately being jailed for 30 years (Crowson, 2017) for the conspiracy to supply firearms and ammunition.

NABIS evidence is also further linked to forensic laboratories across part of Europe, through the International Ballistic Intelligence Network (IBIN). The IBIN system also relies on the IBIS system owned by the FTI. IBIN serves to connect ballistic evidence from the following countries: Spain, Portugal, the Netherlands, UK, Ireland, Denmark, Sweden and Norway (this was true at the end of 2010, however it is expected that other INTERPOL member countries will now be involved) (INTERPOL, 2017). In a 2017 report published by the National Crime agency, it was reported that a key concern in firearms supply are routed from Belgium, the Netherlands and France, which are considered 'key nexus points' for the importation of illicit firearms (National Crime Agency, 2017).

Operating alongside IBIN, the National Integrated Ballistics Intelligence Network (NIBIN), serves the USA in the storage and comparison of ballistic toolmark evidence, and is once again run using the IBIS correlators. The network was set up in 1999, and by August 2009 contained 1.5 million acquisitions. Within these acquisitions, 32,000 hits have been reported. However, it is unknown how many of these hits can be considered 'cold hits' (no firearm information) or 'warm hits', where the evidence can be directly linked to a known firearm, through test fires of a seized firearm being added into the IBIS system (De Ceuster, Hermsen, Mastaglio, & Nennstiel, 2012). As of March 2016, 2.8 million images have been uploaded into the NIBIN system, and reported hits have increased to 74,000 (ATF, 2016).

All of the above systems are used to be able to save evidence into libraries, with subsequent correlation systems designed to minimise the library sizes using class characteristics and comparison of individual characteristics, thus showing the expert examiner a number of possible matches. As such, it is still within the remit of the expert examiner to decide whether or not evidence can truly be considered a match, and therefore techniques used in the systems can still be considered pseudo subjective (De Ceuster, Hermsen, Mastaglio, & Nennstiel, 2012).

The correlators used within any IBIS system are considered commercially confidential of the FTI, therefore there is very little research into either the methodology used in the system or the efficacy of the correlations.

One study (De Kinder, Tulleners, & Thiebaut, 2004) details the efficacy of the correlations, but was unable to state the correlation methodology being used. The study used 600 9mm SIG Sauer pistols, with each pistol being shot seven times. This created a reference database containing 4,200 cartridge cases.

One Remington cartridge fired from each pistol was considered the reference cartridge, and had known matches within the database, one known match being from the same cartridge manufacturer and six other known matches from different cartridge manufacturer. Correlations were performed, to ascertain whether or not the known matches would be indicated when in a large database.

The results found that 72% of known matches would appear in the top ten of a hitlist when a cartridge manufactured by the same company was used. When a different manufacturer's cartridge was used, only 21% of known matches would appear in the top 10. Therefore, it was concluded in this research that the results would not be acceptable in a forensic context (De Kinder, Tulleners, & Thiebaut, 2004). The measurement systems currently being sold by the FTI will be discussed further in the next section.

Even with such databases cataloguing firearms and ballistic evidence within the UK, it is still very difficult to be certain of the number of illegal firearms in circulation. Estimations have reported anywhere between 200,000 and 4 million possible firearms (Bennett & Holloway, 2004). This is partly due to the unknown quantity of "guns for hire" across the UK.

Guns for hire pose a larger problem in the UK than the USA, due to less availability of legal firearms within the UK. Therefore, it is more likely that firearms will be used in multiple crimes in the UK, in the USA firearms are more likely to be discarded after use in one crime. An example of guns for hire was reported in Birmingham, where two gangs were using the same firearm against one another, each hiring the gun in turn from a separate third party (The Economist, 2013). As such circumstances within the UK result in the same firearm being used in multiple crimes, it supports the need for a system that is able to correlate with a greater level of efficiency, as links between ballistic evidence would become clear, thus creating the ability to track the guns for hire effectively.

2.2 History of ballistic toolmark identification

It is reported that the first instance of ballistic toolmark evidence occurred on the second of May 1863, when General Stonewall Jackson was shot, and later died, as he returned to his own troops. During medical intervention, a .675 spherical ball was retrieved, and investigations into the caliber and type of ammunition used found that ball must have been shot from a confederate smooth bore musket, and thus it was concluded that the General died as a result of friendly fire (Mackowski & White, 2013).

There is also a reference to ballistic toolmark evidence reported in June 1900, by the Surgeon Albert Llewellyn Hall who published reports on the possibility of identifying ballistic toolmark evidence, and similarly in 1907, when Frankfurt Arsenal staff were asked to identify bullets during the soldier riots in Brownsville, Texas (Warlow, 2012).

n 1912, the theory of ballistic toolmark evidence evolved when Balthazard became the first to use photography in the comparison of these marks (Bell, 2012).

During 1916, dubious firearms evidence presented in a trial resulted in Mr Waite, of the attorney’s office, travelling around America gaining a better understanding of the manufacturing processes used to manufacture firearms. Using this information, three of his associates were able to develop the understanding of ballistic toolmark evidence.

Major Goddard, a firearms expert, Phillip B. Gravelle, a trained microscopist and John E. Fischer, a machinist, set up a forensic firearms laboratory, and in 1925 they bought a comparison microscope to use in the identification of ballistic toolmarks. Goddard went on to publish many articles on the subject and testified in courts numerous times. Goddard later became the head of the Scientific Criminal Investigation laboratory in 1930 and was considered an expert in the field. It was not until Goddard was visited by gunmaker Mr Robert Churchill that the methodology of visual comparison of evidence under comparison microscopy was adopted in the UK (Warlow, 2012). Figure 2-2 (Leeds Micro, 2013) shows a traditional 2D comparison microscope, in which there are two stages to be able to simultaneously compare two separate pieces of evidence. Each stage is monitored by a camera so that software manipulation allows evidence to be overlaid and viewed side by side on a computer screen.

Figure 2-2: Image of a modern 2D comparison microscope

Since its inception, the techniques used in ballistic toolmark evidence have changed little. Forensic laboratories still rely on the use of comparison microscopy and visual pattern matching to ascertain whether or not two cartridge cases/bullets could have been fired from the same gun.

Both the toolmarks imparted on the base of the cartridge case and the striations on bullets are viewed by comparison microscopy. Two objects are viewed side by side, by using the optical bridge (Figure 2-2), and oblique lighting is used to accentuate the peaks on the surface while shadowing the valleys. The subjectivity in this method comes mainly from the lighting techniques, as using different lighting angles on incidence will change the shadowing on the surface. As it would be difficult to ensure that the exact same lighting conditions are used across all forensic laboratories, it is considered a non-repeatable technique (Baiker, et al., 2014).

Using the standard examination method for identification, as outlined by the Scientific Working Group for firearms and toolmarks (SWGGUN), four steps must be carried out in an examination as follows (NIST, 2014):

- Evaluate the class characteristics present. This allows evidence with non-matching class characteristics to be excluded

- The comparison of subclass and individual characteristics using comparison microscopy
- Conclusion of findings. Is there sufficient agreement between the evidence to consider a match?
- Independent verification of findings by another toolmark examiner

It has been reported that using this methodology, should 20 individual characteristics be found to match across two pieces of evidence, there is a 1 in 193,730,707,456 chance that the object was not fired by the same gun (Bolton-King, 2016), however such circumstances have not been reported to have actually happened. Using Biasotti's research on consecutive matching striae in bullets, it was found that in some cases there would be no more than four consecutive matching striae present on bullets known to be fired from the same firearm. Due to the possibility of non-matching bullets also containing 4 consecutive matching striae (CMS), Biasotti (Biasotti, 1959) set a threshold of 6 CMS (when comparing greyscale images) to determine whether or not two bullets were fired the same gun. Statistically, this would mean that there was a 1 in a 110 chance that the bullet was not fired from the same gun (Bunch, 2000).

2.2.1 Ballistic identification systems

A ballistic identification system describes the combined hardware and software needed to acquire images/datasets of toolmark evidence, along with the subsequent saving, sharing and comparison of evidence. A ballistic identification system usually compromises of a measurement system and dedicated software, a database able to save images/ datasets acquired and share the data between laboratories, and a software system able to compare the data. Each element of a ballistic identification system will be able to run separately from one another, and so it is possible to acquire the measurement of evidence in one laboratory, and compare the evidence in a different location.

The first instance of such a system set up to share ballistic toolmark evidence was established in 1992, when the ATF (Bureau of Alcohol, Tobacco, Firearms and Explosives) implemented 'CEASEFIRE'. CEASEFIRE was set up as an information sharing database, so that the information on firearm related crime such as suspected caliber of firearm, location of crime and other intelligence could be shared across any number of ATF laboratories. In 1993, FTI released their first ballistic identification system, known as 'BulletProof', which could acquire digital 2D greyscale images of bullets only, and compare images to a library of evidence (Walton R. , 2006).

During this time, the FBI also released their own identification system known as 'DRUGFIRE', which was able to acquire and compare 2D images of cartridge cases. With two large scale identification systems now being used, it was found that there were issues in sharing data across forensic laboratories, as the two systems were not compatible, both in terms of the software and the fact that the systems were exclusive to either the identification of bullets or cartridge cases. To overcome this FTI released another system known as 'BrassCatcher', which focussed on the identification of cartridge cases. The combination of BrassCatcher and BulletProof became known as IBIS- the Integrated Ballistics Identification System. This system is now known as IBIS Heritage, due to FTI systems becoming more advanced (Walton R. , 2006).

After the implementation of the IBIS systems, the DRUGFIRE system did continue to be manufactured. Owned by the FBI but developed by Mnemonic systems, in 1996 the system was installed in 80 laboratories across the USA. The DRUGFIRE system was originally implemented as a library system to enable the large backlog of ballistic evidence to be inputted into one online system, meaning that evidence entering the lab could be compared to evidence in any OCF more efficiently. The system used 2D digital image capture by microscopy to acquire images of the evidence, which could then be compared within the software. Guidelines on the correct lighting and positioning of evidence were issued, however it was found that due to the ability to change the lighting settings measurements still resulted in acquired images that were suboptimal for comparison (Jones & Guerci, 1997).

Once images were uploaded into the system, the software was able to convert the images to binary imaging, in which each pixel could only have two possibilities, either black or white. These binary images were then used for the comparison of evidence (Jones & Guerci, 1997). Binary image methods using 2D imaging were also implemented by the FIREBALL system, developed by the Edith Cowan University in Australia (Huang & Leng, 2010; Li, 2003).

The IBIS Heritage system, also retailing in the mid-1990s, cost forensic laboratories $500,000. The system could compare evidence and give a hitlist of any potential matches, which were then compared by a forensic examiner. FTI claimed that this system would be able to compare a library of 1,000 bullets to a bullet in under an hour, thus greatly reducing the time needed to decrease match possibilities by a forensic examiner. During this time FTI claimed that the FBI were unfairly cornering the market with their own DRUGFIRE system. This resulted in FTI hiring a lobbyist to ensure that they could contact a fair share of the market when this was discussed in the US courts. Since then FTI began to sell the IBIS system worldwide, and they are now present in a larger percentage of forensic laboratories than any other system (Sutherland, 1996). As DRUGFIRE and IBIS were incompatible, the two systems were combined by the FBI as the result of a memorandum of understanding between the FBI and ATF as part of the NIBIN project (Thomson, 2017).

More recently, other systems have entered the market. The 'CONDOR' system used a camera-based technique to acquire 2D digital images, and was able to stitch these images together in the x and y axes to create larger images if needed. The CONDOR system was a predecessor to the Evofinder system, which is considered one of the more modern systems (Evofinder, 2017). Similarly, the Russian company Papillion also offers a system known as 'ARSENAL', which is based on 2D image capture and correlation, and uses a variation of lighting settings to acquire numerous images of one piece of evidence to ensure all surface information is captured (Papillon Systems, 2017).

The Evofinder system was used in a study in which an older 2D system was compared to a newer system to determine the increase in efficacy as computing power increased. Mentioned previously in Chapter 2.1.3, the IBIS Heritage system was used to compare 4,200 cartridge cases fired from 600 9mm IGSIG Sauer pistols. Using the same cartridges as in this previous study, the data was entered into the newer Evofinder system. While 19% of the known matches entered were ranked in first position in the IBIS system, this increased to 38% using the Evofinder system. However, it was also found that 36% of the known matches were placed outside of the top 30 probable matches. It is believed that the low matches were due to differences in the cartridge case material (De Ceuster & Dujardin, 2015).

Since this research Evofinder have released a newer system, which claims to use 3D acquisition[2] in the measurement of ballistic toolmark evidence using methods similar to focus variation, however, there is no published evidence of this. It is the opinion of the author that correlation offered by this current system relies on 2D images based on information from the supplier website, however this also cannot be supported (Evofinder, 2017).

Another current example of a system offering 3D capabilities is the Czech system Balscan. This system also claims the acquisition of 3D datasets with the ability to search and compare within the database. The Balscan system uses a digital camera connected to a telecentric lens with a "laser focus", and quotes a resolution of 3µm per pixel (Balscan, 2017). When approached by the author at a conference, Balscan were unable to demonstrate the correlation software offered with the system. Based on the image capture hardware, it is the author's opinion that data acquired is not 3D, but rather photometric greyscale. Photometric greyscale images are able to infer the surface topography using the difference in light intensity across an image.

Cadre forensics are currently offering the TopMatch-GS 3D system, which relies on GelSight imaging for acquisition. Datasets can then be compared, using side by side visual comparison or software generated 'heat maps' of geometrical similarity. In May 2017, it was

[2] It is worth noting that in the case of forensics, the use of the terms 3D and areal have become interchangeable. While from a surface metrology point of view this is incorrect, the author will quote the use of ''3D' as mentioned by the source. It is reasonable to assume that from a surface metrology point of view, the term 3D used to describe these systems will actually refer to areal topography.

announced that the FBI would use this system in their investigations (Cadre Forensics, 2017). GelSight uses a clear elastomer covered by a reflective elastic skin. Cartridge cases are pressed into the skin, thus leaving an impression in the skin. A camera acquires images of the impression from the opposite direction, using a series of red, blue and green lighting. This lighting allows acquisition of height data using photometric stereo, in which the height of the surface is inferred by the light intensity across each point (Johnson & Adelson, 2009). The newest FTI system, known as 'BulletTrax-3D' (bullet identification system) and 'BrassTrax-3D' (cartridge case identification system) also claims to have moved to the acquisition on 3D data using confocal microscopy (FTI, 2017). When approached by the author, it was stated that new IBIS systems would be able to correlate with images in the IBIS Heritage systems. This suggests that in the correlation stage, it must be that 2D images are still being used when correlating between the two systems.

With the exception of the Alias system (which will be covered in detail in Chapter 3.5) the only system that has published research on the use of 3D acquisition systems is SciClops. The SciClops system is able to use confocal microscopy to acquire datasets of the areal topography of bullets. While in this case it is definite that advanced measurement systems have been implemented, the correlation relies on 2D data. Using the areal datasets acquired, the system will average the surface topography of a bullet into one 2D profile, which is then used for further correlation (Bachrach, Development of a 3D-based Automated Firearms Evidence Comparison System, 2002).

It should be noted at this point that the author has neglected to detail any of the comparison methodology used in a ballistic identification system. This is due to the fact that comparison algorithms used within any commercial system are commercially confidential, and as such there is no published information on the comparison methodology. The author did directly contact two of the system manufacturers and was refused any information due to this. NIST, however, do have an open approach to correlation methods used in their research, which will be discussed further in Chapter 2.8.

2.2.2 The rank order approach

The commercial systems detailed above rely on a rank order approach for the correlation of ballistic toolmark evidence. In a rank order approach, the correlation process follows the following steps:

1. Evidence is acquired and uploaded into comparison system
2. Potential matches within the database can be filtered by using class characteristic information i.e. a 9mm cartridge will not be correlated against a .22 cartridge as they cannot be fired from the same firearm
3. Correlation occurs using the reduced database based on matching class characteristics

4. The identification system gives a hitlist of possible matches, which must be confirmed by the examiner (Brinck, 2008)

In the case of FTI systems, the correlation of evidence involves three steps. Firstly, the database of evidence is reduced by disregarding evidence containing different class characteristics. A first pass correlation then occurs using subsampled datasets, in which only the larger scale features are compared. Further evidence can be disregarded based on whether or not these larger scale features match. A second correlation is then applied to the remaining evidence, which focusses on the smaller scale features within the toolmarks. Correlating using the smaller scale features is more time expensive than using the larger scale features, and so the first pass correlation is used to reduce the overall correlation time. The results of this correlation will give a match score, with a higher score indicating a better match (Cork, Nair, & Rolph, 2010).

The match score given in correlation is ambiguous, with different systems using different methodology. While the FTI approach gives a higher score for a better match, other systems are known to give a lower score for a better match (Thomas J. , 2011). This shows that the level of ambiguity in the correlation systems excludes the possibility of interoperability, as the methodology used by the systems are different. As methodology used for correlation in each system is commercially confidential, a direct comparison is not possible.

2.3 Criticisms in investigative methods

The criticism of methods used in forensic science as a general subject has existed since 1920, and through these criticisms the reliability of forensic science overall has increased (Committee on Identifying the Needs of the Forensic Sciences Community , 2009; Roady, 2017).

On November 28th, 1920, a man was shot in the temple and killed in Washington, D.C. The pistol was left at the scene and fingerprints were found on the bricks of the building, but there was no other physical evidence. No information on the murder was gained until August 23rd, 1921, when, while being interviewed for a separate crime, a Mr Frye admitted to the murder due to self-defence. When interviewed by his own legal counsel, Frye then recanted his statement, saying that he had only confessed to the murder to be able to share the reward money with the detective. With no evidence linking him to the crime, Frye was found guilty but was not given the death sentence. While imprisoned, Frye took a lie detector test, which inferred Frye was telling the truth in saying he was innocent. There were arguments concerning the evidence, and it was decided in court that scientific evidence could only be deemed admissible when the methodology had been generally accepted within its field (Kaufman, 2001; Weiss, Watson, & Xuan, 2014). This decision became known as the Frye standard, and was applied to all US criminal cases until 1993. The ambiguity of the Frye standard was highlighted in the case *Merrell vs Dow pharmaceuticals*, in which Merrell was suing Dow pharmaceuticals over claims that the morning sickness drug, Benedactin, was causing birth defects. Merrell had eight expert witnesses testify in court that the drug did cause birth defects, however the evidence was deemed inadmissible. Evidence brought forward by Merrell was not based on any published research, and experimentation carried out was not deemed reliable (Bertin & Henifin, 1994). Following the ruling, the US courts outlined a new standard concerning the admissibility of evidence, known as the Daubert standard, in which it was stated that:

- For a judge to be able to deem the admissibility of evidence, sufficient references and resources must be supplied
- The general acceptance of a scientific methodology does not substitute for the demonstration of reliability in the method
- Statistics can only be used to reinforce conclusions
- The terms "absolute certainty" and "to the exclusion of all others" are not scientific
- Observer bias must be minimised
- Subjective conclusions are only permitted when they are based upon objective methods
- Any competency testing must replicate normal circumstances
- Experimentation must be based on scientific standards or must replicate the conditions of the crime
- Experts must have sound understanding of their field, and testify only within their field of knowledge
- Experts must adhere to the standard practices in their field, and notes on findings must be contemporaneous (Page, Taylor, & Blenkin, 2011)

With regards to the Daubert standard, a research report from the National Research Council (USA) (Committee on Identifying the Needs of the Forensic Sciences Community , 2009) went into detail of shortcomings in forensic disciplines. In the case of ballistic toolmark evidence, it was found that current methodologies:

- Are unable to take into account variabilities in toolmark transfer among individual firearms
- Cannot specify the number of points of similarity that would be needed to ascertain a level of confidence in matching
- Have no precisely defined process
- Do not consider variability, reliability, or repeatability of the measurement methods
- Continue to have a heavy reliance on subjective techniques

A report published in September 2016 by the President's Council of Advisors on Science and Technology also reported a lack of black box studies being completed on ballistic identification systems. Where studies are taking place, a small database, usually minimising variables in manufacturer/ material is used. This may lead to a misrepresentation of false positive rates. It was also noted that there is a need for proficiency testing in these systems (Roady, 2017).

Given the information above there is a distinct need to develop the research of ballistic toolmark identification. Further research would ensure that measurement can be based upon accurate and traceable techniques, automated in such a way to ensure repeatability. Further analysis of toolmarks should be dependent on quantitative techniques, giving an easily traced value, based on standardised methods rather than the ambiguous match score, in which correlation algorithms are not known.

Due to the subjective methodology being used, it is not only possible to alter the results through differentiation in measurement techniques, for example the angle and intensity of the lighting, but using a subjective method can also introduce subjectivity through user bias. User bias can be described as a phenomenon in which a forensic examiner's conclusions can be altered by other information. Bias can occur through observer effects such as: the context of the situation, the mental state of the examiner, employer's opinions and personal beliefs. Confirmation bias can occur when an examiner may selectively (whether consciously or not) process evidence which fits his/her existing beliefs (Kassin, Dror, & Kukucka, 2013).

In a forensic discipline, the most common sources of bias are when extra information has been relayed to the examiner, either with the original evidence, through communication with other examiners, or at a later stage of the investigation. It is also possible to introduce bias through the prosecution demanding a reassessment of the evidence where original results have been argued against (Kassin, Dror, & Kukucka, 2013).

Bias in the identification of ballistic toolmark evidence could be minimised by using methodologies that have no dependence on the user. This would mean using a measurement method where the user was unable to change the settings, and analysis that uses mathematical algorithms to quantitatively assess the similarity of the toolmarks, and rank the results in an open and transparent manner.

The variability of imparted toolmarks has been mentioned in this thesis. Currently research suggests that while some variability does exist, it is not to an extent to which toolmarks become unrecognisable. For example, five different 9mm Parabellum Turkish pistols were fired between 1,000 and 5,000 times, and the first and a subsequent 250th cartridge cases were measured to ascertain whether it would be possible to match toolmarks imparted after extensive firing. The pistols included a Canik 55, Kanuni 16, Sarsılmaz Kılınç 2000, Yavuz 16 and Şahin 08. It was found even between the first and 5000th firing that there was not a significant change in either the class or individual toolmarks, and thus the comparison would not be affected (Sarıbey, Hannam, & Tarımcı, 2009).

There have been studies into the reliability of toolmarks, with regards to the ability to differentiate in cases where toolmarks will be very similar. Bonfanti and Kinder provide an extensive literature review in which the results of toolmark identification in cases where differentiation between consecutively machined or sequentially numbered firearms was needed. With regards to bullets, it was found in the majority of research papers at the time (23) that there is sufficient evidence in the individual characteristics that the barrels can be differentiated, and that the bullet can be matched to the correct weapon.

With toolmarks imparted on the cartridge case, most papers (11) agreed that there was differentiation in the toolmarks between consecutively machined firearms. There was significant evidence to suggest that firing pin impressions are easily distinguishable, but the same cannot be said for breechface impressions. There was significant disagreement in the research as to whether or not toolmarks imparted by consecutively machined breechfaces are distinguishable from another, if they are not distinguishable false positives may occur where the cartridge case is matched to the incorrect firearm (Bonfanti & De Kinder, 1999).

Therefore, while research has been conducted to try and overcome both the Daubert challenge and the comments made by the National Research Council (Committee on Identifying the Needs of the Forensic Sciences Community , 2009), there is still a significant amount of research needed to be able to fully overcome these criticisms.

2.4 The likelihood ratio

When explaining evidence in a criminal investigation, the weight of the evidence must be explained in terms that are understandable by laypersons of the jury, i.e. without the use of technical language, and in such a way to reduce the risk of confusing statements.

When presenting forensic evidence in a court of law, the use of a Bayesian approach has become established in other forensic evidence types such as DNA evidence, however the same cannot be said for ballistic toolmark identification. Bayesian methodology exists when evidential value is combined with prior assumptions based on the defence and prosecution hypotheses to determine posterior odds on the likelihood of the evidence being present. The

Likelihood Ratio (LR) approach currently used to convey evidence in court is based on Bayes' theorem; where odds are formed based on two relevant propositions which are mutually exclusive (Bell, 2012).

For example, in a ballistic toolmark case, two mutually exclusive hypotheses may be: "toolmarks present on a bullet were created by the firearm belonging to the suspect" and "toolmarks present on a bullet were not created by the firearm". In this case, the prior statement would be known as the prosecution hypothesis (Hp), and the latter the defence hypothesis (Hd). Both hypotheses are collectively known as the 'prior odds'. Prior odds used in calculation must be shown to the jury, as the probability is conditional based on information the jury would otherwise not know. As the jury are laypersons their knowledge will differ from the experts, and without knowledge on the prior odds the jury may become confused (Edmond, 2015).

The likelihood ratio describes the probability of the evidence occurring given the defence/prosecution hypotheses and is denoted as L(E|Hp) and L(E|Hd) respectively.

The 'posterior odds' (Equation 1) give an indication of the strength of the defence/prosecution hypothesis with regards to the evidence found in the forensic examination. Pr(Hp|E) denotes the probability of the prosecution hypothesis given the evidence (Hp is once again substituted for Hd when considering the defence hypothesis). Posterior odds are calculated by multiplying the prior odds by the likelihood ratio as shown in Equation 1 (Nordgaard & Rasmusson, 2012):

$$\frac{\Pr(Hp|E)}{\Pr(Hd|E)} = \frac{L\ (E|Hp)}{L\ (E|Hd)} \times \frac{\Pr(Hp)}{\Pr(Hd)}$$

$$posterior\ odds = likelihood\ ratio\ \times prior\ odds$$

(Equation 1)

This logical framework leads naturally to unifying principles in the evidence, and the weight of the posterior odds calculated is given using the correct terminology which has been agreed by ENFSI (European Network of Forensic Science Institutes, 2010) and is shown in Table 2-1:

Table 2-1: List of verbal associations to posterior odds

Likelihood ratio	Verbal associations (two recommended options)
1	Does not support one hypothesis over the other Evidence provides no assistance in addressing the issue
2-10	Provides weak support for the first proposition (prosecution hypothesis) relative to the defence hypothesis Findings are slightly more probable given one proposition relative to the other
10-100	Provides moderate support for Hp Evidence more probable given Hp than Hd
100-1000	Provides moderately strong support Evidence is appreciably more probable
1000-10,000	Provides strong support Evidence is much more probable
10,000-1,000,000	Provides very strong support Evidence is far more probable
Over 1,000,000	Provides extremely strong support Evidence is exceedingly more probable

Due to the hierarchy of propositions, the use of the terms 'guilt' or 'innocence' are not within the province of the forensic scientist, and so the verbal association listed above is an effective way of ensuring that this will not be the case.

Within an investigation, there are three levels of proposition, starting with the lowest, source, then activity, and finally offence.

It is within the remit of the forensic examination to determine source level propositions, for example "is the toolmark likely to have been made by this firearm". Activity level propositions address the question of how the evidence came to be part of a crime scene, and should be addressed by expert examiners only. The offence level proposition addresses whether a crime has been committed and are within the remit of the jury only (Lucy, 2005; Nordgaard & Rasmusson, 2012; Evett, 1998).

To be able to apply the Bayesian framework to advanced ballistic toolmark evidence, it would first be necessary to gain an understanding of the error rates involved in the systems, as they will have an effect on the likelihood ratio.

The likelihood ratio is the odds of the evidence given the defence/prosecution hypotheses. So, in a ballistic analysis system which theoretically always ranks a known match as the best match within the system, the theoretical error rates will be very low. However, in a system where known matches are not always ranked best match, the error rates of the evidence will be larger. With regards to the likelihood ratio, a system with larger error rates will lower the odds of the evidence given the prosecution hypothesis, as there is a greater chance of false positive matches.

To ascertain the strength of the evidence for a ballistics analysis system, it would be necessary to find how often the system will give a false positive result. Such studies would involve entering a large sample size of the same bullets and cartridge cases into each system and testing the proficiency of each systems capability to determine the likelihood of false positives, and therefore the weight of the evidence with regards to the Bayesian framework. As there are currently over 2.8 million images within the NABIS system, the author believes that sample size should not be limited, and should instead be an ongoing study. In Chapter six, the author has completed a study on the strength of evidence based on findings in Chapters four and five using likelihood ratios. However, it should be noted that the study aimed for comparison between the various methods only, and should not be used in a forensic context (NABIS, 2017).

2.5 Current research into the use of advanced metrology techniques

The following section will go into detail of the advanced metrology techniques that are currently being applied to ballistic toolmark identification but have not yet been proven for use in a court of law.

Each measurement technique has been widely accepted in the field of surface metrology, and are based upon various ISO standards (such as ISO 4288) to ensure that the measurements acquired are repeatable and reliable. Coupled with the ability to acquire areal datasets with high resolution, the measurement techniques are all able to advance ballistic toolmark measurement, and the advantages/disadvantages of each will be discussed below.

2.5.1 Stylus measurement

Stylus profilometry uses direct contact to a surface to determine its texture. A spherical tipped cone is traced across a surface, with the height described by the tip path of the surface. As the tip is deflected in the z axis by the surface texture, the vertical movement of the stylus is transformed into electrical signals by a transducer. The changes in electrical charge through oscillation are converted into a digital signal, and are used to plot the height of the surface, as the tip is scanned in the x and y plane by precision stages. The lateral resolution of the instrument is dependent on the user, who is able to decide the distance between measured points on the surface, and the minimal resolution is determined by the sensitivity of the transducer.

The vertical range of stylus profilometry depends on the dynamic range of the transducer, and the resolution is limited by mechanical vibrations and thermal noise within the instrument. However, the millimeter range and nanometer resolution of this instrument is adequate for the measurement of ballistic toolmarks. As the only system reviewed that will make a direct physical contact with the surface, it will not suffer from light reflectance issues found in optical measurements (Whitehouse, 2011).

There are two main disadvantages to using stylus profilometry in the measurement of toolmark evidence. Firstly, the physical contact may result in the surface being scratched. Any damage to forensic evidence must be avoided, as it will affect any further measurements needed during court proceedings (Vorburger, Song, & Petraco, 2015). Secondly, the dimensions of the tip used to measure the surface may impact the measurement result. Should there be very fine surface texture, for example the striation marks on a bullet, the tip may be too large to fully penetrate valleys. In steeper peaks, the tip may measure the surface on its flanks rather than its lowest point. As such, it is possible for the measurement to become 'rounded', where deep valleys have been cut off and the steeper peaks are measured as a rounded point rather than a pointed one. Therefore, in some cases the surface measured by the profilometer is not always a true indication of the surface, as demonstrated in Figure 2-3 (Whitehouse, 2011; Thomas T. , 1999; Faden, et al., 2001).

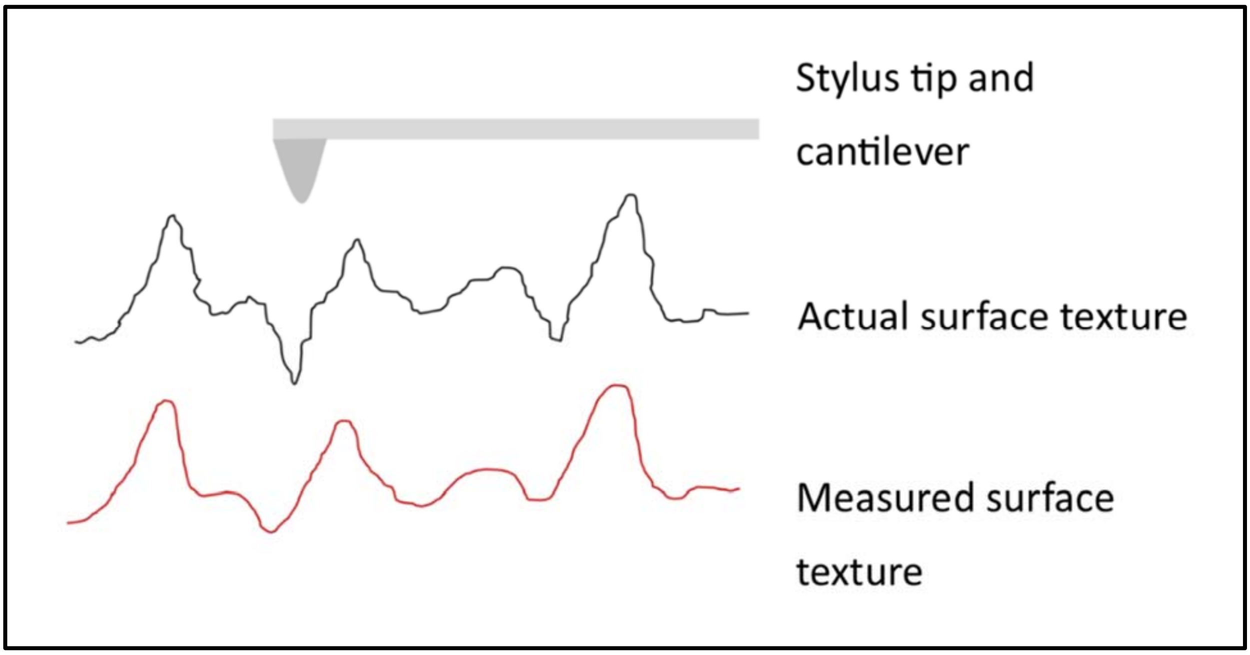


Figure 2-3: Mechanical filtering of surface using contact profilometry

2.5.2 Confocal Microscopy

Confocal microscopy differs from traditional microscopy in that areal topography of the surface is acquired by measuring the path length of a light/laser beam reflected from the surface, via a form of focus point monitoring.

Illuminating light passes through a pinhole and through a beam splitter before being focussed by an objective lens to a diffraction limited spot of light at a given focal plane. Scanning the objective lens vertically determines the intensity of the light, with the highest intensity being the best focus on the surface (Vorburger, Song, & Petraco, 2015).

Reflected light is then collected and focused onto a photodetector by moving the objective lens in z. Moving the stage in x and y allows for a larger area of measurement assuming a motorised stage is available. The path length of this reflected light gives height information for the point being measured. For each point the measurement stepper motors enable the x, y and z information to be stored, thus a true areal dataset is acquired. The technique is known as confocal as the focal planes of both the illumination of the surface and the imaging of the reflected light is coincident.

As each measurement acquires surface information only on the illuminated surface area, which tends to have a spot size of around 500nm, the stepper motors enable the raster scan of a surface in the x and y directions, thus building a topographical area of the surface (Hamilton & Wilson, 1982; Udupa, Singperumal, Sirohi, & Kothiyall, 2000).

While confocal microscopy can acquire dense areal datasets of a surface, there are some disadvantages to the technique. Due to the small spot size and optical methods, reflectance issues can result in outliers, where highly reflective surfaces result in optical spikes appearing as part of the surface, or dropout: where the angle of reflectance results in light not being collected by the photodetector, which is around 15° for confocal systems. Due to this, confocal microscopy may not be able to measure steep flanks, for example within a firing pin impression. As confocal microscopy utilises a very small spot size to acquire information, the acquisition of a full toolmark surface can take a longer time than methods with a larger field of view, and as such the author surmises it may be too time expensive in a forensic laboratory with a high turnover of evidence. These instruments are calibrated using step height standards (Evofinder, 2017; FTI, 2017).

In Figure 2-4, a disk scanning confocal microscope is described. Pinholes present on the disk are arranged in various patterns. Light passes through each pinhole and is reflected back to the detector through the same pinhole (Leach, Optical Measurement of Surface Topography, 2011):

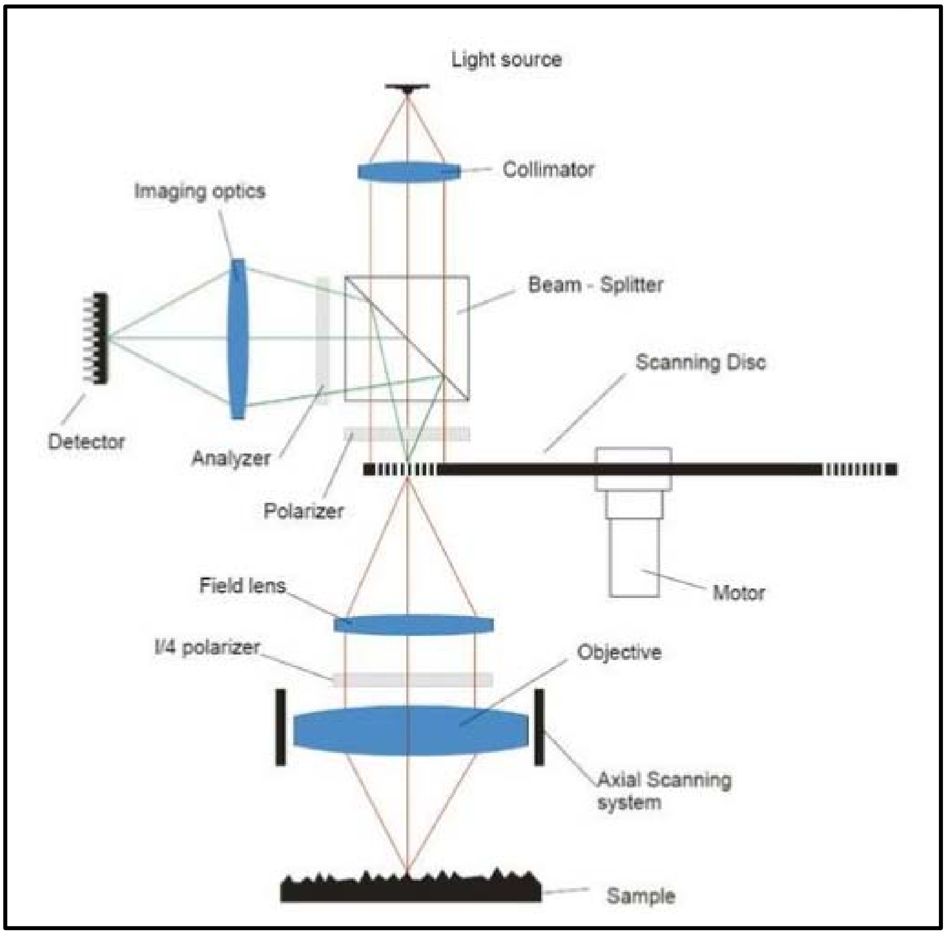


Figure 2-4: Typical schematic diagram of a disk scanning confocal microscope

2.5.3 Photometric stereo

Photometric stereo is not a true areal technique, in that it does not acquire information on the x, y and z position of each measurement point. Instead, it uses variations of lighting across a surface to infer the topography of the surface.

A ring light consisting of several light sources (possibly six or eight) is placed around the surface to be measured. An image of the surface is acquired using each light source separately, so that the shadowing of the surface texture changes dependant on the incident angle of the light source. In comparing the differences in lighting across the images, the topography of the surface can be inferred (Sakarya, Leloğlu, & Tunalı, 2008; Vorburger, Song, & Petraco, 2015). A schematic of the principle is pictured below in Figure 2-5 (Pernkopf & O'Leary, 2003):

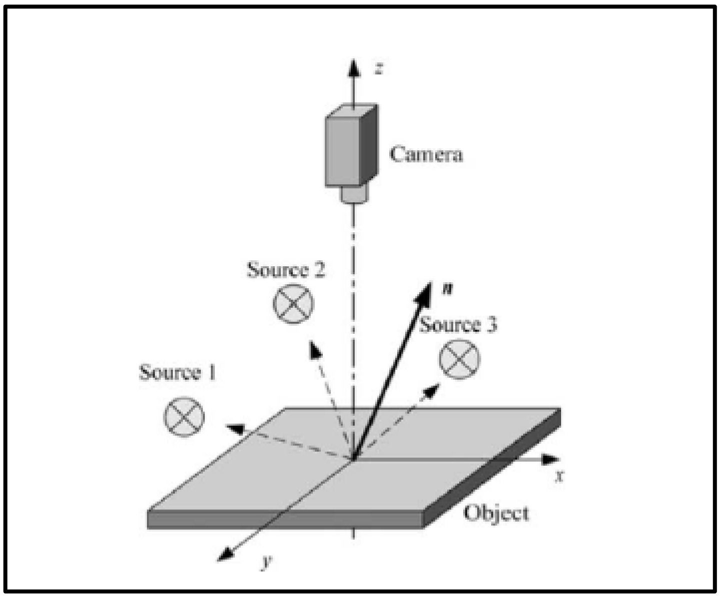


Figure 2-5: Schematic of photometric stereo principle

While photometric stereo is currently employed in some commercially available ballistic identification systems, due to its speed and ability to create good looking images (Cadre Forensics, 2017), it should not be considered as a true topographical technique. As the measurement relies upon lighting and reflectivity conditions, it is considered that both repeatability and traceability has not been proven. Therefore, it should not be considered as a technique able to advance ballistic toolmark evidence according to the Daubert standard. A comparison of the techniques is shown below in Table 2-2.

There are two other measurement systems, coherence interferometry and focus variation, that are also considered for ballistic toolmark identification. They will be described in further detail in Chapter 2.6 as the measurement techniques were used directly for research in the thesis.

Table 2-2: Comparison of measurement technique capabilities

	Contact measurement	Areal data acquisition	Dependant on user settings	Able to measure steep flanks	Measurements repeatable across laboratories	Suitable resolution
Stylus measurement	✓	✓	✓	Dependant on stylus	Yes: with standard operating procedures	✓
Confocal microscopy	✗	✓	✓	✗	Yes: with standard operating procedures	✓
Photometric stereo	✗	✗	✗	Can image steep flanks in 2D	Measurements affected by differences in environmental lighting	N/A

2.5.4 Current developments in ballistic identification systems

The John Jay College of Criminal Justice was involved in a study in which the individuality of Glock 9mm toolmarks were evaluated. During one month in 1997, 617 Glock pistols were test fired and the cartridge cases were evaluated using the IBIS Heritage system. All the cartridge cases were correlated against one another and it was found that there were no misidentifications, i.e. false positives. Over the next five years an additional 1,015 cartridge cases from different Glock pistols were also evaluated, using traditional comparison microscopy. It was also found using comparison microscopy that there would be no misidentifications, and therefore it was shown that Glock pistols could be differentiated.

The college then went on to use more advanced methods to create an automatic identification system. Using confocal microscopy, the breechface toolmarks on cartridge cases bases were measured, and the datasets were filtered so that only the salient components of the surface remained.

Principal Component Analysis (PCA) was then used to reduce the dataset into groups of derived variables, and it was found that PCA was able to differentiate toolmarks imparted by different firearms. The PCA could then be applied to machine learning such that the correlation of toolmarks would be automated. It was found in preliminary studies that the error rate for the automated system is 0%, meaning there would be no false positives identified (Gambino, et al., 2011).

The National Institute of Standards and Technology (NIST) in the USA have published on the advancement of the measurement and identification of ballistic toolmark evidence since the early 2000s. During this time, they manufactured standard reference materials of both bullets and cartridge cases, which have been used in the following studies.

In 2009 (Chu, et al., 2010), NIST collaborated with the ATF, to determine the traceability of identification. NIST manufactured 2,460 bullets and 2,461 cartridge cases to be used as standard reference material (SRM) in ballistic toolmark identification. The SRM bullets were manufactured from reference profiles acquired from a fired bullet using stylus measurement techniques, which were used as tool paths to control the machining of the striations onto blank bullets. This results in both physical SRM and virtual profiles measurements to be used for comparative methods. The cartridge case SRM was manufactured by using electroforming of a fired cartridge case to create a negative, which was then electroformed onto blank plates to result in a replication of the original toolmarks.

Using both stylus instruments and confocal microscopy, measurements of the SRM were acquired and correlated to the master using the cross-correlation function (to be described in Chapter 2.8). It was found that there was a high level of correlation- over 90% match- between both the different SRM samples and the two measurement methods used, therefore the system could be used for the validation of ballistic imaging systems. Using the SRM samples it was also found that describing the correlation of toolmarks using the cross-correlation function resulted in accurate methodology, as using surface texture parameters alone was misleading due to the fact that significantly different surfaces may have similar surface parameters (Song J. , et al., 2009; Vorburger T. V., et al., 2011)

Using the SRM material, NIST were able to investigate methods that would allow for the automation of current manual techniques. In one study (Chu, Song, Vorburger, & Ballou, 2010), the effective correlation area for a bullet striation was investigated. It was reported that the width of a Land Engraved Area (LEA), if calculated accurately, could provide a class characteristic that is able to distinguish different firearm manufacturers. A system was then presented that was able to select the effective correlation of a bullet, i.e. separate areas containing striations, calculate the twist angle of the rifling and extract the average profile across the area.

Further studies also showed the ability to predict whether a bullet could be used for effective correlation based on the quality of striations imparted on the bullet. Firstly, bullet measurements were acquired using confocal microscopy and any outliers and dropouts were removed from the dataset. A Gaussian filter was applied with a long cut-off value of 0.25mm and short cut-off value of 0.0078mm, and a top hat transform was applied to suppress data that did not relate to the striations on the bullets. Canny edge detection was then used, which binarises images based on light and dark areas. Dark pixels were designated as 0, and light pixels as 1, with 1 values being considered as edge elements. Based on the vertical length of neighbouring pixels with the value 1, the edge of a striation can be automatically detected, while disregarding shorter edges that have arisen through smaller scale toolmarks. Using a ratio of edge element pixels to the total number of pixels in a given area, the density of striations can be calculated, thus giving an indication of whether there is enough toolmark information on a surface to allow effective correlation (Chu, et al., 2010).

Following on from this, NIST then revisited the Consecutive Matching Striae (CMS) method first detailed by Biasotti (Biasotti, 1959), who described the ability to match striations based on how many consecutively (without break or dissimilarities between) matching striae were found between two bullet surfaces. He concluded that either one set of six and above CMS or two sets of three and above CMS when comparing 2D images of bullets would prove that the bullets were fired from the same gun.

While this technique has been generally accepted, it is still a subjective technique in that the number of CMS is based on the opinion of the examiner. Therefore, NIST conducted a study to determine the efficacy of an automated CMS system. Using an averaged 2D profile of a bullet surface, striae were extracted based on the position, width and height values within the overall profile. In a preliminary study, 10 matching bullets- and therefore 60 land impressions- were matched using the automated system, and 29 pairs were considered a match based on the 6 CMS method. In a later blind study, 29 of 30 matching pairs were identified with no false positives (Chu, Thompson, Song, & Vorburger, 2013).

In an effort to understand the validity of current identification systems being used, the National Ballistic Imaging Comparison (NBIC) project was developed by both NIST and ATF, to evaluate the traceability of the measurement techniques currently used in ballistic investigation by using standard references to establish calibrations and measurement uncertainty. SRM material was sent to the ATF, so that examiners would be able to acquire measurements using the modern IBIS systems. A total of 13 laboratories took part in the study with 13 examiners involved.

There were three phases in the study: phase one included 10 repeat measurements over one/two days by one examiner using the same system settings, phase two included four weekly tests by one examiner using different settings and phase three extended this to 12 monthly tests.

Within this study, it was found that quality problems arose as a result of IBIS examiners incorrectly setting correlation areas for measurement, thus ultimately changing lighting settings used to acquire the images, and effecting correlation between laboratories. It was also found that there were some discrepancies and bugs in the IBIS correlation system, where stitching of images to create one larger image per bullet can result in disjointed images through issues with stage translation (Song J. , et al., 2012).

Current research at NIST has focused on creating efficient methods of ROI detection and pre-processing of surfaces to ensure only salient information is being used for correlation. Congruent Matching Cells (CMC) were developed at NIST in 2015 to enable differentiation between salient and non-salient areas of the cartridge case surface (Song, 2015). The cells are essentially a grid which is overlaid onto a cartridge case dataset. The cell size of this grid must be small enough that a mosaic of cells can separate valid and invalid correlation areas, but large enough that there is sufficient data in each cell. Using a rule of 6 congruent matching cells for identification between two datasets, validation test of the method suggests sufficiently low error rates, with combined probability of false positive/false negatives rate quoted at 1×10^{-6}, to be considered as a forensic technique. These findings are currently being used to implement a new identification system, known as the NIST Ballistic Identification System.

A review paper published by both NIST and the John Jay College of Criminal Justice discussed the instruments and pre-processing methods that were directly applicable to ballistic toolmark identification (Vorburger, Song, & Petraco, 2015). Measurement methods such as confocal microscopy, coherence scanning interferometry and focus variation were discussed. Pre-processing methods discussed included Gaussian filtration, to remove short wavelength noise components and long wavelength form components from the surface, and the use of surface scale decomposition. Using wavelet transforms, it was found that the band pass properties of the fourth order Coiflet transform, the identifying striation patterns would be successfully decomposed from other surface texture.

Research currently being conducted at NIST shows a definite trend towards the automation of ballistic identification, and so far, it has been proven that not only will such systems minimise subjectivity, but will also result in a more accurate identification system. However, the advanced measurement techniques available are not being fully taken advantage of. While NIST have proven the capabilities of acquiring areal datasets, for bullets an average profile is being used for further analysis. This may lead into issues with data being skewed by damaged/noisy areas on the surface, and an adverse effect on any further correlation.

Bolton-King et al. (Bolton-King, et al., 2010) published a concise study on the efficacy of various measurement systems to acquire toolmark topography. Using a standard reference bullet manufactured at NIST, measurements were acquired using various systems and the results were compared.

Firstly, it was found that a measurement system must have at least a vertical resolution of 0.1µm and a lateral resolution of 1µm, to ensure the fidelity of the dataset even if it should be subsampled for further analysis.

Four measurement systems were compared, interferometry, laser profilometry, confocal microscopy and focus variation.

Laser profilometry uses a focus following principle and in operation similar to contact profilometry, with the stylus replaced by a laser beam, was found to have insufficient lateral resolution to acquire the quality of dataset needed in further analysis. This was due to surface texture changing the incident angle of the beam, thus expanding the beam footprint and skewing the measurement.

White light interferometry, which relies upon the interference patterns of two light sources to determine the surface roughness, was also found as an unsuitable technique. Firstly, the initial focusing on the interference pattern on the surface proved difficult due to the rougher

surfaces dispersing the light of the fringes. Also, due to a small working distance between the surface and the lens, it was found that it was not possible to measure deeper toolmarks such as the bottom of the firing pin impression. Most importantly, standard white light interferometry is highly slope limited, thus resulting in excessive data dropout. Additionally, phase ambiguity caused by rapid slope change can induce optical spikes.
Confocal microscopy, which is purported to be used currently in the newer IBIS systems, was found to acquire datasets of a high enough quality to be used for further analysis. However, due to slope limitation issues it was unable to measure steep flanks such as those within a firing pin impression, which leads to dropout of data.
It was found that the focus variation technique offered the best solution for measurement, as it has generally larger working distances than confocal microscopy and the ability to measure step flank angles means all topography of the toolmarks can be acquired. Focus variation techniques are more time expensive than confocal microscopy, however the ability to measure slopes of up to 85° based on micro reflectance is a clear advantage.
In other studies conducted by Bolton-King, advanced analysis techniques were applied to 2D imaging to determine whether the transitional areas between the lands and grooves of a barrel can be considered a class characteristic (Bolton-King, 2012; Bolton-King, et al., 2012).
The barrels of firearms were replicated, and the replicas were imaged using microscopy with image stacking to ensure full focus of the surface by extending the focus depth.
The images were then converted to binary and a 2D fast Fourier transform was applied to gain a spectral image that defines the frequencies of the image. Using principal component analysis to separate variables in the pattern of the transitions, it was found that there was a differentiation between the transitional areas between separate manufacturers, which could further be differentiated by using a weighted Euclidian distance between the mean score of an image to an image of the same manufacturer, and the mean score of an image to an image of a separate manufacturer.
The study proved that the transitional area between the land and groove did change between different manufacturers, and thus can be considered as a class characteristic.

2.6 Measurement techniques to be used in present study

The following section will detail the advanced measurement systems used to acquire datasets for studies in this thesis. The techniques were chosen as a comparison between general purpose commercial systems and those that are specific to the measurement of ballistic toolmark evidence.

The Advanced Ballistic Analysis System (Alias) was chosen due to an existing relationship with Forensic Pathways Ltd. Forensic Pathways were interested in research relating to ballistic toolmark investigation. As such, Forensic Pathways Ltd. were instrumental in securing a ballistic imaging system on loan to the University of Huddersfield during the course of the research. The author also had unlimited access to an Alicona G4 instrument which is owned by the University of Huddersfield, and therefore it was decided due to availability and time constraints, (i.e. the time needed to measure the entirety of the Odyssey collection on each instrument)) the research should focus on two measurement systems.

There is a significant difference in cost between the two systems, with the general-purpose instrument being cheaper. With the added ability to measure other evidence types within the forensic laboratory, it may be considered the better choice. However, the data fidelity must be considered. Should one measurement system offer better acquisitions with less outlier and dropouts, then it should be considered as the better choice, as from a forensic point of view such evidence will be more reliable and repeatable and therefore admissible in a court of law.

2.6.1 Advanced Ballistic Analysis System (Alias)

The Advanced Ballistic Analysis System (Alias) was developed by Pyramidal Technologies, Ottawa, as an alternative system to traditional comparison microscopy systems. It is a system dedicated to the measurement and comparison of ballistic toolmark evidence (Barrett, Tajbakhsh, & Warren, 2011).

The system software allows each user to add a cartridge case or bullet to the system, by firstly entering information on the evidence, as seen in Figure 2-6. This includes the caliber, material and manufacturer of cartridge cases and bullets, and includes the number of rifling marks and twist direction in the case of bullet evidence. Optional extra information input includes the geographical area from which the evidence was recovered, the staff involved in the case and any other information i.e. connections to certain suspects.

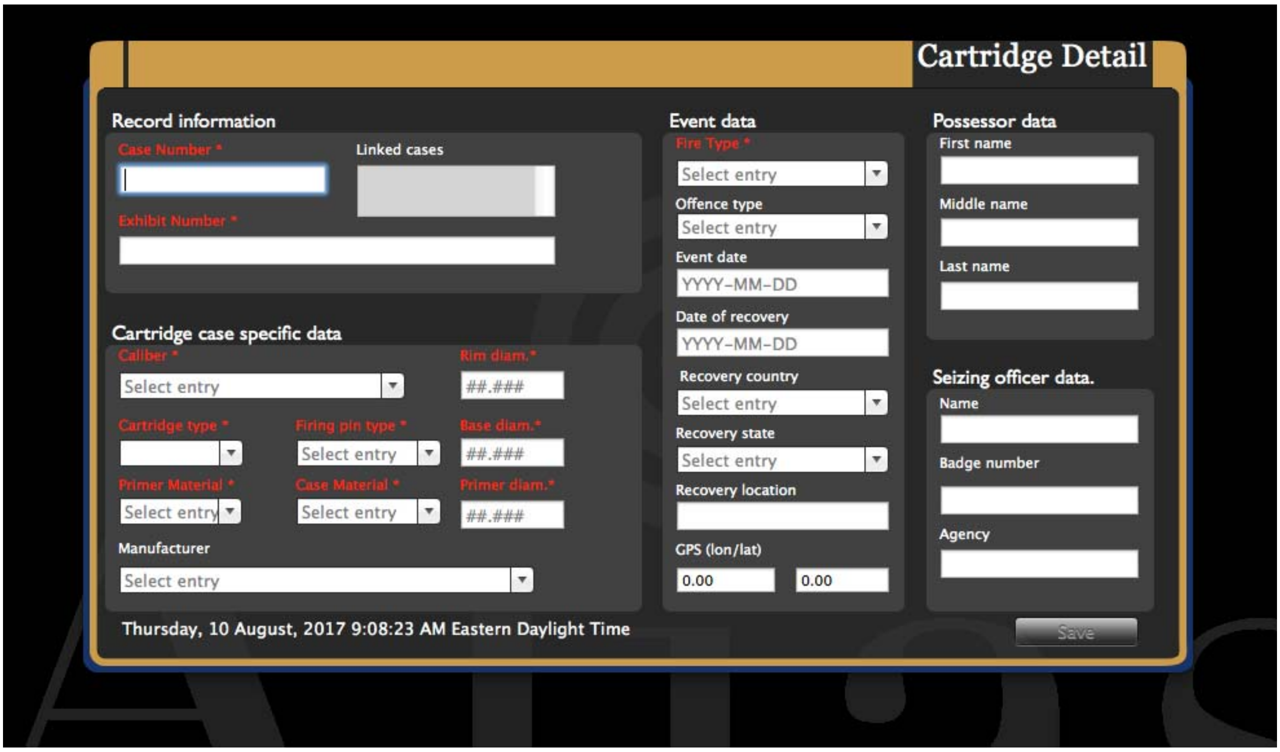


Figure 2-6: Example of information input screen in Alias software

The Alias system allows the export of data in various file formats, meaning datasets can be imported into or from other commercially available software for analysis, or into a ballistic imaging system for comparison. The open software architecture of the system also allows for the importation of comparison algorithms designed by the user, meaning the software can be changed to suit the individual preferences of the end user.

The visualisation of datasets in the Alias system can be manipulated should the need for visual comparison arise. Evidence can be viewed side by side as in traditional comparison techniques, with added manipulation of lighting settings and z scale to visually highlight toolmarks on the surface (Barrett, Tajbakhsh, & Warren, 2011).

The data acquisition is achieved using a Linnink parallel Optical Coherence Tomography (pOCT). POCT relies upon the interference patterns created between light reflected from a reference mirror and light reflected from the surface (Figure 2-7) (de Groot, 2011). When the optical paths of the reference light and surface light are the same (known as the coherence plane), interference patterns are formed on the sensor and the height of the point on the surface can be recorded. If either path is moved more than the coherence length of the light source used, the interference patterns will disappear, and no height information is recorded.

The Alias tomograph uses a low coherence super luminescent diode (SLD) as the light source, with a centre wavelength of 800nm and Full Width at Half Maximum (FWHM) of 40nm. An objective lens is placed in front of the SLD to focus the beam, which is then split with one beam going to the reference mirror and one to the sample surface.

The beams are reflected from each surface and combine, pass through the camera objective lens, and the resulting interference patterns are imaged on a camera. The camera measures the optical intensity of the light, which vary dependant on the difference in path lengths of the reference beam and surface beam. When the paths are the same, this is known as the coherence plane, and the interference intensity can be used to measure the height of the point of the surface. Moving the reference mirror allows the path length to change, thus the surface height can be measured over a longer range. Stepper motors enable the instrument to scan in both the x and y axes, thus the entire topography of the toolmark can be acquired by stitching images.
The height of each point is measured in relation to neighbouring points, and the topography of the surface is built up as the instrument moves in the x, y and z axes (Lambelet, 2011; Barrett, Tajbakhsh, & Warren, 2011; Vorburger, Song, & Petraco, 2015; Leach, 2010) .

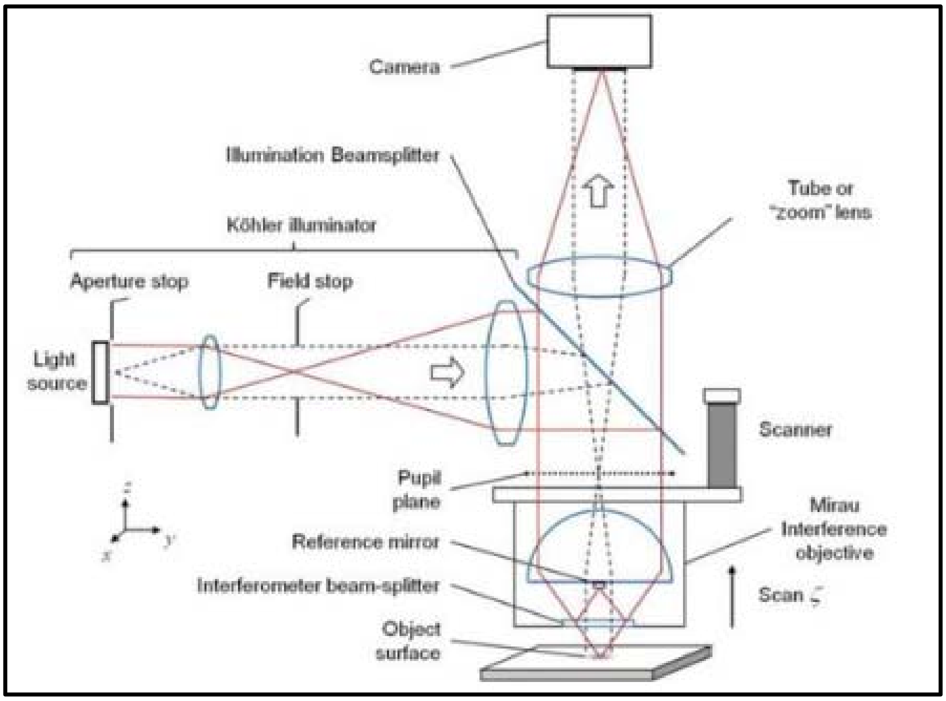


Figure 2-7: Principle of pOCT

The intensity data for each pixel is recorded in successive frames on the camera, building height point data as the optics move in the z axis and the surface moves in the x and y axes. This effectively offers an autofocus at every point in the field of view, as only the highest intensity of interference will be recorded at each data point P (de Groot, 2011).
POCT interferometry offers advantages over other optical measurement systems due to its ability to measure complex surfaces. As both an aperture stop and field stop are used within the interferometer, the numerical aperture and illuminated surface area can be controlled. This effectively reduces scattered light and reflectance issues, and therefore minimises 2π

errors, in which interference fringe order had been misclassified due to light dispersion (de Groot, 2011; Leach, Brown, Jiang, & Blunt, 2008).

2.6.2 Alias correlation algorithms

Through discussions with the staff of Pyramidal Technologies, some information on the current correlation algorithms used in the Alias software was obtained, however this information is not exhaustive as it is subject to commercial confidence.

Firstly, the database of potential matches is minimised by using the information on class characteristics input by the user at the time of acquisition. With regards to cartridge cases this means that correlation will only occur within the same caliber and type of firing pin impression (rimfire/ centrefire). Bullet correlation is based upon the caliber, number of rifling twists and direction of the twists.

Cartridge case correlation firstly relies upon the automated differentiation of firing pin impressions and breech face marks, in both centrefire and rimfire cartridges. This is achieved in the Alias software by calculating the height of each data point and assigning a cut off value when the height has exceeded that expected in the firing pin impression. This is originated by using an automatically generated seed (one data point) as a starting point and growing the area of neighbouring points until the height cut off value is found.

Once the breech face and firing pin toolmarks have been separated, each are correlated using the Alias Alpine/Dead Sea comparator. The algorithm distinguishes a fixed number of the highest and lowest points on a surface, and where they appear within the coordinate system.

These points are then compared across the surfaces, which results in a list of best matches with the lowest score being the best match.

The correlation of bullet toolmarks is firstly dependant on the correct differentiation of engraved areas on the bullet. In the Alias software, this is achieved by determining where the largest step height difference is between neighbouring points, once again acquired by using a seed and growing the seed based on a change in height between neighbouring points. The lower surface between two step heights corresponds to the engraved areas and is extracted.

The extracted area is then decomposed into the separate frequencies using Zernicke moments present on the surface, in which the surface can be described as a finite number of polynomials, with lower frequencies corresponding to waviness and form of the surface and higher frequencies corresponding to the rougher surface texture. Once the surface is decomposed, frequencies corresponding to the topography of the toolmark are compared to one another, by overlaying one surface over the other and comparing differences between the height points. A hitlist of best matches is prepared with the lowest score being the better match.

2.6.3 Focus Variation

The Alicona G4 focus variation measurement system will be used in the following studies to determine the efficacy of using a general purpose advanced measurement system for the acquisition of ballistic toolmark evidence, where correlation is facilitated by third party algorithms.

The Alicona combines the use of objectives with a small depth of focus and the ability to scan the objectives vertically to acquire areal datasets of a surface. This can further be combined with a fourth rotational axis to acquire true areal measurements (data acquired on information of object in space along with height point of surface).

As pictured in Figure 2-8 (Vorburger, Song, & Petraco, 2015), white light is focused on the surface through the objective, and reflected light is collected onto a sensor. As the objective is scanned vertically the variation of focus for each point on the surface is analysed, and when the point is considered in focus the information is used to build up the topography of the surface. The focus of each point relies upon the micro reflection of peaks and valleys upon the surface. Such contrast is quantified by calculating the standard deviation of the light intensities compared to neighbouring points, where the focus is determined by the largest standard deviation possible. Therefore, it is not always possible to measure smooth surfaces using Alicona as there is not enough contrast between neighbouring points (Leach, 2010; Vorburger, Song, & Petraco, 2015).

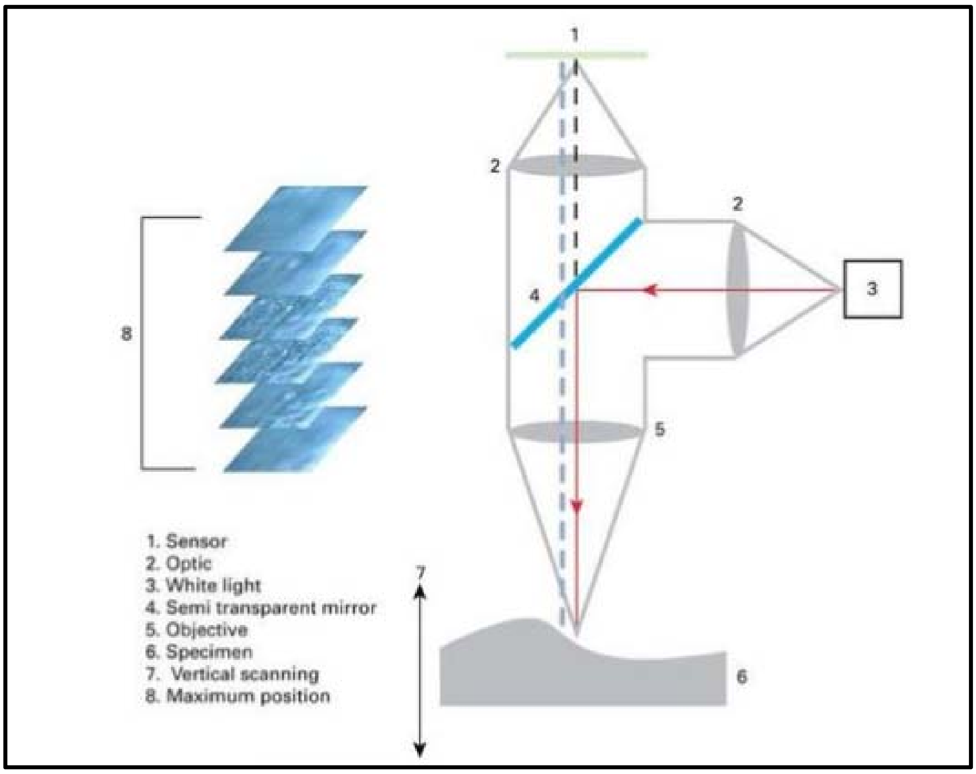


Figure 2-8: Schematic diagram of focus variation

The use of a modular light system enables the measurement of various materials, as light intensities can be changed dependant on the reflective properties of the sample. In highly reflective materials such as metallic surfaces, the light can be polarised to ensure the high levels of reflectivity does not result in optical spikes appearing on the surface. Incident light can be generated either coaxially or through a ring light, which has the added benefit of overcoming the numerical aperture limitation of the objective and therefore making it possible to acquire data from surface angles of over 80°. Slope limitation in optical measurement techniques is caused by the use of coaxial light and lenses with a low numerical aperture. With a low numerical aperture, the angle of reflectance of light that can be captured is low. Therefore, if the surface is angled, reflected light may be at a higher angle than that at which the numerical aperture can accept. The use of a ring light means the reflectance angles of the light can be shifted so that light is captured past the limits of the numerical aperture, thus overcoming issues using smaller numerical apertures and capturing light at larger angles. The vertical resolution and scan range depends on the objective chosen by the user, but the system is capable, although not used for studies contained within this thesis, of vertical resolutions up to 10nm. The ability to move the surface in both the x and y axes also means that areas of up to 100mmx100mm can be acquired (Leach, 2010; Vorburger, Song, & Petraco, 2015; Danzl, Helmli, & Scherer, 2011; Hiersemenzel, Petzing, Leach, Helmli, & Singh, 2012).

Due to the ability of the user to change measurement settings of the instrument, the technique could be considered as having an objective aspect. Using different lighting conditions, objectives and resolutions will all have a direct effect on the quality of the measurement. Therefore, to ensure that this technique could be used in a forensic context, there should be a standard operating procedure in place to ensure the same settings are used for each measurement and across all forensic laboratories, thus making the technique reliable and repeatable (Hiersemenzel, Petzing, Leach, Helmli, & Singh, 2012).

While there have been a number of studies in which the Alicona has been used as a measurement technique in various disciplines (Bello & Soligo, 2008; Ender & Mehl, 2013; Mahat, Aris, Jais, Yahya, & Ramli, 2011; Schroettner, Schmied, & Scherer, 2006), currently there has been little published on the measurement of ballistic toolmark evidence. The one published study on the use of the Alicona to measure ballistic toolmark evidence found that through the measurement of 50 .177 caliber air pistol pellets the measurement system was capable of producing repeatable results. Rachel Bolton-King also successfully used the Alicona to define transitional areas between LEA and GEA in the barrel. (Bolton-King, 2012).

As the Alicona is a general-purpose instrument the software does not include any correlation algorithms that are fit for purpose for this study, therefore in this study measurements will be exported into other software packages, including MATLAB and Surfstand for further analysis.

A comparison of the measurement techniques detailed above is shown in Table 2-3 below:

Table 2-3: Comparison of Alias and Alicona measurement techniques

	Contact measurement	Areal data acquisition	Dependant on user settings	Able to measure steep flanks	Measurements repeatable across laboratories	Suitable resolution (Bolton-King, et al., 2010)
Alias	✗	✓	✗	Can measure steeper flanks than traditional interferometry but Alicona can measure steeper angles	✓	✓
Alicona	✗	✓	✓	Yes: >80°	Yes: With standard operating procedures	✓

2.7 Processing and analysis techniques to be used in the study

The following section will detail both the processing techniques used to ensure the datasets include only salient data, and the analysis techniques used for the correlation of evidence.

2.7.1 SURFSTAND software

The SURFSTAND project was a multi partner project which aimed to evaluate surface parameters and develop new surface filtering techniques. Using an increase in practical evidence, problems with original definitions of areal surface parameters were solved, which results in a decrease in ambiguity in parameter definitions. The initial 'Birmingham 14' areal surface parameters, extended from existing 2D parameters, lacked practical evidence of their applicability to real life scenarios, and thus SURFSTAND was created to demonstrate this (Blunt & Xiang, 2003).

This resulted in a software package in which measured datasets in various file formats can be imported and processed. With regards to experiments completed within this thesis, SURFSTAND was used to be able to crop datasets and use filtration techniques to separate salient information contained in the original dataset.

2.7.2 Surface processing- filtering of salient data

Within a dataset of ballistic toolmarks, all scales of information of the surface topography will be acquired. This will include the overall form of the surface, for example the curvature of a bullet, transitional areas between land and groove engraved areas, and the surface roughness in which individual characteristics will be found. Lower frequency wavelengths of surface information relate to the overall shape/form and waviness of the surface, while higher frequency wavelengths relate to the roughness and smaller scale topography on the surface (the International Organisation for Standardisation, 2012).

As the individual characteristics of toolmarks are the only characteristics which can differentiate between firearms, it is only this information that is salient in further correlation. Using other information from sub class or class characteristics will result in correlations scoring higher than expected, and therefore potentially introduce false positives, as these characteristics are shared in more than one firearm. Using optical techniques may result in optical (high frequency) spikes or noise occurring on the surface, which must be removed before correlation as they may skew the results. Applying a low pass filter to a surface will remove lower frequency data from the measurement while keeping frequencies higher than the wavelength designated by the filter. Applying a high pass filter will remove frequencies higher than those defined by the filter. Therefore, a band pass approach should be applied to ballistic surface data, to leave only salient frequencies. Testing of filtering methods can be seen in Appendices 3 and 4.

Datasets must be processed before correlation to remove extraneous data. The methods of processing used in this study are as follows:

2.7.3 Gaussian filtering

Gaussian filtration is the traditional method of extracting surface features from a measurement for further analysis and has been defined as the standard technique in standards ISO 25178. Areal filtration techniques are directly extrapolated from the 2D profile filtration technique, and as such it is considered a robust technique for filtration. (the International Organisation for Standardisation, 2012; Zeng, Jiang, & Scott, 2010).

The transmission properties of a filter, i.e. the resultant filtered surface, is determined by its weighting function. In Gaussian filtration, the weighting function is in the form of a Gaussian probability function, which averages a height point dependant on the height information of its neighbouring points. The resulting filtered surface is a convolution of the originally measured surface and the weighting function.

The cut off wavelength used in Gaussian filtering does not completely disregard wavelengths outside of the set frequency. Instead the cut off value is the frequency at which data will be attenuated at 50% of the original surface. The attenuation from the cut off value is steadily decreased to 0% for frequencies below the cut-off value and transmission is steadily increased up to 100% above the cut off value for low pass filtration, and vice versa for high pass filtration (Digital Surf, 2017).

Areal Gaussian filtration uses two cut off values, one for the x direction and one for the y direction. As averaging of each point depends on its neighbouring points, it is possible for averaging to be skewed by 'freak' values. Freak values are high peaks or low valleys on the surface, and when they are used to average a neighbouring point, the resulting average will become higher or lower than expected than in a neighbouring point with less height difference. As such, a robust Gaussian filter is used which adds an additional weighting function on the height of each point. Freak values will have a proportionately lower weighting, and thus will not disturb the average (Blunt & Xiang, 2003).

It is also possible to use Gaussian filtration in the frequency domain. In such a case, an areal surface can be converted into a frequency domain using the Fast Fourier Transform, and the frequencies of the surface are plotted against the amplitude. Gaussian filtration can then be used to remove lower or higher frequencies in the plot before the FFT is reversed, thus resulting in a filtered surface. Using the FFT approach is computationally more efficient, and therefore is usually preferred in filtration software.

In the present study, robust Gaussian filtration will be used as both a high pass filter and a low pass filter. For most surfaces, the cut-off values are determined based on ISO standards which outline a set of standard wavelength frequencies. However, the ISO standardised filtration techniques do not necessarily separate salient and non-salient data in toolmark datasets. In the terms of ballistic toolmark evidence, applying standard filtering may result in some of the salient data regarding individual characteristics being removed, or surface texture relating to subclass and class characteristics being retained in the filtered surface and skewing further correlation. As such, it is important to determine the correct cut-off values to ensure only salient data is retained in the filtered surface (see Chapters 4 and 5 and Appendix 3).

2.7.4 Multi scale decomposition using wavelets

Wavelet decomposition varies from Gaussian filtering, as instead of removing certain wavelengths from a surface to create one filtered surface, the surface is decomposed into banks of different surfaces that contain defined frequency bands. In ballistic toolmark evidence, such a technique may be able to separate individual characteristics from other surface features, as individual characteristics are contained in a higher frequency bands than other features.

Using a wavelet transform, the original measured surface can be described with varying resolutions. The surface is compared to a series of wavelet functions to determine the various frequencies contained within the surface (Wang, et al., 2017).

A wavelet ($\Psi(T)$), has an integral of 0, and has support in both the space and frequency domains. As it is compact, a wavelet exists only for a finite amount of time. A surface can be considered as being made up of multiple wavelet signals in various frequency bands and is known as the mother wavelet. The function of the wavelet must satisfy (Scott, Zeng, & Xiang, 2011) Equation 2:

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{\left|\widehat{\Psi}(\omega)^2\right|}{|\omega|}\, d\omega < \infty$$

(Equation 2)

Should the wavelet satisfy the admissibility condition described above, then the inverse wavelet transform can exist. The inverse transform function, in which the mother wavelet is broken down, results in several versions of the mother wavelet that have been rescaled and shifted with changes in the scaling coefficient a, and the translation coefficient, b (Capri, 2011). The transform is described in Equation 3:

$$Wa,b\,(f) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \Psi \left(\frac{t-b}{a}\right) dt$$

(Equation 3)

Where:

f(t) is the signal input, a is the translation parameter that covers various frequency ranges, and b defines the location of signal events.

a is a scaling coefficient, which determines the size of the wavelet, as can been seen in Figure 2-9 (Zahouani, Mezghani, Vargiolu, & Dursapt, 2008). Dilation and constriction of the wavelet results in decomposition of different scale information on the surface.

b is the translation coefficient, which allows movement of the wavelet across the surface.

In current wavelet decomposition of surfaces, the lifting wavelet transform is used, as is it can be applied to both regular and irregular datasets. Using the lifting scheme, the surface is first split into even and odds sets using a lazy wavelet. The lazy wavelet is then lifted to produce filter banks, in which the desired wavelet scaling functions are created. This results in the construction of filter banks within the spatial domain (Abdul-Rahman, Xiang, & Scott, 2013).

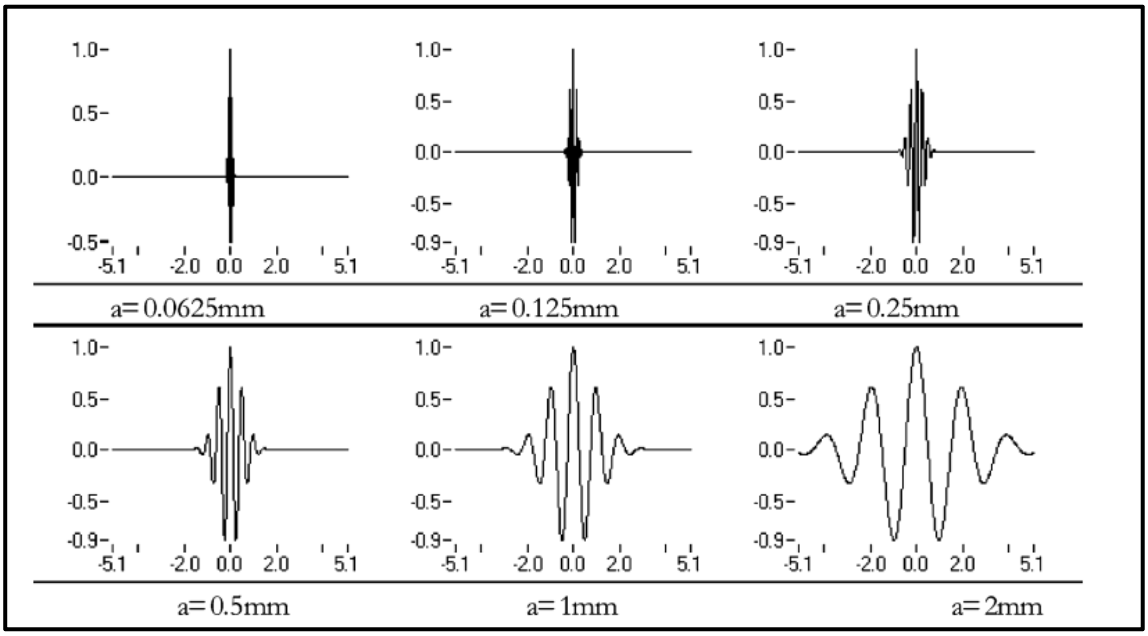


Figure 2-9: Using the scaling coefficient to resize the scale of the wavelet

The wavelet function is able to divide the original signal, or mother wavelet, into separate locations and frequency bands. Using a high pass filter and a low pass filter the original surface is progressively decomposed, and the resultant surface is rescaled. The resultant surfaces are denoted as follows in Equation 4, where S is the original surface, A is a course approximation of the surface and D surfaces are the detailed surfaces in varying frequency bands, as shown in Figure 2-10 (Capri, 2011; Zou, Li, Kaestner, & Reithmeier, 2016):

$$S = a_n + d_1 + d_2 + d_3 + \cdots$$

(Equation 4)

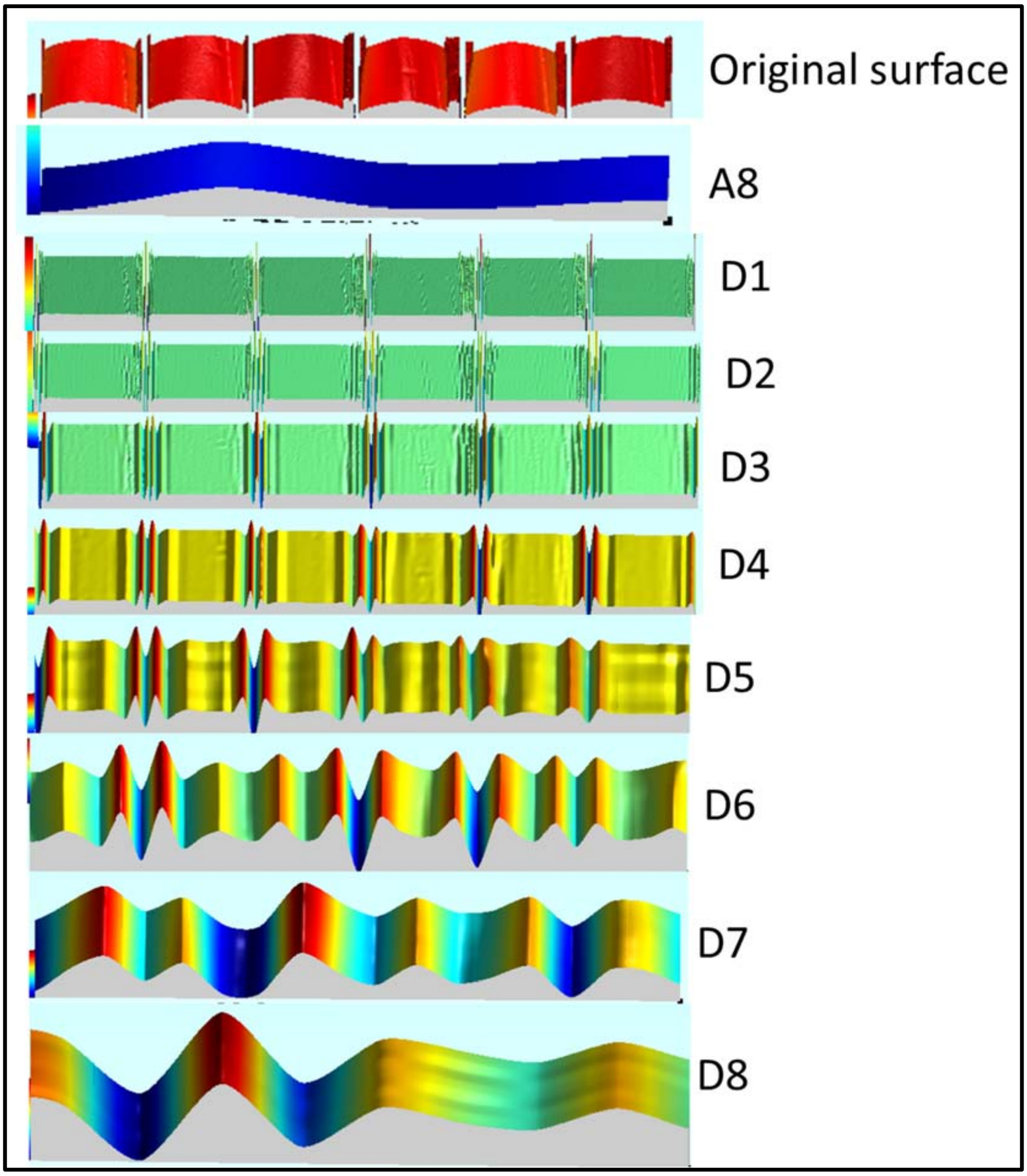


Figure 2-10: Wavelet decomposition of an unwrapped bullet dataset

In this case, the entire frequency range contained within the surface can be described within eight D levels and one A level band.

Wavelet decomposition of surfaces delivers the potential to implement correlation of defined frequency banks, which contain only the spatial properties relating to ballistic toolmark information.

2.8 Correlation algorithms

Cross-correlation is measuring the similarity of two series, as a function of the lag between the two series. The lag between series can represent time, shift distances and degrees of rotation amongst other variables.

The use of the cross-correlation function has been a long-standing method in signal processing, to be able to quantify the relationship between two signals in the time domain. Cross-correlation is calculated by shifting one signal by τ with relation to the other signal to estimate a time average product (Fahy, 2001).

While cross-correlation is traditionally used in signal processing, it can also be applied to surface texture to determine the degree of similarity between two surfaces. The cross-correlation function can be applied in any number of dimensions; thus it can be used for 2D profile correlation of profiles extracted from surface data, or as an areal correlation using the height points across two surfaces. The two profiles are aligned, and where there are points on one profile that align vertically to points on the other profile, the values for the heights of the profile are multiplied together. Each set of aligned points are multiplied together, and the sum of these values corresponds to the cross-correlation.

One profile is the shifted in the x axis, and this represents the lag between the two profiles. The cross-correlation is calculated at varying shift distances to determine the maximum cross-correlation function with regards to the shift distance between two profiles.

In such cases the coefficient is a measurement of the relationship between the size and directions of variable height on a surface. When correlation results in a positive integer, it shows a positive correlation between the two surfaces. A negative integer infers a negative correlation, and a value of 0 shows no correlation between the two surfaces (Lyon, 2010).

The first published application of the cross-correlation function (CCF) to ballistic toolmark evidence was by NIST. As bullet signatures can be considered random profiles, the correlation between such profiles will decrease as one is shifted against the other. Due to this, cross-correlation is advantageous in ballistic toolmarks as correlation will only reach a maximum when toolmarks correspond (Song, Ma, Whitenton, & Vorburger, 2005).

In 2D profiles, used by NIST to correlate extracted 2D profiles of bullet striations, CCF can be described as shown in Equation 5:

$$CCF\ (A, B, \tau) = \lim_{L\to\infty} \left(\frac{1}{L} \int_{-L/2}^{L/2} Z_A(x) Z_B(x+\tau) dx \right) / [Rq(A)\ Rq(B)]$$

(Equation 5)

Where:

τ is shift distance of one profile against another, Rq is root mean square roughness of the profile and Z_A and Z_B are the two profiles used in correlation.

In using the root mean square of roughness across the two surfaces, the CCF becomes insensitive to scale differences. Therefore, if two profiles have the same pattern but different amplitudes, the CCF will still give an indication of high correlation between the two profiles, even when there is large scale difference.

Therefore, NIST described a secondary function, D_s, known as the difference function, to be able to differentiate between scale differences, which is described in Equation 6:

$$Ds\ (A,B) = \frac{Rq^2(A-B)}{Rq^2\ (A)}$$

(Equation 6)

Where A and B represent the two profiles (Vorburger T. V., et al., 2011).

2.8.1 Areal cross-correlation

It is also possible to correlate the entire areal topography of a toolmark using the height data of a dataset. In this case, rather than using 2D profiles extracted from the dataset, the vertical heights of a series of x, y data points common to each dataset are compared.
As in the cross-correlation of 2D profiles, one dataset is shifted with regards to the other multiple times and the cross-correlation is calculated at each shift until the maximum value is found.
For the cross-correlation of areal datasets (ACCF), the equation is described as follows in Equation 7:

$$ACCF_{max} = \frac{\sum x,y\ (Z_A(x,y) - \bar{Z}_A)(Z_B(x,y) - \bar{Z}_B)}{\sqrt{\sum x,y(Z_A(x,y) - \bar{Z}_A)^2}\ \sqrt{\sum x,y\ (Z_B(x,y) - \bar{Z}_B)^2}}$$

(Equation 7)

In which:
Z_A and Z_B describe the two datasets being compared, $Z_A(x,y)$ describes the height data for points of dataset Z_A, $Z_B(x,y)$ describes the height data for points of dataset Z_B[3]
The $ACCF_{max}$ value is a percentage, where 100% indicates two identical surfaces.
In correlation of 2D profiles, the percentage match would be in terms of the lag, or shift distance, between the profiles. This is not the case in correlation of areal datasets, where instead one surface is shifted point by point with regards to the other surface.

2.8.2 Confirming the cross-correlation of areal datasets with areal D_s

The cross-correlation of datasets does not take into account any difference in scaling factors. Therefore, while CCF may indicate a high similarity, actual similarity may be lower when scaling factors have been taken into account (Song J. , et al., 2009).

[3] In this case, two points x and y from the dataset are being compared. It is possible to correlate between any number of points of the dataset by adding the data to the equation. I.e. $Z_B(a \dots n)$. The algorithm used for this research correlates against every point in a specified region of interest.

To combat this, a complimentary algorithm is used to differentiate between scale differences that may be present between the two datasets. This is known as the relative topography distance, and is described as shown in Equation 8:

$$D_s = \frac{\sum((Z_A(x,y) - \bar{Z}_A) - (Z_B(x,y) - \bar{Z}_B))^2}{\sum x,y\,(Z_A\,(x,y) - \bar{Z}_A)^2}$$

(Equation 8)

A low value of D_s indicates that the scale of the two datasets are similar, whereas a high value will indicate a larger difference. Used in conjunction with the cross-correlation function, a more accurate comparison of similarity between the two datasets can be achieved.

2.8.3 Considerations

While the cross-correlation of two datasets is a valid technique to measure similarity, the following must be taken into account:

- There is no differentiation between salient and non-salient data
- Impressions relating to both individual and class characteristics are correlated
- If used on cartridge cases, datasets must be rotated to be correlated. This is time consuming, and difficult to achieve with accuracy.

The fact that the cross-correlation is unable to differentiate between salient and non-salient data or class/individual characteristics shows the correlation is insensitive, and therefore the calculation is not weighted towards data of salient information.
Therefore, it is expected that data pre-processing and surface filtering is critical before the cross-correlation is calculated.
When correlating cartridge cases, it has been usual in the past to use the circular form of the dataset to perform the correlation. Due to this, one dataset must be rotated with regards to the other to find the maximum cross-correlation. This rotation is more time consuming than shifting the datasets in one axis.

2.9 The Odyssey collection

The Odyssey project was an EU funded project, with which 12 members throughout Europe sought to resolve issues with data interoperability of forensic evidence across forensic laboratories. While some forensic providers within the USA rely upon NIBIN as a means to easily share data across different regions, such a system does not currently exist within Europe.
As there is no information sharing system within Europe, it was found that sharing data between different regions proved difficult, as different laboratories use various systems in

ballistic toolmark identification. Due to these different systems, file formats are often different and therefore cannot be transferred into other systems. S Should there be a need for forensic comparison of one bullet across providers, estimated cost including travel and accommodation for the forensic examiner is around €9,000, however INTERPOL now support a double casting of evidence for shipping which can minimise costs (INTERPOL, 2017).

Project Odyssey aimed to integrate data from across Europe, importing into one information sharing system that would be accessible to all forensic providers. The broad outline of requirements expected of the Odyssey project were to:

- Create a system able to store and exchange information securely
- Enable the automatic identification of potential information links between regions
- Facilitate communication between examiners without the need for costly travel between regions.

Odyssey research resulted in a system compromising of two main databases, local and central. Local databases are maintained by each laboratory, which are then uploaded into the central database. From the central database, it is possible to "mine" the data, by searching using keywords to show any information that may be relevant. This allows an examiner to gain knowledge on evidence in other regions that may be related to their own information. It also possible to create an automatic alert, which will notify the user of possible related evidence based on the keywords describing the evidence inputted by the user. While it was also within the remit of the Odyssey project to create a system that would correlate evidence uploaded from different regions, a solution was not developed (Wilson, Jopek, & Bates, 2010; Yates, Akghar, Bates, Jopek, & Wilson, 2011).

During the Odyssey project, a sample of ballistic evidence was created to be able to compare differences in commercially available systems, with an aim to be able to create interoperability between them. A series of 9mm pistols were used for test firing, as they were identified to be the largest sample size of evidence within most ballistic evidence databases. 390 9mm Luger fired bullets and cartridge cases were entered into two commercial systems, Evofinder and Arsenal as part of another doctoral thesis (Thomas J. , 2011). The 390 cartridges used were of eight manufacturers, to try and minimise effects created by using various materials while creating as large a dataset as possible. Each 9mm pistol was of pristine condition, and as such the sample can be considered as a 'stress test', as less individual characteristics through improper storage and cleaning will be present upon the firearm, thus making correlation more difficult than in most actual forensic cases.

Each of the 390 bullets and 390 cartridge cases were measured and correlated by the two systems, to ascertain whether there is any interoperability, and how the hitlist of best matches compared between the two systems. Due to differences in these hitlist results, i.e.

where the known match was placed within each list, it was found that full interoperability would not be possible. This could be remedied by communication between each system manufacturer, however due to systems containing trade secrets such communication is unlikely.

The full correlation hitlist results will be made available alongside results from studies by the author, for ease of comparison between each system (Thomas J. , 2011).

2.10 Progression from previous publications

The previously published results above show that while there are various methods available for the areal measurement of ballistic toolmark evidence (Bolton-King, et al., 2010; Vorburger, Song, & Petraco, 2015), studies undertaken are not yet taking full advantage of the methods or are not encompassing variabilities in toolmark creation. Most sample sets used in past research have been restrictive, either using small sample sets, samples not encompassing a range of materials/manufacturers, or toolmarks artificially manufactured using a CAD model (Brinck, 2008; Song, 2015; Vorburger T. V., et al., 2011). As such correlation results may appear positive due to minimisation of natural variability in samples used for measurement, it is unclear how efficient methods would be when applied in real-life situations. Advantages in using the Odyssey collection include a comparatively larger sample size, with variables in material and manufacturer of both cartridge and firearm. While it is suspected by the author that this will decrease correlation efficacy compared to published results, measurement and pre-processing methods may be more robust when taking variables in sample into account.

Capability for measurement of the areal surface of bullets has been well established, both in research systems such as confocal microscopy at NIST, and the newer FTI Bullet-Trax system (Brinck, 2008; Hamzah, 2016; Bolton-King, et al., 2010; Vorburger, Song, & Petraco, 2015). However, there are no published methods on the correlation of areal datasets. Published methods either rely on 2D images, using methods such as canny edge detection, or an averaged profile of the areal topography (Chu, Thompson, Song, & Vorburger, 2013; Chu, et al., 2010). As firing pin impression correlation does not rely on 2D correlation, the author decided that a novel method for areal correlation of bullets would enhance current research in the correlation of bullet toolmarks. It is believed with the correct pre-processing methods to accurately separate individual characteristics within an LEA, areal correlation will be possible. This will allow for a shift from 2D to areal correlation, in which the surface topography information will be fully taken advantage of.

2.11 Aims and objectives

This thesis aims to convey the possibility of a shift from using comparison microscopy/ photometric imaging with visual comparison to open access advanced techniques for both the measurement and mathematical correlation of ballistic toolmark evidence. The literature survey conveys that while the capabilities for the shift exists, it has yet to be accepted as forensic evidence. This is due to a lack of transparency in current systems, both in the measurement techniques and subsequent analysis. The Daubert challenge facing forensic evidence in court recommends that the scientific principles behind forensic evidence be both reliable and repeatable. While advanced systems are predicted to satisfy the Daubert challenge, currently there is not enough research to conclusively prove this is the case. Correlation of ballistic toolmark evidence is currently treated as commercially confidential by all commercially available systems. As such the data processing techniques and correlation algorithms used cannot be directly compared and the standards used for correlation are unknown.

This study will investigate the need for the pre-processing of datasets, the differences in measurement systems and how toolmark impressions may vary under different conditions. The results of the studies will give insight into:

- The advantages of using areal measurement systems

It is expected that shifting to areal measurement will offer various advantages. Firstly, objectivity in the measurement set up will be reduced in the use of systems that are either set up so that user interaction cannot change measurement settings, or a measurement protocol can be set up so that the correct measurement settings are known to the user and can be easily repeated. In comparison microscopy, the measurement protocol is not repeatable, in that it is almost impossible to recreate the exact lighting conditions across laboratories.

- The application of open mathematical algorithms for the effective correlation of toolmark evidence

While the literature survey indicates that some of the commercial systems available have correlation software available, there is no available information on the algorithms/ pattern matching techniques being used. Therefore, a unified open source correlation algorithm should be introduced to enable a protocol that could be used across all forensic laboratories.

- Differences in dataset quality using various optical measurement techniques

Different optical techniques will have different advantages. While interferometry does not require user input with regards to lighting and resolution choices, white light focus variation allows the acquisition of steep flank angles using modulated light. There is also a significant difference in time needed for each acquisition between the systems, and as such the time expense must be considered. The efficacy of both approaches should be compared to ascertain which would offer the most advantage in data acquisition with regards to stipulations that must be met in a forensic laboratory.

- How correlation is affected in the use of different optical measurement techniques

Correlation algorithms used to correlate forensic evidence must be proven to be efficient, in that the risk of false positive matching is as small as possible. For example, a system that can theoretically consistently offer known matches in the top 5 of the hitlist will be preferred over one that is known to match erratically. The less time an expert examiner is needed to confirm matches indicated by a correlation system, the more advantageous the system is to a forensic laboratory. It is hoped that the quality of data proposed in this thesis will result in a correlation system that will need little validation from an expert examiner.

- Direct comparisons of correlation efficacy between 2D and areal systems

Current research shows little comparison into the correlation efficacy between the older systems relying on 2D pattern matching techniques and the more advanced systems. Only a few system comparisons have been completed, including those between IBIS and IBIS Heritage (Brinck, 2008). To gain a better understanding into how correlation efficacy is affected by a shift in measurement techniques, it is important to obtain a direct comparison between the techniques. In this thesis, 390 bullets and 390 cartridge cases have been acquired by four different systems. The results of two 2D systems have been included from a previous study, along with two advanced systems using different measurement methods. As the exact same sample set was used, it is possible to compare correlation between the methods. It is expected that as areal acquisition is much more detailed, and includes true height information for a surface, this will result in more data that is not relevant to the toolmarks being captured, for example waviness characteristics on the surface and possible outliers. As such, it is important to be able to differentiate between salient and non-salient data of the surface and be able to separate the two.

- Variations in toolmarks impressed in various ammunition materials.

It has been briefly discussed in several publications that the material of cartridge cases/ bullets will affect the toolmarks imparted into it. While there have been studies conducted using striations imparted by screwdrivers (Bachrach, Jain, Jung, & Koons, 2010; Baiker, Pieterman, & Zoon, 2015), there are very few studies that look into the difference in ballistic toolmark evidence. Brinck at al. did complete a comparison of IBIS and IBIS Heritage instruments based on their correlation efficacy in copper jacketed and lead bullets, however the study was limited to a small number of samples, and correlation between different materials was not considered (Brinck, 2008). To combat this, studies have been included within the thesis that use surface analysis to determine the topographical difference in toolmarks imparted into different substrate materials, and how it may affect the correlation of the toolmarks.

2.12 Novel Contributions

Firstly, novel contributions within the thesis firstly include a direct comparison of commercially available systems based on 2D imaging and matching, commercial systems based on areal measurement and correlation, and multi-purpose instruments capable of areal measurement of a wide range of surfaces.

Secondly, pre-processing techniques were put forward that have not been applied to the correlation of ballistic toolmark identification previously. The filtering techniques used on firing pin impressions are unique to this study, in which a Robust Gaussian filtration with cur off values of 75µm and 450µm were applied and have proved successful. The correlation of bullet toolmark impressions were completed using wavelet decomposition and allowed for the areal surface to be correlated, using both the D5 and D6 wavelet bands of a Discrete Wavelet Transform using a spline wavelet technique. This is to the author's knowledge the first instance in which correlation of bullet LEAs has been successfully completed using the areal dataset, rather than an averaged profile.

Finally, this is the first instance in which a system has been created which would result in a transparent and completely objective technique. While most current commercial systems rely on a 'match score', it means that the user is unable to ascertain how good a match has been found, as a match score tends to be an arbitrary value. However, in this case a percentage match is achieved, meaning the user gains a better understanding on the success of the correlation.

CHAPTER THREE

DATA FIDELITY

In the previous chapter, areal measurement methods applicable to ballistic toolmark identification were discussed. The Odyssey collection of 9mm Luger bullet and cartridge cases were also introduced.

Following from this, two measurement systems available to the author, the Alicona G4 focus variation instrument and the Alias ballistic imaging system were used for measurement of the Odyssey collection. This allowed the author to complete a comparison of a general-purpose measurement instrument and a system built specifically for the measurement of ballistic toolmark evidence.

Main considerations of a data fidelity comparison must include the overall quality of the measurement, the resolution available in each system and the presence of dropouts and optical spikes. As measurement systems are being considered with regards to forensic measurement of ballistic toolmark evidence, considerations must also be made on the time and cost efficiency of each instrument with regards to the casework of a forensic laboratory. These elements will be discussed in the following chapter.

3 Data Fidelity

3.1 Introduction

In this chapter, the measurement techniques used in the study will be discussed. While the two measurement systems used both rely on optical data, the acquisition techniques vary. Therefore, a comparison of the measured dataset is needed to determine the differences in quality of dataset.

In terms of ballistic toolmark data collection, the following chapter will detail the fidelity of data acquired using the Alias pOCT interferometer and the Alicona G4 Focus Variation (FV) technique. This includes the presence of noise and dropouts, the ability to measure the entire topography of a toolmark, the resolution and time taken.

A measurement system used in ballistic toolmark evidence should satisfy various conditions, to ensure that the acquired datasets are of the correct quality and the methodology is shown to be both reliable and repeatable. Optical measurement techniques have been chosen, due to the non-contact method being unable to damage forensic evidence. However, any optical measurement technique can be affected by issues such as reflection from a surface and local angle of the surface (Bolton-King, et al., 2010; Vorburger, Song, & Petraco, 2015).

The presence of optical spiking and dropout must be minimised to ensure the measurement is an accurate representation of the surface. The numerical aperture of the objective lens determines the maximum acceptance angle at which reflected light can be collected. Should there be a slope angle present on the surface that exceeds the numerical aperture of the lens, reflected light will not be collected and therefore the point will not be measured. This results in a phenomenon known as dropout, where no data will be recorded (Leach, 2010). Where a highly reflective material is to be measured, such as a metallic surface, the level of light that is reflected back to the camera is increased, when compared to a less reflective surface that will either absorb or disperse some light. Increased reflectance levels can result in an optical spike appearing on the surface, where more specular reflected light has been recorded on the camera and has caused a spike on the surface to be recorded (Nayar, 1989).

Due to the nature of individual characteristics being high frequency surface components in the micrometre scale, the resolution of the instrument used to acquire datasets must be sufficient to capture the small-scale topography perturbances. As previously stated in the literature review, a measurement system capable of a vertical resolution of greater than 0.1μm and lateral resolution of less than 1μm would be capable of acquiring the surface data, which would be suitable for analysis and correlation (Bolton-King, et al., 2010).

As ballistic toolmark identification takes place in busy laboratories who may have a large backlog of evidence to measure along with incoming evidence, it is also important to consider the time it will take to measure each bullet/cartridge case. While a measurement method with a longer scan time will acquire datasets of higher resolution, the extra time taken in acquiring the dataset may be too expensive for the forensic laboratory. Therefore, it could be acceptable to acquire datasets of a slightly lower resolution should the time taken for acquisition be significantly reduced. Another consideration would be the overall cost of the system, as a lower price instrument with capabilities suitable for measurement will be preferred over an instrument which may have slightly better capabilities at an increased cost.

3.2 Overview of focus variation

The Alicona G4 is a general-purpose measurement technique, and as such the system requires a high level of user input to ensure the fidelity of data acquired. User input required by this system includes:

1. Type of lighting used

The modulated lighting of the Alicona G4 allows the user to choose various aspects of the lighting conditions used to illuminate the object. Coaxial or ring lighting can be used, with ring lighting be able to overcome the numerical aperture of the lens and thus acquire surfaces with steep flank angles. Ring lighting is also advantageous in the measurement of reflective surfaces as lighting becomes more diffuse and thus less light is reflected back to the sensor, therefore minimizing any optical spiking caused by increased specular reflection. Further input on the exposure time of the light source can also change the amount of light available to be reflected back to the camera (Leach, 2010). Figure 3-1shows the difference in acquisition in variations of lighting conditions. Using the same measurement area of a firing pin impression, the exposure time of the lighting was changed. As shown in the two pictures, there is a variation in dropout dependent on the lighting conditions used. The percentage of valid points as quoted by the Alicona software when used by the author can vary from 98.49% to 99.99%.

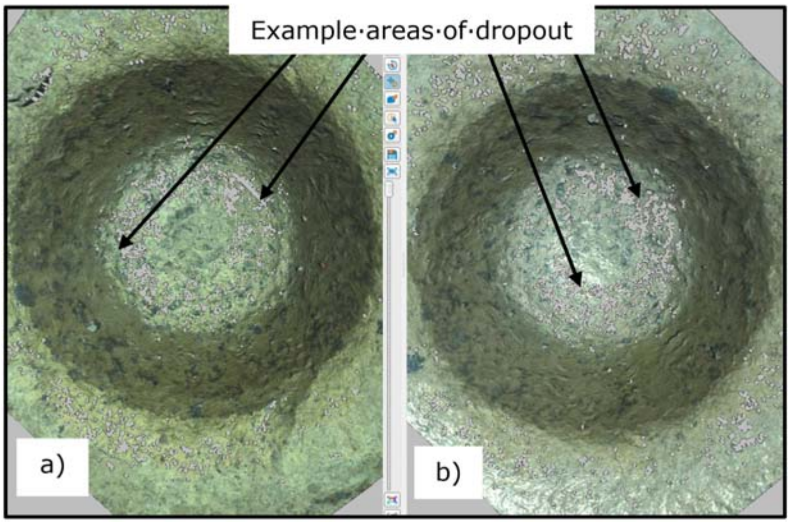


Figure 3-1: Variations in acquisition using varied lighting techniques, where b) has an increased exposure level compared to a).

2. Magnification of objective lens

The objective lens can be chosen manually, between 5, 10, 20, 50 and 100x magnification, since the choice of the lens effects the measurement. With a higher magnification, there is a higher numerical aperture for the lens. As such, a larger angle of light can be collected from the surface being measured but a smaller surface area can be measured at any one time. However, the choice of objective lens will determine the highest resolution possible, with higher magnifications having higher resolutions, and taking the longest time to acquire a measurement due to a smaller sampling distance. Therefore, there is a trade-off between the time taken to acquire a measurement and the resolution possible. In a busy forensic laboratory, an objective lens must be chosen that allows acceptable resolution of datasets within the shortest time possible.

Pictured in Figure 3-2, acquisitions of firing pin impressions were gained using the 50x and 20x objectives. While the numerical aperture of the 50x objective is inherently larger than the 20x objective, the area measured at any given time is smaller, and thus there is less illumination available. It is then possible to acquire more dropout in the measured dataset, despite a higher numerical aperture, due to issues with correctly illuminating the surface.

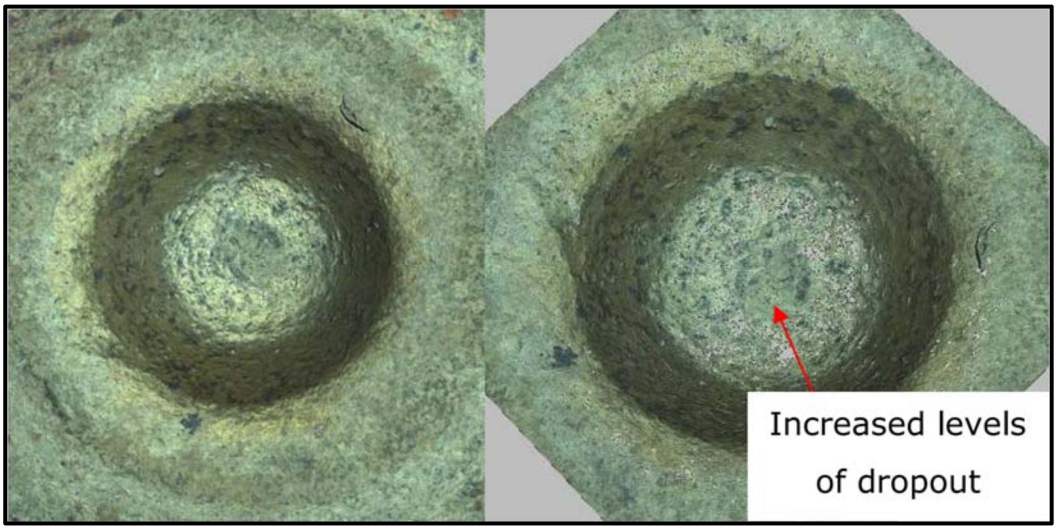


Figure 3-2: Differences in data between 20x objective (left) and 50x objective (right)

3. Lateral and vertical resolution

After choosing an objective lens, the resolution of the datasets can be further controlled by selecting a resolution within the range of the lens. The resolution range is determined by the objective lens used, with specific resolutions determined by the user, and instructs the software how often to acquire images after vertical movement of the optics and x/y movement of the surface. The software will select this resolution automatically dependent on the objective lens chosen, however if the user decreases the chosen resolution it is effectively subsampling the number of data points within an acquired dataset. The range of resolution available for each lens is shown in Table 3-1 (Alicona, 2017):

Table 3-1: Available resolution of the Alicona per objective lens

Objective	Lowest vertical Resolution (µm)	Highest vertical Resolution (nm)	Lowest lateral resolution (µm)	Highest lateral Resolution (µm)
5x	23.07	410	23.48	3.49
10x	5.71	100	11.74	1.75
20x	2.73	50	8.8	0.881
50x	1.19	20	6.4	0.640
100x	0.478	10	4.4	0.440

4. Measurement volume

As the Alicona relies on a small depth of field in the measurement of the surface, each single measurement is limited by the depth of focus and numerical aperture of the lens.

Therefore, to be able to acquire a dataset of the bullet topography or the base of the cartridge case, the user must specify a volume in which the Alicona software will acquire and stitch data together determined by user input regarding x,y and z dimensions of the measurement (Leach, 2010). However, if the user selects a volume that does not fully contain the region of interest, the data will not be acquired. This is most likely when choosing the vertical envelope, as the user may not take into account deep valleys or high peaks on the surface, in which case the software will crop this topography.
Considering the Daubert standard, the amount of user input needed in measurements acquired using the Alicona is too high. Having so many variables in user input that will all influence the measured dataset means that the method may not be repeatable. Therefore, to be considered as a technique available for forensic analysis it would be vital to implement a measurement methodology that a forensic examiner would have to adhere to, thus minimising repeatability issues. Ideally, a measurement procedure would be put in place that would not be affected by the slight variation in reflectivity in materials used in cartridge manufacturer, which has been presented in this study.

3.2.1 Calibration of Alicona system

The calibration of the Alicona G4 system is performed annually by a professional calibration company. In calibration, a set of standard calibration artefacts are used to ensure the measured values of the surface are within an acceptable range of the actual values of the surface. A measurement artefact designed specifically for focus variation was used, which contains a regular sinusoidal profile with peak spacing of 50µm and peak to valley height of 1.5µm, giving a resulting profile roughness of 0.5µm A step height artefact with steps of 1000µm is also used (Danzl, Hemli, Rubert, & Prantl, 2008). As of the most recent calibration (Appendix 6), the instrument used had 0nm deviation in lateral accuracy, 200nm deviation in the curvature of field, and 440nm deviation in vertical calibration. These results are within specified and acceptable tolerances.

3.3 Overview of Alias system

The Alias measurement system is a dedicated system for the measurement of ballistic toolmark evidence. As the system uses pOCT interferometry to acquire datasets, there is no need for the user to select measurement settings, as the illumination levels are constant and other settings are controlled by the measurement software.
In the first instance user input is required to enter forensically relevant information regarding the evidence into the software. This includes caliber, firing pin impression shape, number of rifling marks and twist direction of the rifling, all of which are class characteristics. As class characteristics are easy to distinguish, little prior training is needed

to input this information. There is also the option to add case information such as location of the crime scene and staff involved in the investigation. Information inputted at this stage can be used to reduce the number of potential matches before correlation takes place.
As a dedicated system, the measurement volume is not set by the user, instead, using class characteristics of the evidence the software will prompt a set measurement volume, which will control the travel of the optics and the x, y stage. However, in cases where the firing impression has caused a large vertical deformation some data may be cropped. To ensure the entire surface can be acquired within the set measurement volume, the user must ensure cartridge cases are loaded level with the stage and bullets are placed in the holder with a minimal amount of coaxial movement, as seen in Figure 3-3 and Figure 3-4:

Figure 3-3: Picture of cartridge case stage used by the Alias system

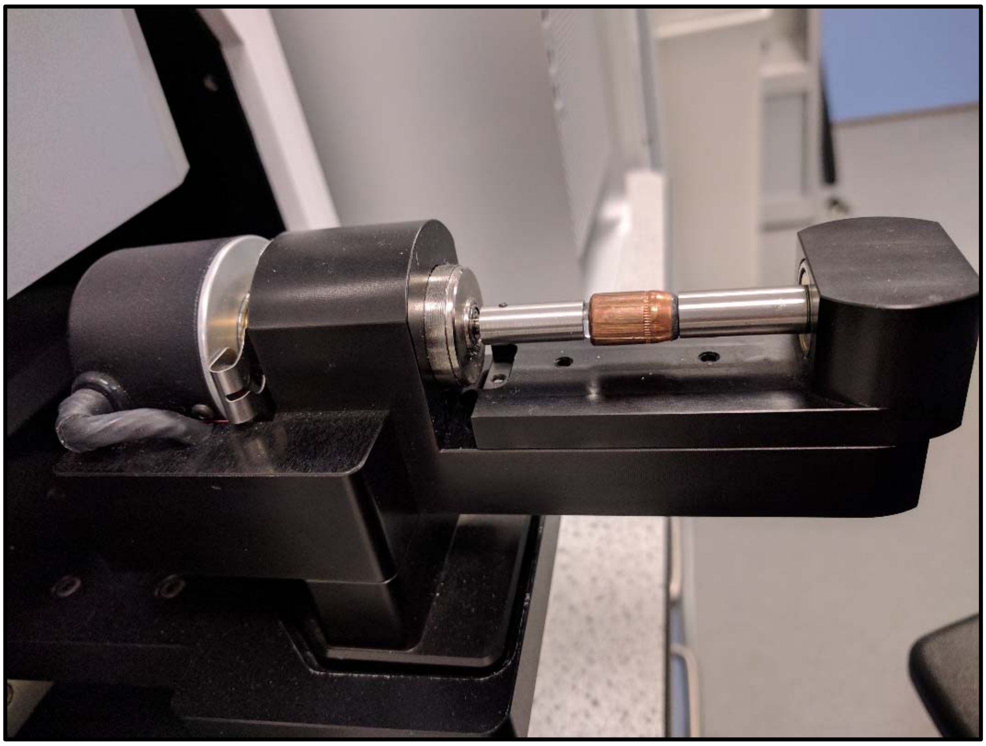


Figure 3-4: Picture of motorised bullet stage used by the Alias system

3.4 Resolution of imaging systems

The choice of resolution from each system is controlled in different ways and for the purpose of comparative studies it was important to ensure that the resolution of datasets from each system was comparable. There is therefore a need to minimise variables to ensure an accurate comparison, and thus resolution of the datasets must be as similar as possible.

Using the Alias system, the resolution of the actual measurement is set by the interferometry method, based on the coherence length of the light source and accuracy of movement in the stepper motors. Therefore, all datasets are acquired initially with a vertical resolution of 200nm and lateral resolution of 2µm. As an average 9mm cartridge case measurement will acquire 64 million data points, each containing the x, y and z coordinates of the point, the dataset potentially will become very large. Therefore, in the context of a forensic laboratory, it was decided to use subsampling of the datasets to ensure systems would be able to cope with dataset size and further correlation of evidence would be more time efficient. The fewer data points, the less space needed to save in a dataset, and correlation will be faster.

However, downsampling must be considered with regards to the scale of interest. Downsampling results in the increase of distance between recorded data points, and as such this distance, or lateral resolution, must not exceed the scale of interest for individual toolmarks. It has been found in previous studies that a lateral resolution of <3μm is needed in the original dataset to be able to downsample without losing information regarding the individual characteristics, therefore it is possible to subsample datasets acquired using the Alias system (Bolton-King, 2012).

It was decided that downsampling the datasets by 2 would be an acceptable compromise. Lateral resolution would decrease to 4μm while vertical resolution remains at 200nm, and thus downsampled datasets still have the resolution needed to differentiate individual characteristics in ballistic toolmark evidence.

To ensure measurements in the following studies would be comparable, it was necessary to ensure that datasets acquired using the Alicona would be as similar as possible, which would allow for a comparison of measurement technique without further influence from differences in resolution of data. Using the Alicona it is firstly necessary to choose an objective lens that has enough magnification to enable the choice of resolution, while simultaneously having a large enough numerical aperture to minimise the risk of data dropout and minimising acquisition times. Both the 10x and 20x objective lenses were capable of offering resolution needed, however the quality of data acquired was affected by the lens chosen. Using the 20x objective sometimes results in data dropout in shadowed areas of the surface. Due to this and the fact that an increase in objective magnification results in a significant increase in the time taken to acquire the dataset, it was decided to use the 10x objective lens.

3.5 Measurement methodology using focus variation

As the Alicona Focus Variation instrument is general purpose, it relies upon user input to ensure correct acquisition of the surface. As it is within the remit of this thesis to reduce effects of user input in ballistic identification, a method would have to be created that would minimise user input as much as possible for the Alicona to be considered a viable measurement technique.

Over the course of several experiments, a method was devised that would allow all cartridge cases in the Odyssey collection to be measured[4] without the needs for the user to change any measurements settings, thus creating a repeatable method. As the Odyssey collection consists of 390 9mm Luger cartridge cases, consisting of both nickel and brass primer caps, the methodology had to allow for both slight variations in primer cap size and materials across various manufacturers. As all samples were metallic it was expected that

[4] In this study measurements of the primer cap surface were acquired. Should the entire cartridge case base need to be measured, the dimensions of the measurement volume would need to be increased appropriately.

the reflectance properties would be similar, however there may be slight variations dependent on the material and condition of the cartridge case.

It was found that using the measurement setting tabulated below in Table 3-2 allowed for the correct measurement of all cartridge cases within the Odyssey collection using the focus variation method. Therefore, these measurement settings can be used for acquisition of 9mm caliber cartridge case primer caps without any further user input. It should be noted that should any other caliber of cartridge case be measured, while the lighting settings will continue to be optimal the measurement volume will vary.

Table 3-2: Measurement settings used to acquire cartridge case measurements using focus variation techniques.

Measurement variable	Designated setting
Objective lens	10x
x measurement range	-2250 to 2250µm
Y measurement range	-2250 to 2250µm
Z measurement range	-25 to 700µm
Subsampling	/4- To give resolution stated below
Resolution after subsampling	Lateral: 4µm Vertical:0.34µm
Lighting conditions	Ring light Polarised
Lighting settings	17.3ms exposure 0.3 contrast

The user must first determine the lowest point of the firing pin impression, centre the measurement at this point (using crosshairs on the measurement screen) and ensure the objective lens is situated where it is completely below focus for all points across the surface, as shown in Figure 3-5. At this point the user must set all measurement coordinates to 0.

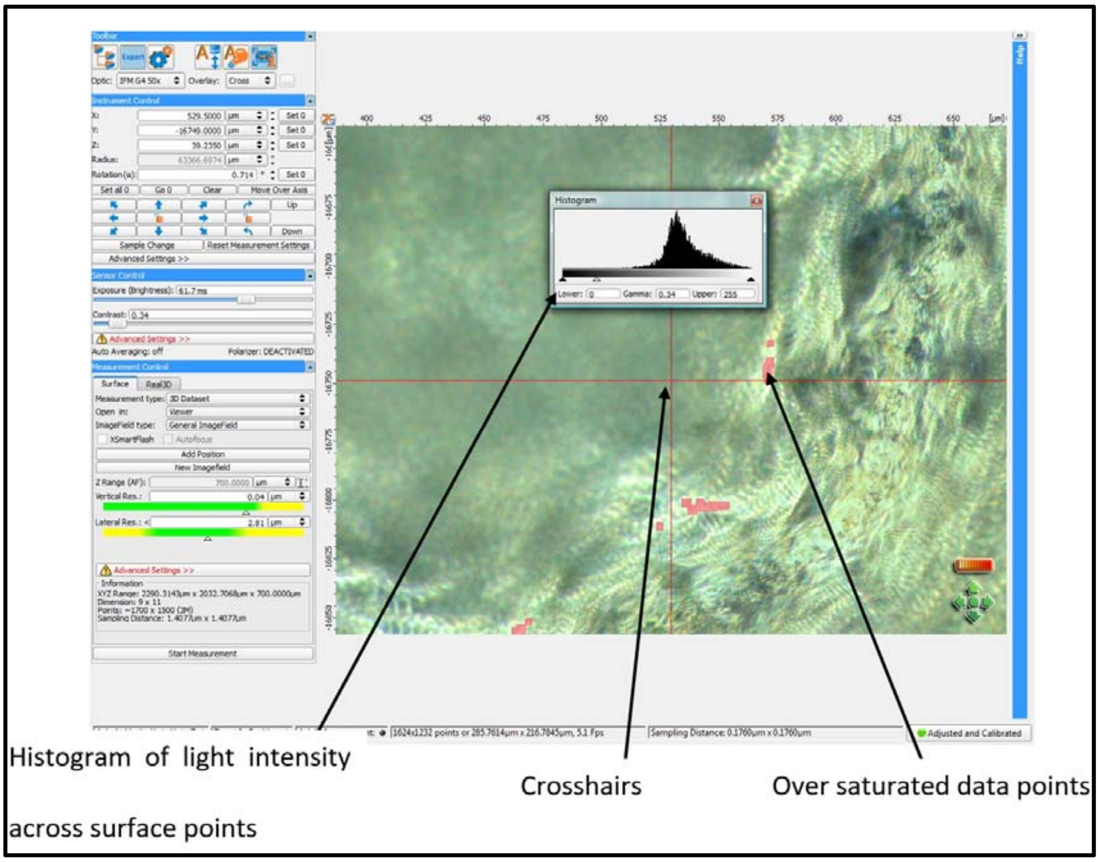


Figure 3-5: Determining the focus range of the measurement

The resulting measurement will have taken 15 minutes to acquire and will contain 1481x1602 points. The measurement area is slightly larger than needed, however it ensures that the area of interest is always acquired.

In the case of bullet measurement, a secondary rotary stage was utilised so that an entire 360° measurement around the bullets surface could be measured without the need for the user to move the bullet manually (Figure 3-6). To mount the bullet in the rotary stage, a steel rod was machined so that a cone of material was removed from the end. The cone was machined larger than the ogive of a 9mm Luger bullet to allow the bullet to be mounted using a two-part silicone material (Microset), ensuring the bullet would not be damaged during measurement while eradicating effects of material creep sometime present in softer materials.

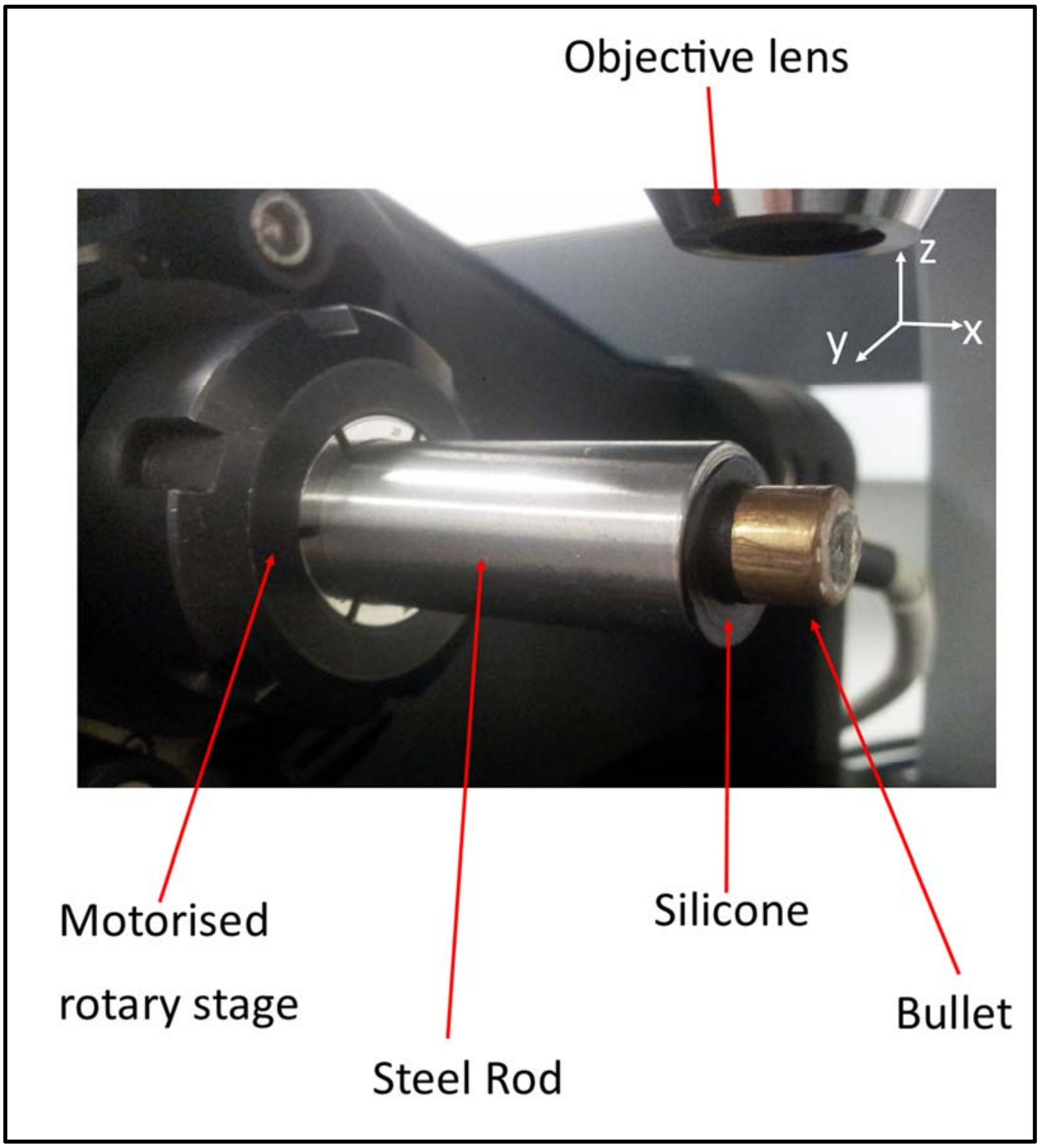


Figure 3-6: Picture of mounting technique for measurement of bullets using the Alicona

Each bullet measurement was acquired by using the x dimension that related to the measurement area of the 10x objective only (1.6mm), rather than stitching a larger measurement area, due to time constraints. However, this measurement area is comparable with that automatically acquired in the Alias system.

The user must ensure that the live onscreen image is moved so that the area of interest is placed on beginning of the striations, while ensuring no striation data is missed. The user will then ascertain the point at which the objective lens is lower than any focused point within this region, and will then set the measurement coordinates to 0.

Using the 390 9mm Odyssey bullets, of varying manufacturer and material, it was found that the following measurement settings would acquire high quality data in all cases (Table 3-3):

Table 3-3: Measurement settings for bullet acquisitions

Measurement variable	Designated setting
Objective lens	10x
Measurement range	Full 360° rotation Y= N/A X= 0 Z= -20µm to 500µm
Resolution	Vertical: 340nm Lateral: 4µm
Overlap between singular measurements for stitching	30%
Lighting conditions	Ring light, Polarised
Lighting settings	Exposure: 44.64ms, Contrast: 0.44

Using the above settings results in a dataset with 811,268 data points and will take approximately 35 minutes to acquire.

3.6 Measurement methodology using the Alias system

The Alias measurement system has been designed as a closed measurement system, and as such the user input needed to acquire each dataset is kept to a minimum. Such a system is better suited in a forensic context than an open measurement system, as it satisfies the repeatability condition of the Daubert principle (Page, Taylor, & Blenkin, 2011).

The first step of acquiring datasets is to select the caliber of evidence to be inputted into the system, the vast majority of which are included in an existing library. In the case of the Odyssey collection, all evidence was entered as being 9mm Luger caliber. Once the caliber has been inputted into the system, the dimensions of the evidence are automatically created by the software. This instructs the measurement system on the x, y and z

dimensions needed to acquire the entire topography of a cartridge case or bullet. There are no lighting settings to be changed using the Alias system, and resolution of the measurement is fixed, therefore user input for the measurement is minimised.
Once the dataset has been acquired, it is possible to subsample for exportation. In this case, the sampling distance is essentially decreased, thus decreasing the lateral resolution and the number of points in the dataset. Subsampling was used for exportation of Alias acquired datasets to ensure resolution was similar to resolution gained using the Alicona instrument. Settings used for all measurements (bullet and cartridge cases) are shown in Table 3-4:

Table 3-4: Measurement settings used by Alias

Measurement variable	Designated setting
Objective lens	20x
x measurement range	Determined by caliber of evidence
Y measurement range	Determined by caliber of evidence
Z measurement range	Determined by caliber of evidence
Subsampling	/2- To give resolution stated below
Resolution after subsampling	Lateral: 3.94µm Vertical:0.2µm
Lighting conditions	N/A
Lighting settings	N/A

3.7 Calibration

The Alias interferometer is calibrated by the manufacturer (Heliotis). ISO standard calibration artefacts are used, including step and spherical artefacts, to ensure measured height points are within an acceptable range of the artefact dimensions.

3.8 Data Quality

The following is a discussion on the overall quality of measurement data acquired from each system.

3.8.1 Alias system

Overall the Alias system acquires high quality datasets in a considerably shorter time than compared to the Alicona datasets, taking around five minutes per both bullet and cartridge case scans. While the Alias does suffer data dropout at high flank angles, as seen in the trench around the primer cap in Figure 3-7, such dropout will have no effect on further analysis (due to cropping the dataset to the ROI) and therefore it is deemed acceptable. Where salient information is contained in steep angles on the surface, i.e. the flanks of the firing pin impressions and the transition to land engraved areas on bullet surfaces (Figure 3-7-Figure 3-8), there have been no instances of data dropout, thus showing pOCT is capable of overcoming slope angle issues in large form deviations using a small spot size and vertical scanning of the optics.

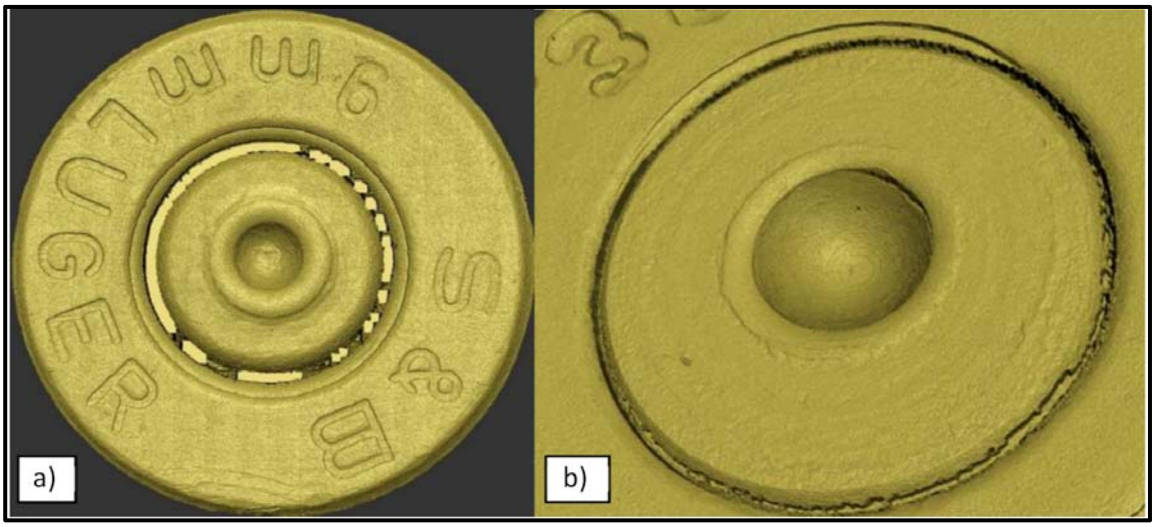


Figure 3-7: Full cartridge case acquisition (a) with visible dropout in the primer cap trench (b), acquired using Alias system

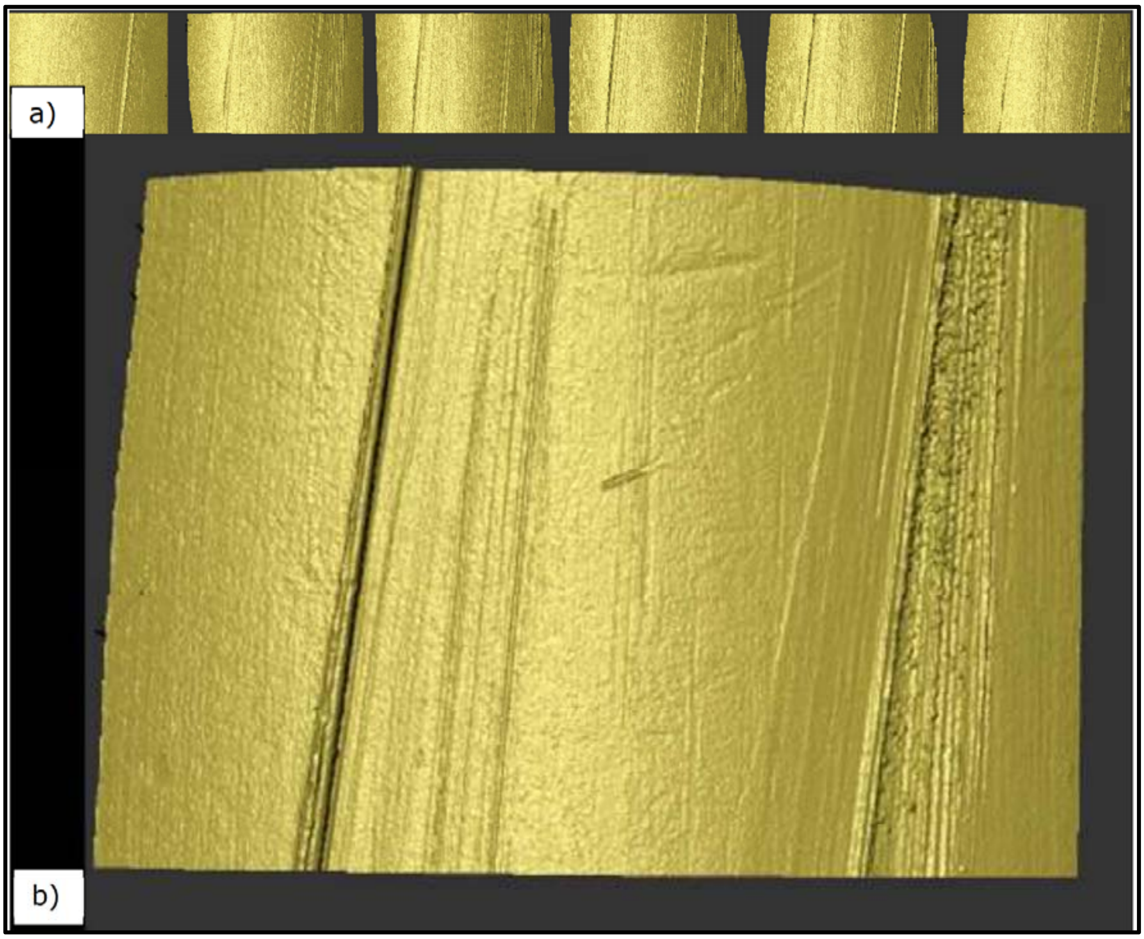


Figure 3-8: Full acquisition of 9mm bullet circumference, acquired using Alias system.

Individual characteristics can be clearly visualised on each dataset, and there is no noise present on the measured surface. Therefore, it is expected that these datasets will perform well in further analysis.

3.8.2 Alicona measurements

Measurements acquired using the settings detailed above resulted in satisfactory quality acquisitions with no instances of optical spiking or data dropout in areas of interest. However, it can be seen that there are some lower frequency measurement artefacts present in the measurement, which appears as waviness- see Figure 3-9 below:

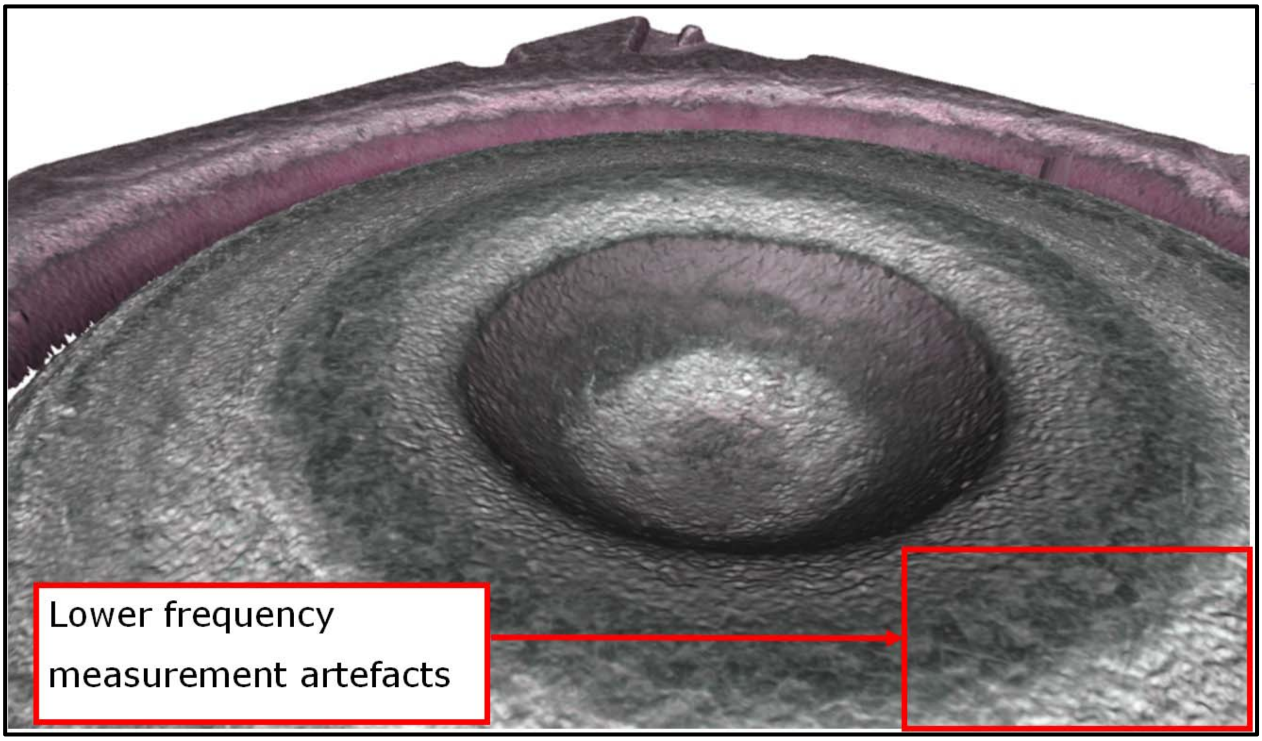


Figure 3-9: Alicona acquisition of firing pin impression

It is expected that the measurement artefacts are due to the reliance of averaging data from neighbouring points for data acquisition. This averaging may result in waviness components being introduced into the surface topography (Vorburger, Song, & Petraco, 2015). The accurate acquisition of height point data in focus variation relies upon a distinguishable difference between a focused and non-focused position. However, lighting conditions and effects such as barrel distortion can affect the difference in contrast between neighbouring points, thus introducing measurement artefacts (Helmli, 2011). Measurements still show the ability to acquire individual characteristics in ballistic toolmarks (Figure 3-9 and Figure 3-10), and therefore there is a possibility that effective correlation can be achieved, however it is expected that in comparison with Alias datasets this correlation will be less effective. Due to the measurement artefacts imparted into Alicona measured datasets, it is suspected that the individual characteristic toolmarks may be masked or skewed by artefacts not present on the original surface. As correlation relies on the exact height of each measured point on the surface, any measurement artefacts will skew the measured height points, and therefore in turn skew the correlation results.

Figure 3-10: Alicona acquisition of bullet topography

3.8.3 Comparative 2D profiles

Using the SURFSTAND software, 2D profiles can be extracted from areal measurements manually. The user defines the line upon which the profile is to be extracted, and the software displays all measured height points from this line as a 2D profile. This was used on a bullet and cartridge case, measured using both techniques, to give a visual indication of differences in measurement quality. Datasets were not subjected to any filtering before 2D profiles were extracted to ensure a direct visualisation of the measurement quality.

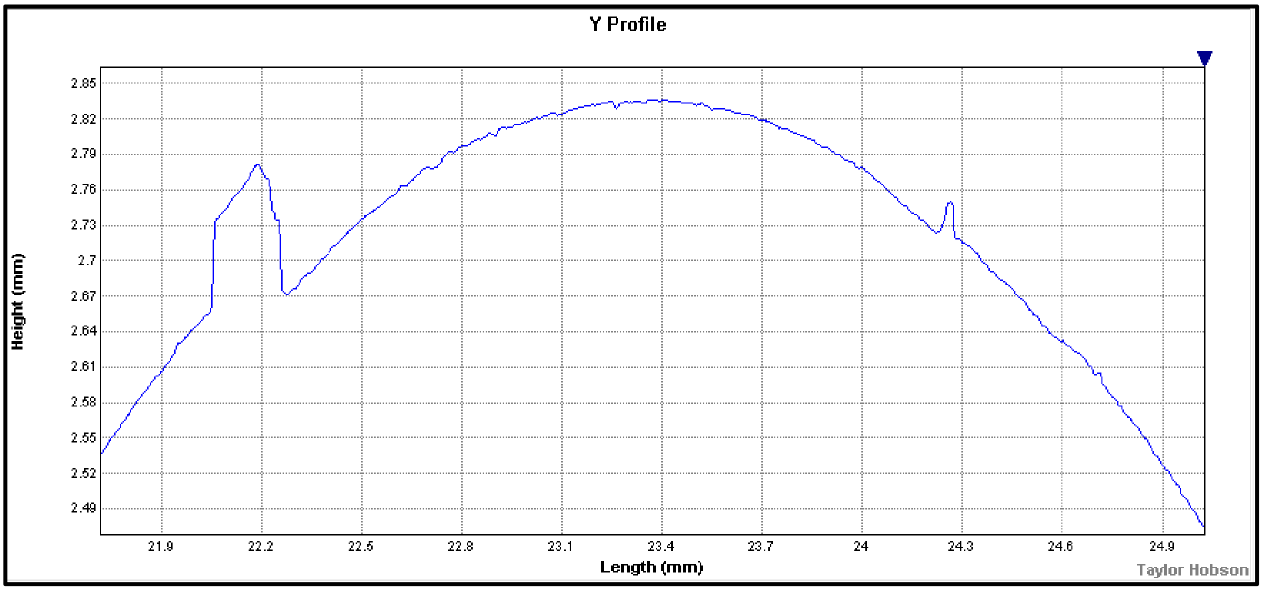


Figure 3-11: 2D profile of a bullet measurement extracted from an Alias acquired dataset

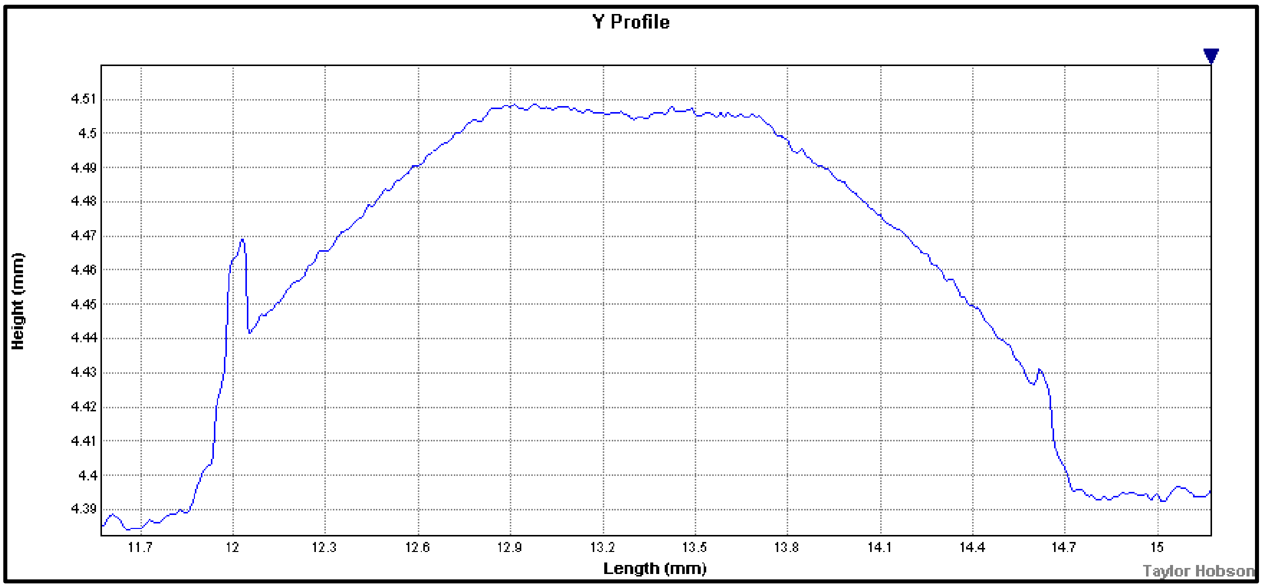


Figure 3-12: 2D profile of a bullet measurement extracted from an Alicona acquired dataset

With regards to bullet measurements, visual inspection of 2D profiles shows show a marked difference in measurement quality. The profiles extracted were taken from the same bullet, and as such the two measurements should be visually similar. Shown above in Figure 3-11, the Alias acquired profile appears to have little inclusion of measurement artefacts, whereas in comparison the Alicona acquired dataset as seen in Figure 3-12 has be affected both in form and waviness due to measurement artefacts. In both cases, the large spike on the left of the measurement and the smaller spike on the right are due to ploughing of material on the bullets surface as the barrel rifling plastically deforms the bullets surface (Xie, Xiao, Blunt, Zeng, & Jiang, 2009).

With regards to cartridge case measurement, there is a significant difference in the measured depth of the firing pin impression. It is expected, as focus variation relies on light intensity to infer height point data, that the loss off light intensity through shadowing of the steep flanks of the firing pin impression has resulted in the incorrect recording of the height points within shadowed areas as seen in Figure 3-14. Increasing light intensity for the firing pin impression would ultimately result in a high percentage of erroneous data points where shadowing does not occur due to light saturation, and therefore this cannot be corrected using the focus variation technique. As the Alias interferometer does not rely upon light intensity, this issue does not occur within the firing pin impression, and therefore height points within the firing pin impression will be more reasonable (Figure 3-13).

While it is accepted that a firm conclusion on variation in data quality would require a direct profile comparison facilitated using a physical cross-section of the toolmark, the author was unable to do so with the samples measured. The terms of using the Odyssey collection for measurement include ensuring samples remain in pristine condition for further research, and it was therefore not possible to destroy any samples to measure the true cross-sectional depth/ topography.

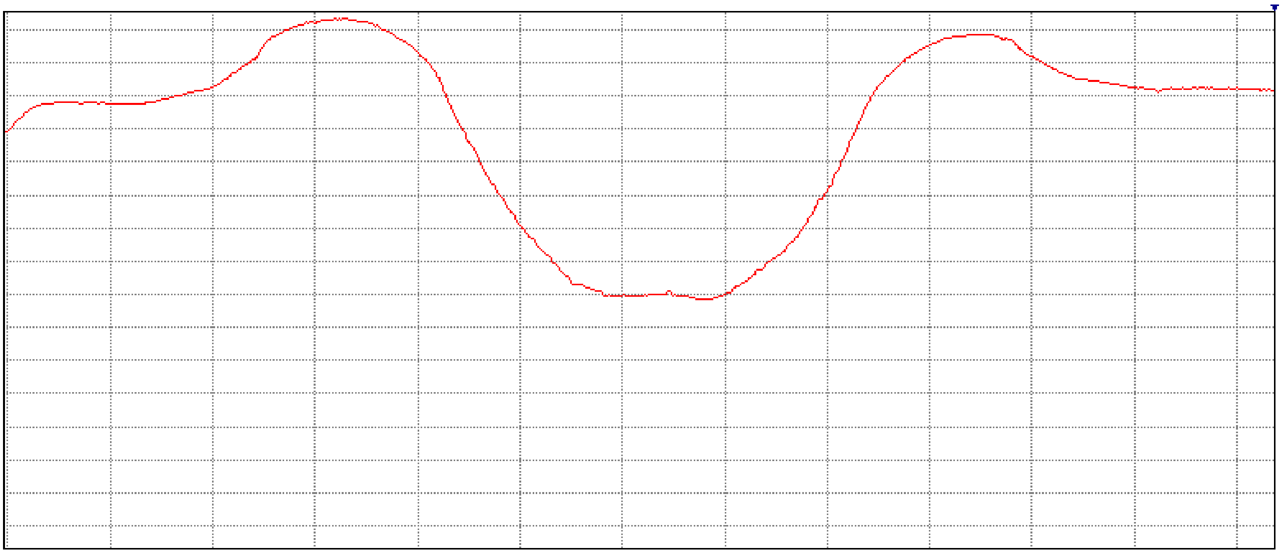


Figure 3-13: 2D profile of primer cap extracted from Alias acquired dataset

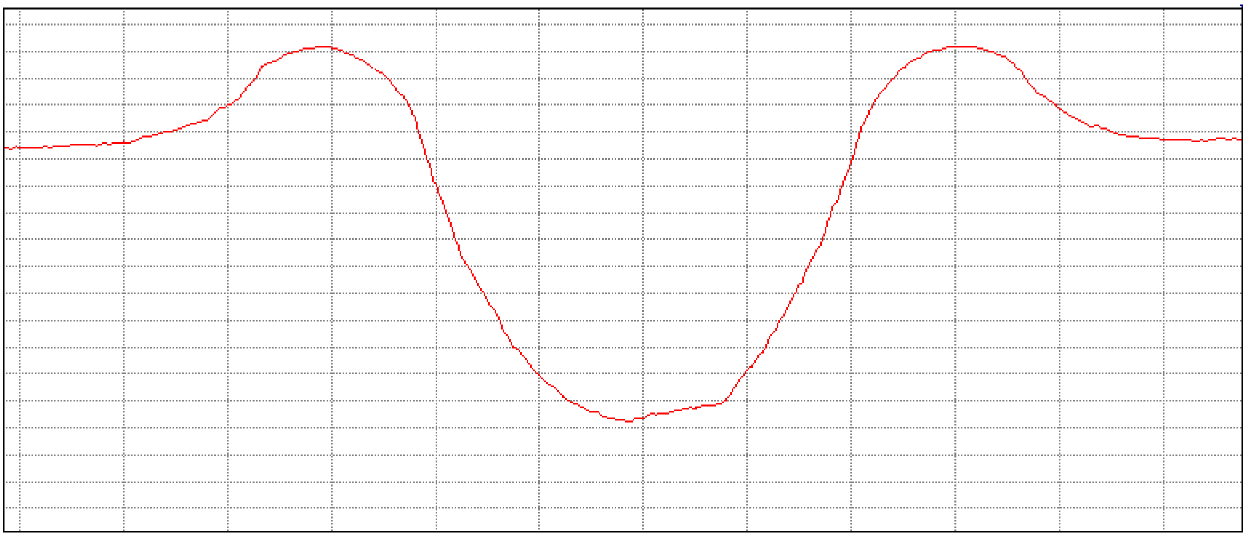


Figure 3-14: 2D profile of primer cap extracted from Alicona acquired dataset

3.8.4 2D microscopy

Using microscopy to gain 2D images of the ballistic toolmarks with the Leica S6D microscope, when compared with the advanced areal measurements systems, the lack of data quality is apparent.

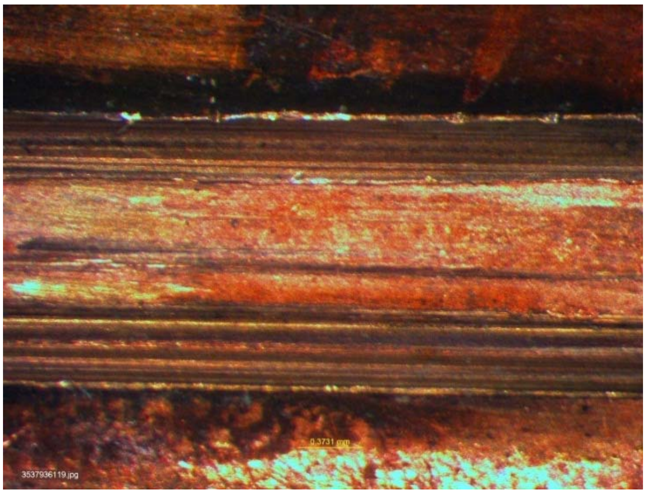


Figure 3-15: 2D image of land engraved area of bullet

Lighting conditions available result in areas where reflectance is an issue, while simultaneously shadowing occurs in areas surrounded by higher topography. Combined with a small depth of focus, it is apparent that visualisation of individual characteristics within the toolmarks is considerably more difficult, as seen in Figure 3-15 and Figure 3-16.

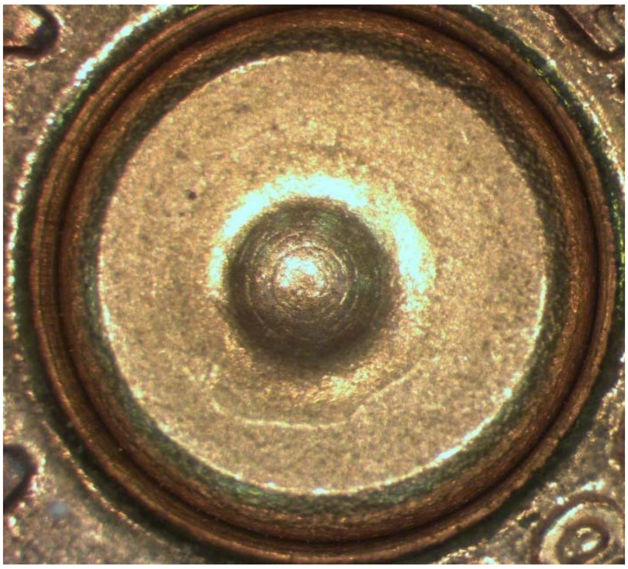


Figure 3-16: 2D image of fired primer cap

3.9 Discussion and conclusions

It has been found within the experiments of this chapter that both the measurement setup and data quality of the measurement systems vary substantially. While the Alias measurement system is automatic, with settings defined by the caliber of the evidence, the

Alicona system requires a high level of user input. According to the Daubert principle, user input should be kept to a minimum to increase repeatability of measurements and to ensure data quality is comparable across various forensic laboratories. Therefore, in this respect, the Alias measurement system is more suited to the measurement of ballistic toolmark evidence. With regards to data quality, there is an obvious disparity caused by the differences in measurement technique. While the Alias relies upon pOCT interferometry, and therefore is less prone to measurement artefacts due to reflectance and slope issues, the Alicona focus variation technique results in measurement artefacts being included in measured datasets due to height point data being skewed by light reflectance issues. As correlation techniques will rely on height point data, it is expected that correlation will be skewed by these measurement artefacts. There is also a significant time difference in data acquisition, with Alias data acquisition taking around 5 minutes per sample, compared to 15 minutes for each Alicona acquisition.

Overall it is expected that the use of Alias acquired datasets will result in a more accurate correlation compared to Alicona acquired datasets due to a higher quality of acquisition (Figure 3-17), however further testing is needed, which follows in Chapters 4 and 5.

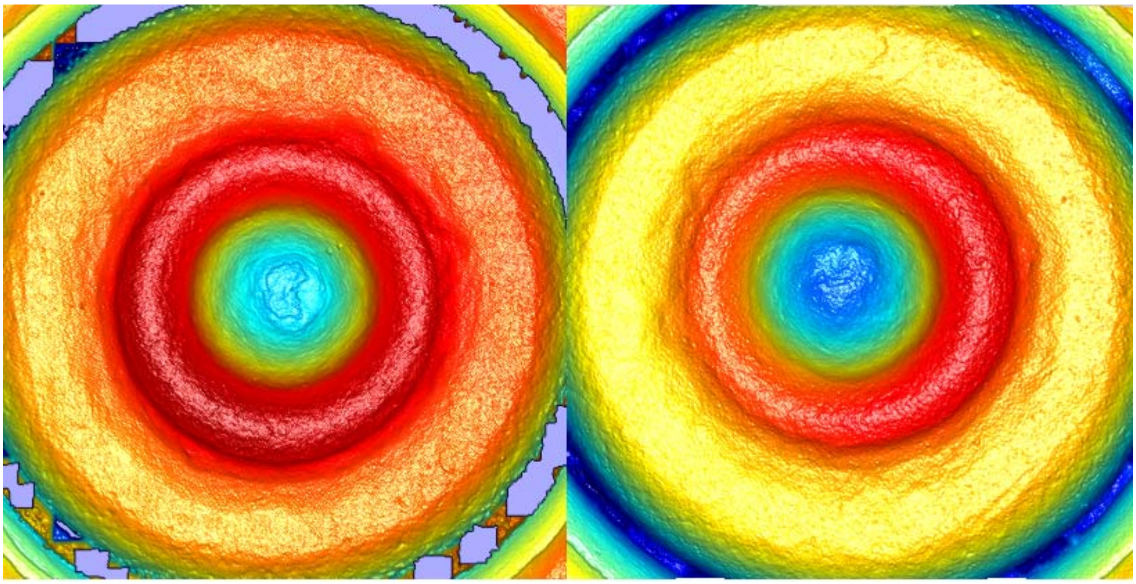


Figure 3-17: Visual differences between Alias and Alicona acquired datasets using the same primer cap for measurement

CHAPTER FOUR

CARTRIDGE CORRELATION STUDIES

It has been discussed in chapters two and three that a shift from greyscale to areal measurement will result in an increase in acquired surface information, where the height point of each x,y coordinate is recorded at micron scale intervals. Within areal measurements, different spatial frequencies present in the dataset will relate to different characteristics of the cartridge case. It is vital for accurate correlation between cartridge cases that only spatial frequencies relating to individual characteristics are used.

In this chapter an investigation into efficient separation of salient spatial frequencies for correlation is achieved. These pre-processing methods were compared across two measurement techniques, and the results from these studies were further compared to those gained in previous Odyssey studies.

A discussion on toolmark variability is also presented, with a focus on whether or not the material composition of the primer cap may affect the topography of the toolmark.

4 Cartridge correlation studies

4.1 Introduction

The correlation of areal datasets of ballistic toolmark evidence relies upon both the quality of the original dataset and the pre-processing methods applied to ensure the correct separation of individual characteristics from other spatial elements of the surface topography. Therefore, the aim of this study was to determine the most efficient correlation of ballistic toolmark evidence, based upon differing measurement and pre-processing techniques.

The Odyssey collection of 390 fired 9mm Luger cartridge cases were used for this study as it would allow for a direct comparison of hitlist efficiency to the 2D identification systems used in previous studies. Previous studies used two commercially available ballistic identification systems (which must remain anonymous as part of a confidentiality agreement), each of which relied upon 2D pattern matching techniques to determine the level of similarity between two separate pieces of ballistic evidence. Previously the Odyssey collection has been used as a research tool to determine the level of interoperability between commercially available systems as part of an EU funded project (Yates, Akghar, Bates, Jopek, & Wilson, 2011), however it was found that there was no interoperability. As such, the Odyssey collection will be used in the following studies to determine the difference in correlation efficiency between different measurement techniques and will also offer a direct comparison of hitlist efficiency between the 2D systems used in previous studies and the areal systems used in the following studies.

Ultimately, this will lead to a better understanding of the effect on ballistic toolmark identification in a shift from 2D to areal systems, the consequences of differing measuring techniques and the effect this will have on the correlation of toolmarks. As the Odyssey collection is comprised of eight different cartridge case manufacturers, each using variations in primer cap material, it is also possible to study the effect of primer cap material on the overall correlation efficiency of ballistic toolmarks. The results of these studies will then be discussed with regards to the Daubert principle and Bayesian framework of evidence in Chapter six to determine how the techniques could be applied to forensic casework.

4.2 Overview of the Odyssey collection: Cartridge cases

The Odyssey collection compromises of 390 cartridge cases. Within the 390 cartridge cases, 22 have been designated as the “test objects”. The test objects are known to have two other cartridge cases within the remaining 368 that are known matches, i.e. they were fired from the same gun. The remainder of the collection (324) are designated as the background sample-which were fired using 184 firearms. Figure 4-1 is an overview of the cartridge cases within the Odyssey collection.

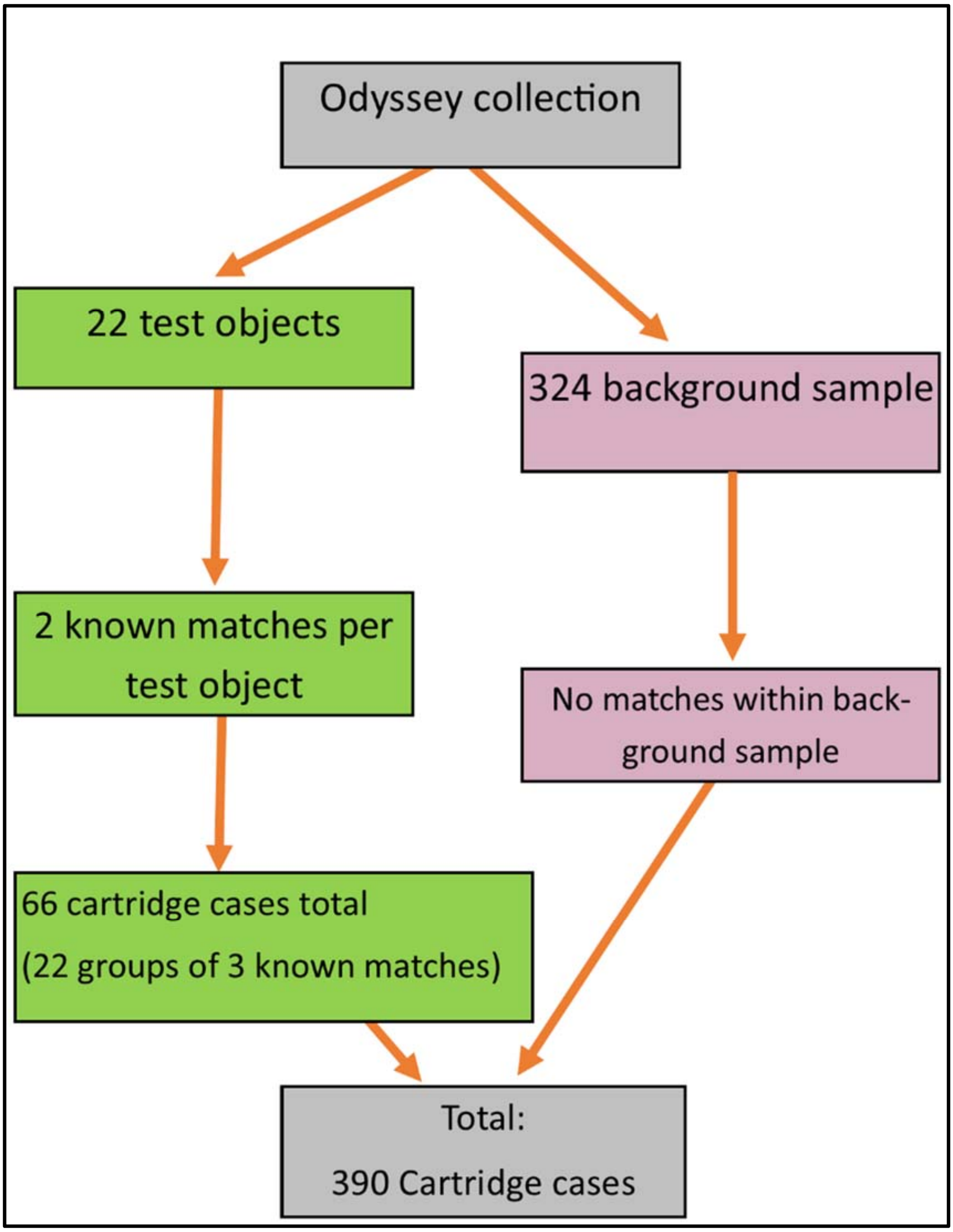


Figure 4-1: Overview of the Odyssey cartridge case collection

Each cartridge case contained within the Odyssey collection was engraved with a four-digit identifying number to ensure the cartridge case could be traced back to the gun it was fired from. However, to ensure the Odyssey collection remains available for further blind testing this information will not be included.

Table 4-1 below shows the manufacturers of both firearms and cartridges. Cartridges were randomly assigned to firearms for test firing, and variation in materials were not considered.

Table 4-1: Manufacturers used in study

Firearm manufacturers	Cartridge manufacturers
Beretta	Geco
Beretta U.S.	S&B
Beretta Italy	GFL
Beretta U.S.A corp	SBP
Pietro Beretta	PMC
Pietro Beretta Italy	CBC
Pietro Beretta Gardone	FC
Pietro Beretta s.p.c	MRP
TARIQ	
Maadi	
MAS	
Helwan	
LIW	
Taurus	
Vektor	

4.3 Determining Pre-processing methods

The aim of the following study was to determine the correct pre-processing techniques to be applied to cartridge case datasets to ensure efficient correlation of ballistic toolmark evidence. The pre-processing techniques must result in the effective segregation of salient data from other surface topography such as waviness, form and noise. It is also important to ensure that a frequency band is chosen for filtration that will not only segregate salient data but will also ensure all salient data is included.

Should data be used for correlation that does not relate to the individual characteristics, for example optical noise or waviness of the surface, this could result in false positives in correlation results. This is due to the fact that the waviness of a surface cannot be considered individual, and may share similarities across ballistic toolmark surfaces that are not actually considered a match. For example, the overall shape of the firing pin impression is a class characteristic, and as such there will be a high similarity across a large group of firearms. Should the class characteristic to be used in correlation, matches may occur in different firearms from the same manufacturer, thus producing a false positive match (Heard, 2013).

The following study used variations in pre-processing to ascertain the effect on the cross-correlation between a subset of the Odyssey collection. The sub set was chosen at random by an independent party, to ensure testing remained blind and user bias was not introduced to testing.

Filtering tests were conducted on the firing pin impressions of a subset of the Odyssey collections with $ACCF_{max}$ correlation to ascertain the best differentiation in correlation between known matches and known non-matches. For each filtering test, two test objects were correlated against both a known match cartridge case and a non-match cartridge case. These preliminary tests were then expanded to include 10 non-matches per test object to validate any filtering tests that appeared to allow for a clear discrimination between known-match and non-match correlation results.

During testing, the following variables were interchanged: type of levelling used, type of filtering and cut off lengths of the filtering. The methodology of steps taken between the original dataset and that used for correlation is shown below in Figure 4-2:

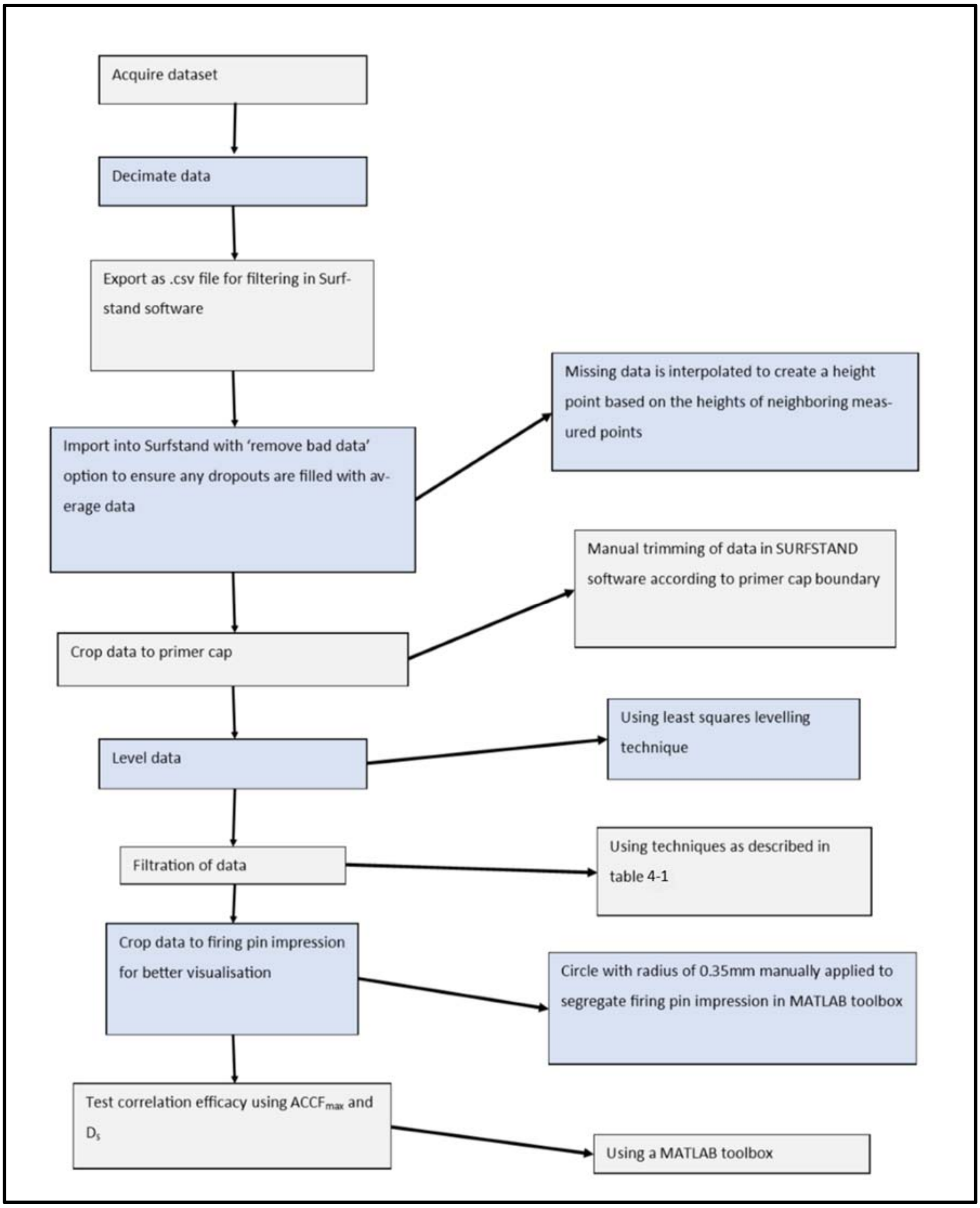


Figure 4-2: Flow of cartridge case correlation methodology

It was determined that the surface frequencies contained within the base of the cartridge base (i.e. manufacturer stamps) and the possibility of a differences in angle and depth between primer cap and cartridge case base could skew both the levelling and filtration results. Therefore, the order of operations described above were chosen to minimise this risk.

Table 4-2 shows the variations in filtering used in the study to ascertain the correct pre-processing methods needed for correlation:

Table 4-2: Table of filtering methods tested

Test	Levelling used	Filtration used	High pass cut-off (μm)	Low pass cut-off (μm)
1-	Least squares	Robust Gaussian	25	2500[5]
2	Robust polynomial- 2nd order	Robust Gaussian	25	2500
3	Robust polynomial- 3rd order	Robust Gaussian	25	2500
4	Robust polynomial- 2nd order	Robust Gaussian	25	800
5	Least squares	Robust Gaussian	50	600
6	Least squares	Robust Gaussian	50	500
7	Least squares	Robust Gaussian	50	450
8	Least squares	Robust Gaussian	75	450
9	Least squares	Robust Gaussian	75	400
10	Least squares	Robust Gaussian	100	450
11	Least squares	Spline	25	450
12	Least squares	Spline	50	450
13	Least squares	Spline	75	450
14	Least squares	Spline	100	450
15	Least squares	Spline	150	450

In filter test number one, the ISO standard filtration was used, to ascertain whether standard filtration would be suitable. It was found however that correlation results were unable to differentiate between known matches and non-matches, thus indicating smaller scale individual characteristics were not fully separated from other surface components. To ensure the issues were not due to inefficient form removal, testing was then completed using various form removal techniques. When it was found that changing form removal techniques did not result in differentiation in correlation results, it was decided to vary the frequency of the cut off lengths to ensure both high frequency noise and low frequency topography were separated from individual characteristics within the toolmark.
It was found firstly in the testing of filtration methods that using the ISO standards (test 1) for the pre-processing in data would result in inefficient correlation, due to incorrect

[5] This test conforms to the ISO standard of filtration for a dataset with the same area (4mmx4mm) as used in this study

differentiation between salient and non-salient data contained within the surface topography.

It can be seen in Figure 4-3 below, while the known matches give very good correlation, with both percentage matches in the high 90th percentile, using these pre-processing methods also result in non-matches giving a percentage match over 85%. Therefore, it can be shown that the pre-processing techniques are resulting in form/ class characteristics remaining as part of the surface topography, and positively skewing the correlation.

In using polynomial levelling on the datasets rather than least squares, it was found that $ACCF_{max}$ scores, while lowered, would remain too similar between a Known Match (KM) and Non-Match (NM). Using polynomial levelling also results in a waviness artefact being introduced into the surface, which results in topography distance D_s scores becoming much higher than those where least squares levelling had been used (Figure 4-4).

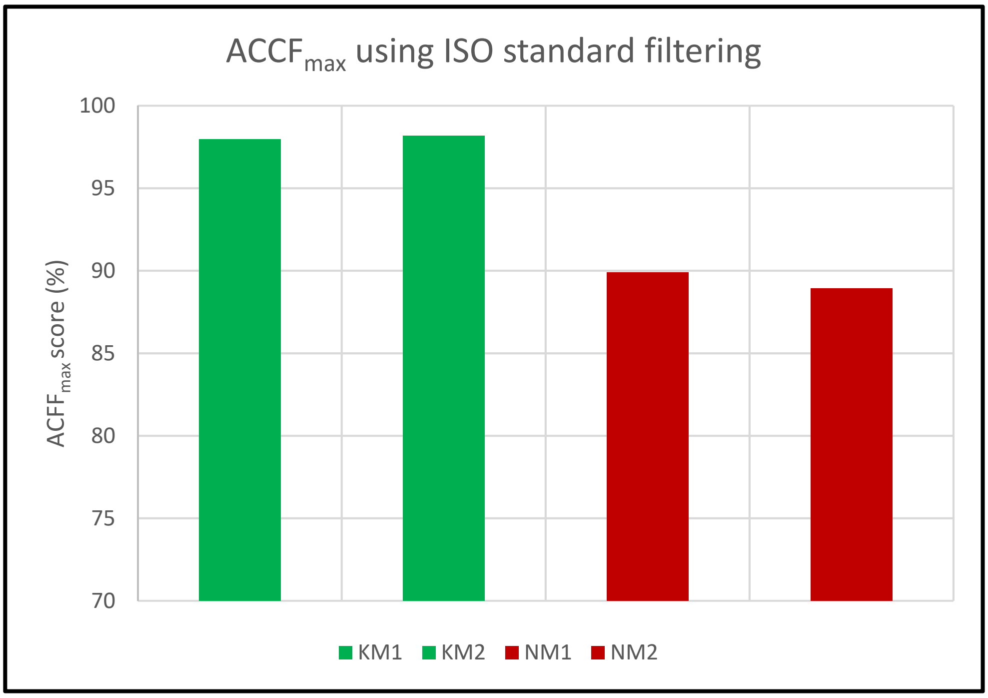


Figure 4-3: Graph of example correlation results using ISO standard filtering and least squares levelling, where KM= known match and NM= non-match

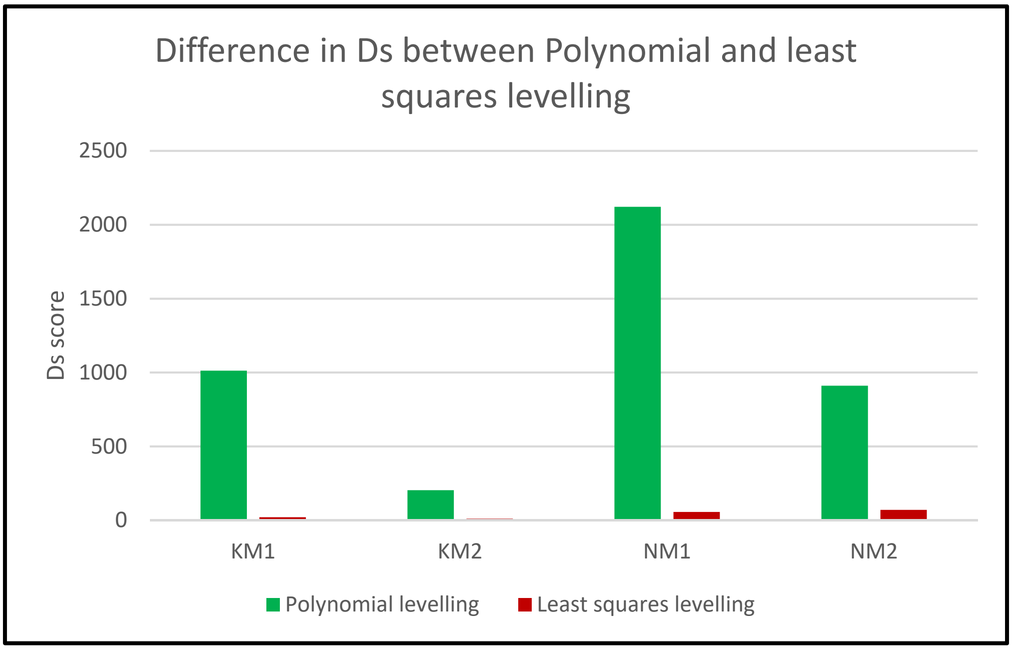


Figure 4-4: Graph of differences in D_s between polynomial and least squares levelling techniques

Of all the filtering methods tested, it was found that using a least squares levelling technique followed by a Robust Gaussian filter with cut off values of 75µm and 450µm would result in the most accurate correlation results, with known matches giving a high $ACCF_{max}$ percentage and low D_s value, as shown in Appendix 3:2, and non-matches having consistently lower $ACCF_{max}$ and high D_s results (Figure 4-3 and Figure 4-4). The average known match $ACCF_{max}$ result was found to be 71.71%, with an average D_s score of 9.10. Throughout testing of the Odyssey collection using each test object to correlate against the rest of the database, it was found that reasonable known match results would give a D_s score of below 100 (Table 4-3). As this trend continued throughout all testing it was decided that using a D_s score of over 100 to disregard matches would be an acceptable tool to minimise resulting hitlists.

Table 4-3: Example of correlation results using optimum pre-processing techniques

	Known match 1	Known match 2	Non-match 1	Non-match 2
$ACCF_{max}$ (%)	90.9141	88.9252	69.9227	35.2683
D_s	1.9687	5.09	100.4064	3055.935

The graphs below (Figure 4-5-Figure 4-7) show examples of the distribution of $ACCF_{max}$ scores when one test object is correlated against all other (389) firing pin impressions within the Odyssey collection using the optimum filtration methods. In each case, the two known matches gave the two highest $ACCF_{max}$ results. It can be seen that there is a common distribution in the $ACCF_{max}$ percentage results, with few results giving a large percentage match:

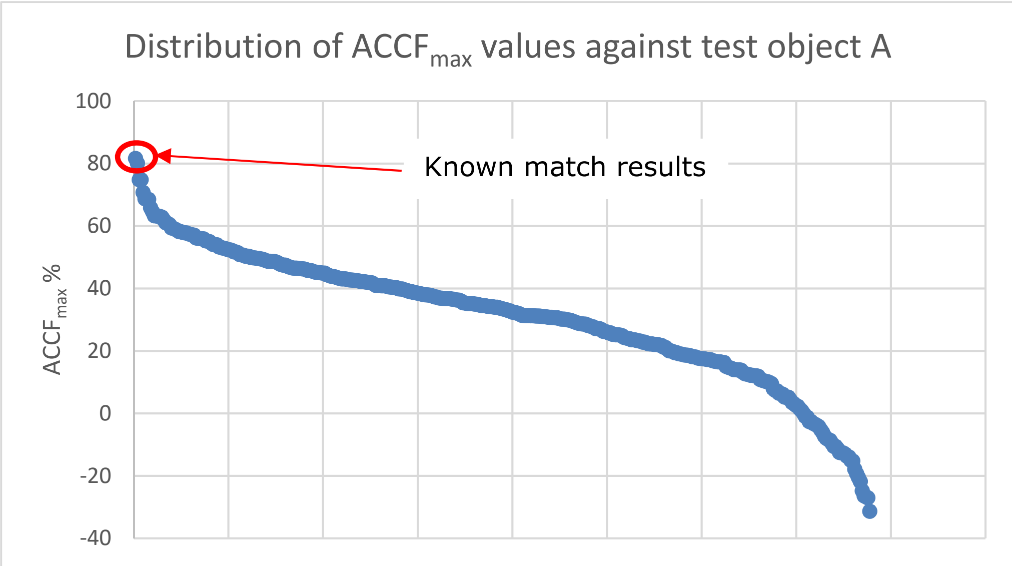


Figure 4-5: Graph of $ACCF_{max}$ percentage distributions when comparing test object A to all other firing pin impressions in the Odyssey collection

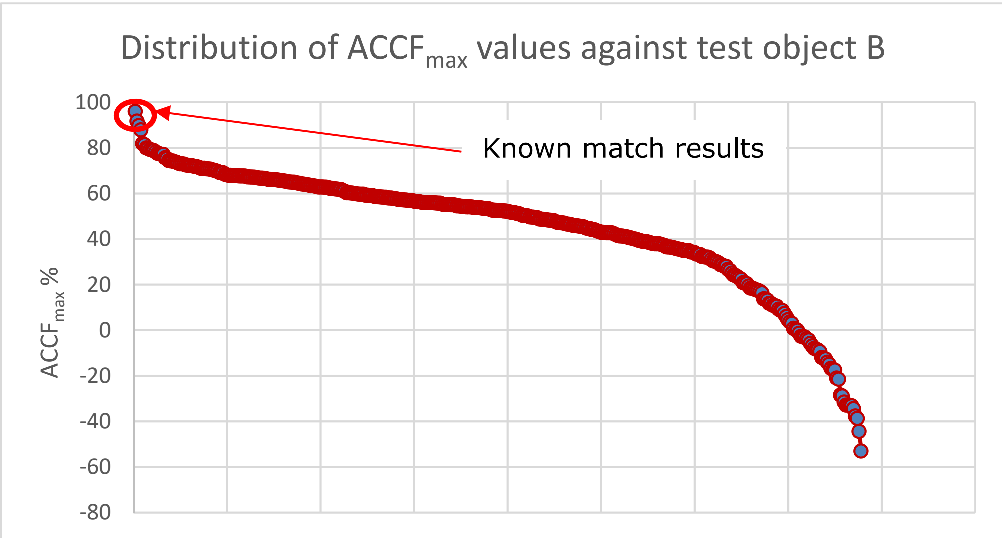


Figure 4-6: Graph of $ACCF_{max}$ percentage distributions when comparing test object B to all other firing pin impressions in the Odyssey collection

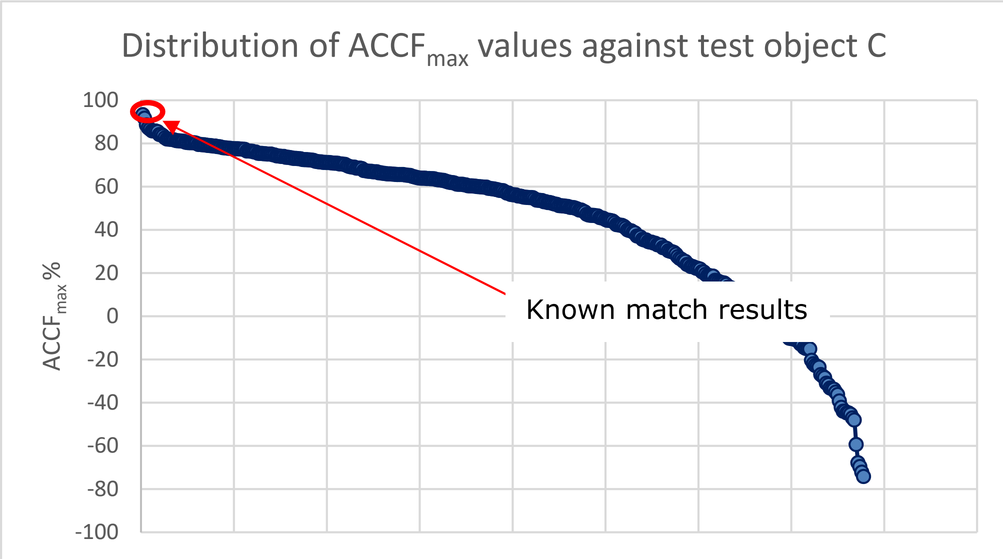


Figure 4-7: Graph of $ACCF_{max}$ percentage distributions when comparing test object C to all other firing pin impressions in the Odyssey collection

$ACCF_{max}$ values gained that were below zero indicate negative matching of some surface characteristics. This occurs when the overall shape of features is similar, but where the feature would appear as a peak on one surface, it is a trough on the other. There are also instances in which $ACCF_{max}$ values in non-matches are high. In cases where the D_s value is also low, this indicated a similarity in characteristics that have been created by different tools, i.e. the machining method used in machining the firearms may be very similar (Heard, 2013). Some cases of high $ACCF_{max}$ results in non -matches where the also resulted in a D_s value would indicate that the match should be disregarded due to large scale differences, where score over 100, which occurs when the overall shape of surface characteristics is similar, however the height scale difference of the peak/trough will have such a large variation that it becomes unlikely the impression was created by the same tool. It was also found through the preliminary testing, that using the above pre-processing techniques would result in very minor difference in D_s between known matches and non-matches. It was found that nearly all D_s scores would be below 100. Therefore, in this case D_s scores, while monitored throughout the testing, were not able to discriminate between known matches and non-matches. However, this suggests that the Robust Gaussian filtering and least squares levelling being used is not resulting in any waviness/form components either being retained from the original surface or introduced through incorrect levelling of the surface.

4.3.1 Transition to polar coordinates

The firing pin impressions of cartridge cases in the Odyssey collection are all round in shape. Due to the circular form of the firing pin impression, issues in correlation can be created through the rotational variance between two firing pin impressions. For instance, if a toolmark present in two surfaces is rotated 90° between the surfaces, effective correlation will not be achieved while the rotational variance exists (Senin, Groppetti, Garofano, Fratini, & Pierni, 2006).

There are three possibilities for the eradication of rotational variance within the toolmark. The first relies on user input at the data acquisition stage, in which cartridge cases must be aligned using other impressions on the cartridge case, for example an extractor mark, before a measurement takes place. During a demonstration of both the FTI and Balscan instruments, the author was advised that the identification systems would rely on such manual alignment for effective correlation of the toolmarks in these systems. While an expert user would be able to minimise rotational issues using this technique, there are various issues involved with regards to a truly automated/subjective method. The technique still has some reliance on the user to ensure the efficient correlation of the toolmarks, and the method relies upon the expertise and knowledge to be able to properly discriminate between the extractor marks and other marks present on the cartridge case. Therefore, this measurement technique can be considered to include some reliance on the user, which in turn introduces subjectivity to the technique, and therefore should be avoided.

It is also possible to remove rotational variance during the $ACCF_{max}$ correlation. In $ACCF_{max}$ correlation one surface is shifted with regards to the other to ascertain the best possible match between the two surfaces, however this usually entails shifting in only the x and y axes, until the $ACCF_{max}$ has been discovered. Where there is rotational variance between two datasets, it is also possible to introduce a rotational shift about one dataset in relation to the other. In this case $ACCF_{max}$ will be calculated using shifts in the x and y axes along with a rotational shift. While this method does not introduce any subjectivity into the technique, unlike manual alignment as mentioned previously, the computational cost of the correlation is significantly increased. As such, compensating for rotational variance using a rotational shift results in the time needed for correlation significantly increasing. Should this technique be used in a forensic context, it may be found to be too time consuming to be beneficial to the laboratory. As commercially available systems do require training for acquisition of cartridge cases with minimal rotational variance between samples, it may be argued that such a time-consuming technique is not needed. However, in a truly objective technique the possibility for measurement variance should not exist.

The final solution is to be able to translate the rotationally variable datasets into a format that removes the variance. This can be accomplished by translating the original surface topography measured in Cartesian coordinates into a polar coordinate scheme. Firstly, the method of correlation based in Cartesian space is shown in Figure 4-8:

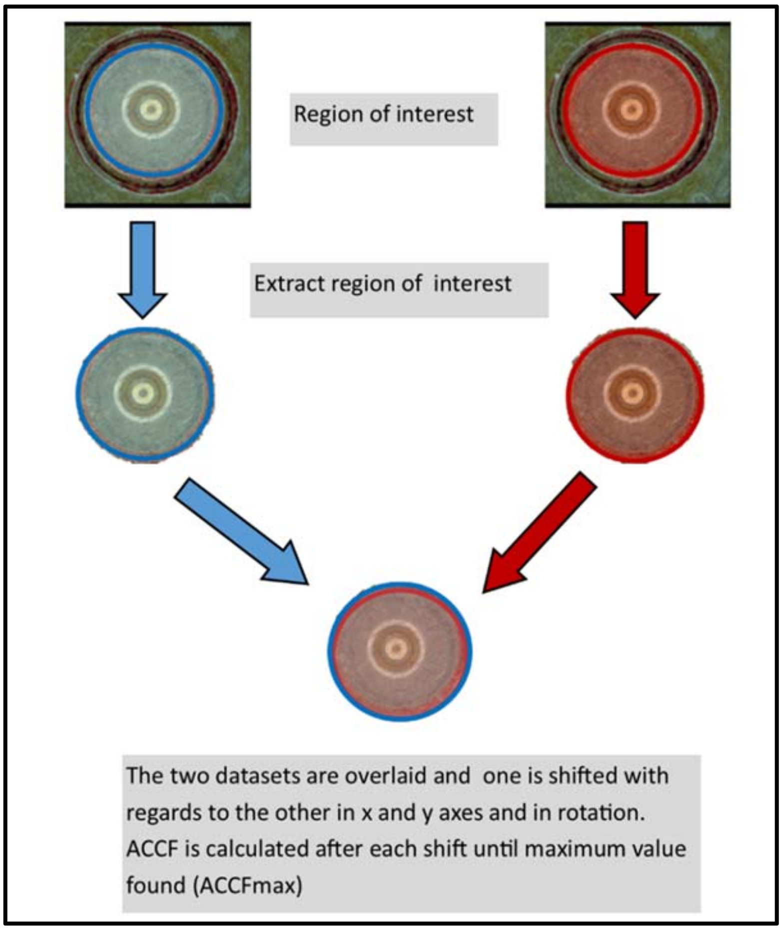


Figure 4-8: Flowchart of correlation in Cartesian format

In this study the circular form of cartridge case surfaces has been removed by translating datasets from Cartesian coordinates into polar coordinates. As polar coordinates describe a point by its distance and angle from coordinates (0,0), a circular dataset will be translated into a rectangular form. When shifting a rectangular dataset with regards to another, there are no rotational issues and shifting only occurs in the x and y axes, thus making correlation computationally more efficient.

Described in Figure 4-8 is the use of datasets in a traditional Cartesian format, however translating into polar coordinates results in the following flowchart, Figure 4-9:

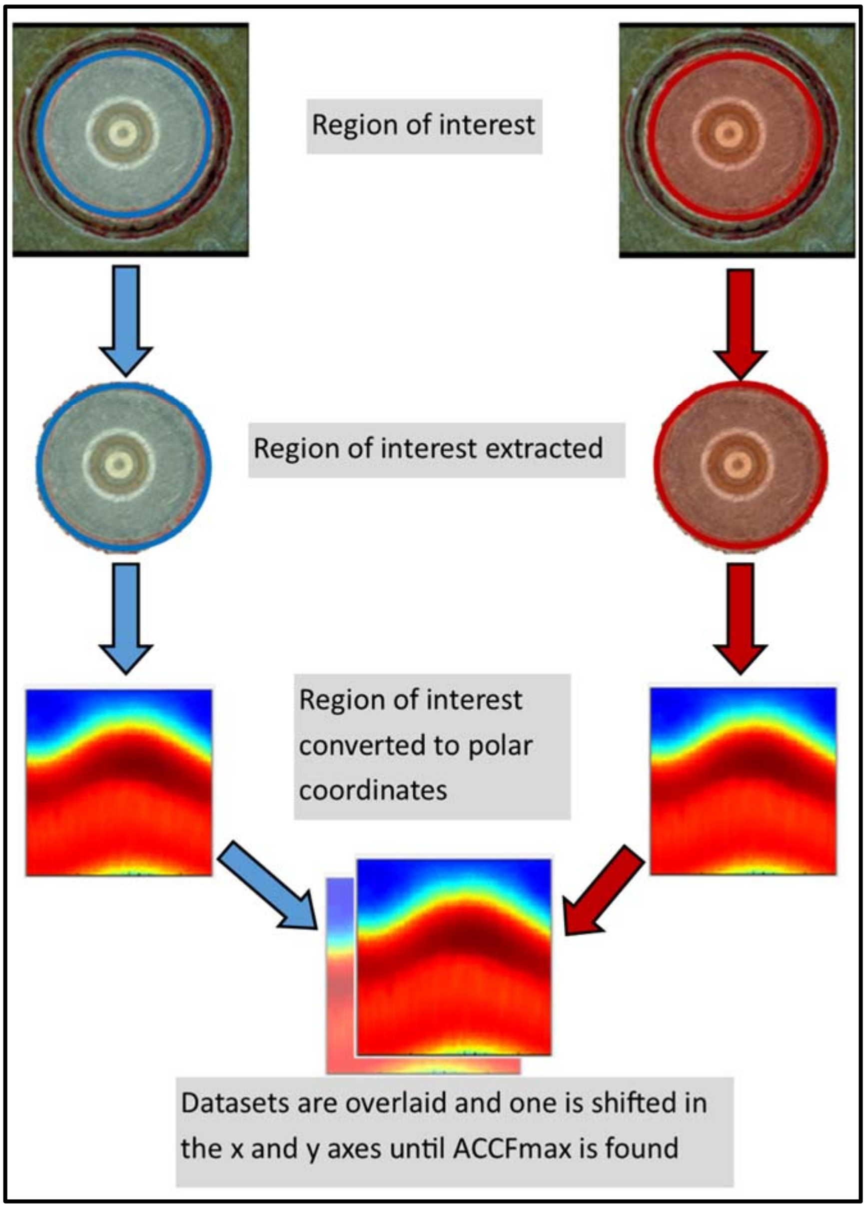


Figure 4-9: Flowchart of correlation using polar coordinates

In preliminary testing, it was found that using datasets in polar coordinates rather than Cartesian had two advantages. Firstly, the decrease in computational cost for the determination of $ACCF_{max}$ resulted in a decrease in correlation time of approximately five minutes per correlation (based on the author's workstation). Secondly, the increased efficiency in correlation with the removal of rotational variance resulted in a more efficient $ACCF_{max}$ correlation. Removing rotational variance resulted in a larger difference between known matches and non-matches, and increased the percentage match score between the known matches, as the eradication of rotational variance resulted in the correlation becoming more accurate. A minimum difference between known matches and non-matches using Cartesian coordinate $ACCF_{max}$ correlation was found to be 2.24%, whereas using polar

coordinates this was increased to 9.03%. As the polar transform relies on the correct positioning of a central pivot point, an additional shift for correlation was introduced. As such, the $ACCF_{max}$ is calculated multiple times as the pivot point is shifted pixel by pixel, and the maximum correlation is recorded. Due to this additional automated shift, versus manual positioning used in the Cartesian coordinate system, correlation has become higher in both known match and non-match. However, the increased differentiation between known matches and non-matches indicates the technique is valid.
Figure 4-10 and Figure 4-11 below show the difference between polar and Cartesian coordinate correlation when using the ISO standardised filtration techniques. Each column represents one correlation, therefore a total of four correlations are shown:

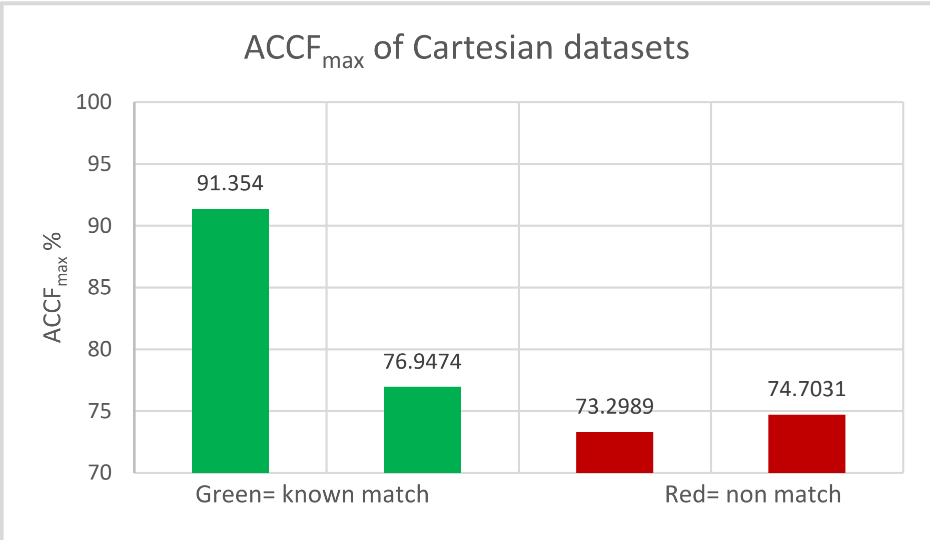


Figure 4-10: $ACCF_{max}$ correlation using Cartesian datasets

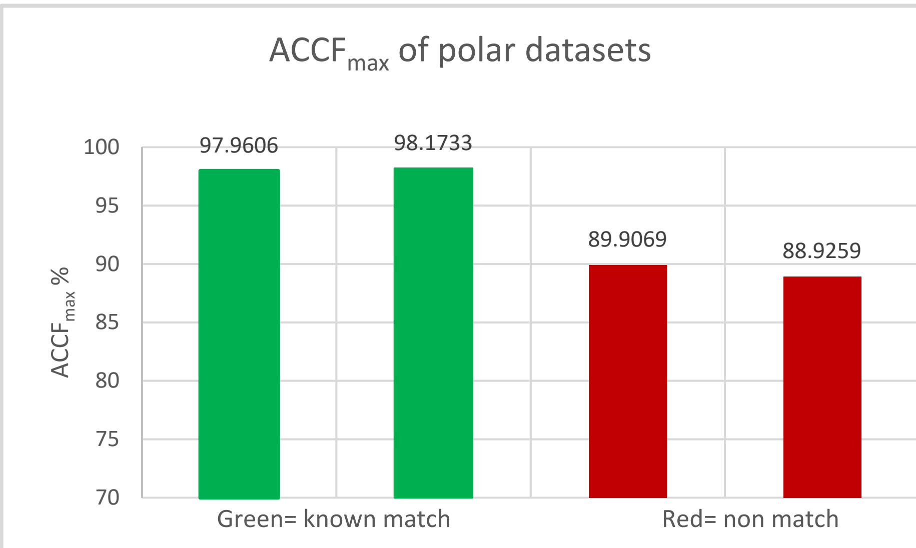


Figure 4-11: $ACCF_{max}$ correlation using Polar datasets

The overall increase of ACCFmax correlation results in non-matches using polar coordinates is due to an introduction of a slight shift

4.4 Overview of full correlation technique- firing pin impression

The pre-processing techniques have been determined in the preliminary studies using the Alias measurement system, and therefore for the further studies the following technique will be used (Figure 4-12):

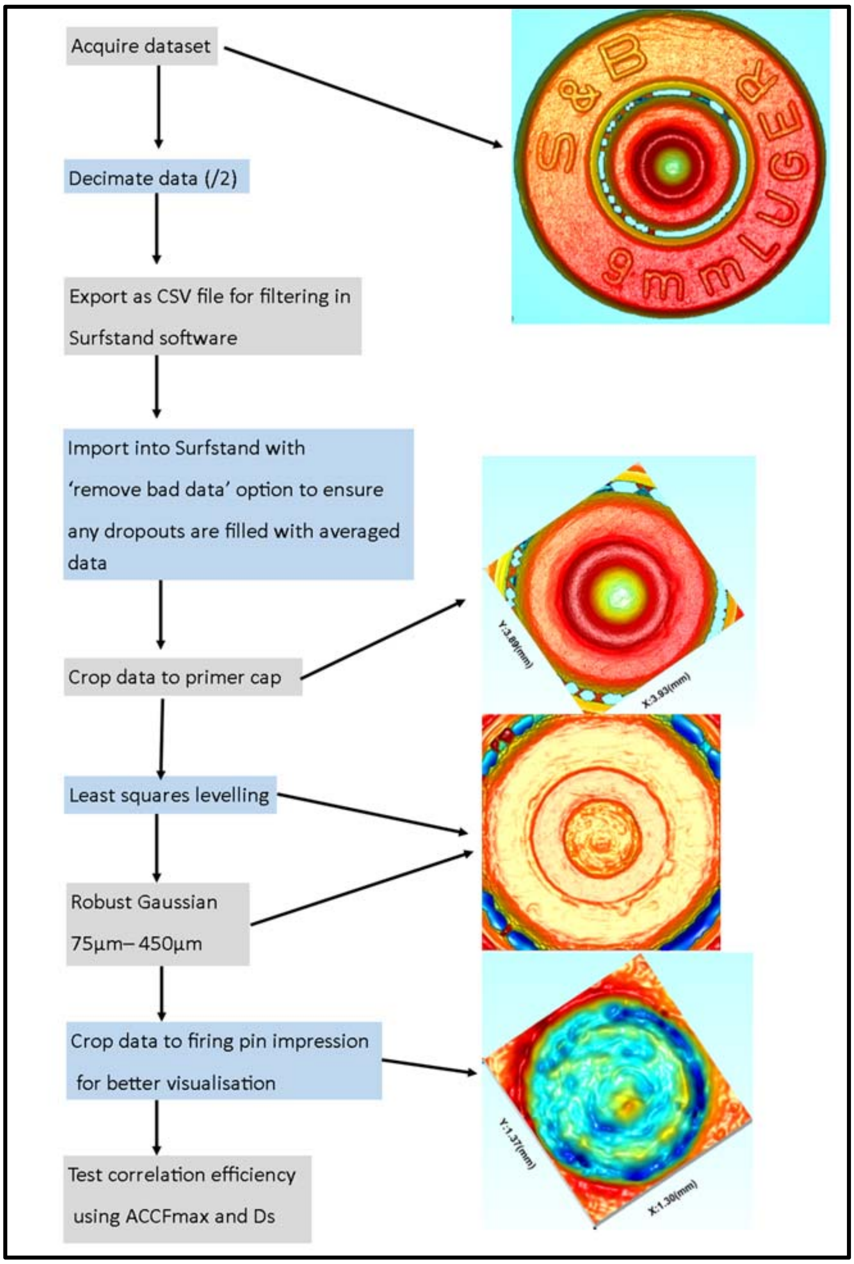


Figure 4-12: Flow of pre-processing techniques to be used in further studies

4.4.1 Firing pin correlation: Odyssey results

Having determined the pre-processing and correlation techniques to be used, a full study of the Odyssey collection can be completed, using the same method as in previous studies to ensure that a direct comparison of the efficiency all of measurement systems could be achieved. Previous studies included the use of the Evofinder and Arsenal 2D identification systems, and where necessary the results of these studies will be discussed anomalously, using the terms "system A" and "system B" for reasons of commercial confidentiality.

The 22 test objects are correlated against all other objects within the Odyssey collection. This creates a total number of 389 correlations per test object, two of which were known matches. The total number of correlations created is 389x22= 8,558. The study is completed twice, once using datasets acquired using the Alias system, and once with Alicona datasets, and therefore a total of 17,116 correlations of firing pin impressions were created.

The correlation results for each test object were firstly checked to ensure the D_s values were acceptable i.e.<100 D_s score, and any correlations with an abnormally high D_s value were not included in correlation hitlists. Acceptable correlations were then ordered from highest to lowest to create the hitlist.

At this stage of the research, the correlation of the firing pin impressions is not currently fully automated. As such, the author had to manually determine the area of firing pin impression to be correlated. MATLAB code was created that would allow the author to choose the area of interest, and the code would then translate this to polar coordinates before correlating the two surfaces. To minimise the effect of manual determination of area, the area is also slightly shifted in increments of one data point in both the x and y directions in one dataset to ensure the same area of interest is being correlated between the two surfaces, which is defined by the highest $ACCF_{max}$ percentage value (Figure 4-13). This effectively minimises any discrepancy in measurement position and rotational variance that may have been introduced in the manual set up of measurement. This results in a robust measurement technique, thus satisfying the Daubert standard (Page, Taylor, & Blenkin, 2011).

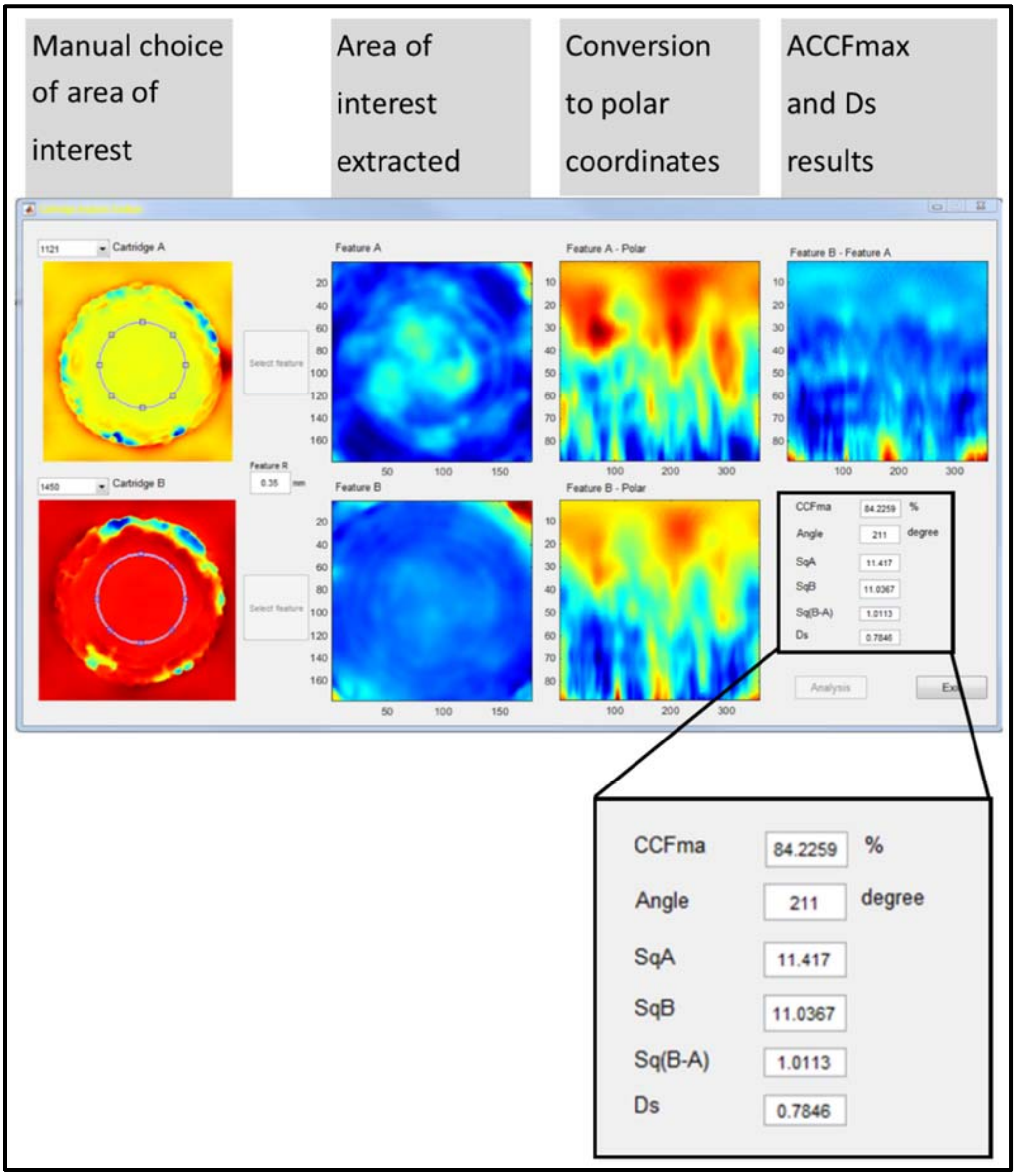


Figure 4-13: MATLAB code used for $ACCF_{max}$ determination.

The results of the correlation for each system are as follows.

4.5 Discussion of results

The following results show the difference in correlation efficacy comparing the various methods used within this study. Table 4-4 shows the hitlist results for both known matches using each method:

Table 4-4: Firing pin correlation hitlist results using each method

	Alias with ACCFmax		Alias with alias correlation		Alicona with ACCFmax	
Cartridge case test object	Known match B position	Known match C position	Known match B position	Known match C position	Known match B position	Known match C position
1A	1	100	162	366	3	109
2A	1	372	365	198	95	289
3A	1	37	20	193	215	1
4A	1	9	8	5	10	5
5A	306	7	125	129	36	122
6A	9	110	19	69	307	337
7A	1	9	20	3	4	13
8A	1	5	168	60	17	1
9A	2	1	113	106	1	190
10A	2	1	278	176	1	No match
11A	1	No match, damaged cc[6]	70	46	218	No match
12A	1	2	14	387	19	112
13A	1	313	7	93	2	235
14A	1	3	176	370	1	30
15A	1	3	243	385	1	2
16A	1	6	No match: did not appear in list of possible matches	66	1	4
17A	1	2	13	177	11	159
18A	1	232	140	84	227	No match

[6] It was found that the firing pin impression of cartridge case 11C had been erased due to the firing pin piercing the primer cap.

19A	3	2	67	96	21	10
20A	2	1	182	342	10	1
21A	1	33	385	135	3	4
22A	1	11	63	211	6	73

It can be seen from the above results that there are inconsistencies in hitlist results between the different methodologies used, with the better results being found using Alias datasets combined with $ACCF_{max}$ correlation and worst correlation results being seen using Alias datasets combined with the correlation algorithms used in the Alias software. Table 4-5 shows the number of known matches placed within hitlists of varying lengths, which corroborates these findings. This is further detailed in Figure 4-14 below, in which a comparative graph gives a comparison of percentage success rate for the presence of known matches within varying hitlist lengths.

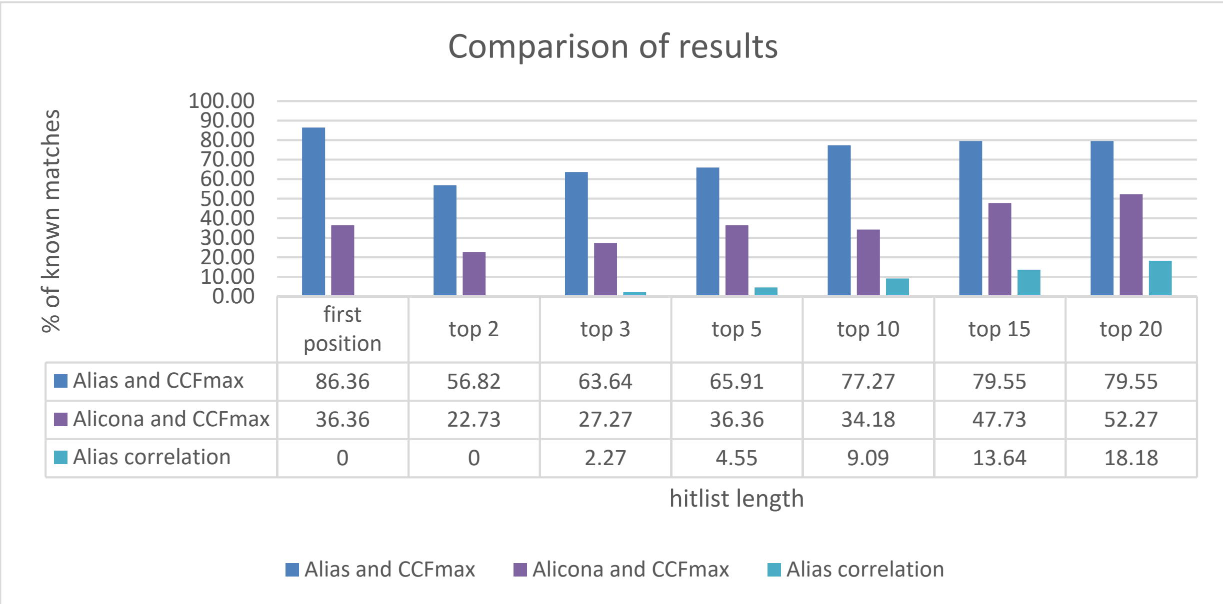


Figure 4-14: Comparison of hitlist results in varying methods

Table 4-5: Known match position results

	First position (of 22)	Top 2 (of 44)	Top 3 (of 44)	Top 5 (of 44)	Top 10 (of 44)	Top 15 (of 44)	Top 20 (of 44)
Alias and $ACCF_{max}$	19	25	28	29	34	35	35
Alicona and CCF max	8	12	12	15	16	21	23
Alias correlation	0	0	1	2	4	6	8

As shown in the graph of correlation efficiency, the largest discrepancy in results can be seen between correlating datasets acquired using the Alias system, but using different algorithms for correlation. In using $ACCF_{max}$ correlation rather than the Alpine/Dead Sea correlation algorithms used in the Alias software, there is a significant increase in efficiency. As the exact same datasets were used in correlation, the difference can be attributed only to the algorithms used. As the Alpine/Dead sea comparator relies upon comparing a small number (5) of the highest and lowest points upon the surface, it can be shown that this method does not utilise enough information regarding the topography of individual characteristics within the toolmarks. Using the Alias correlation method, it was found that no known matches would be placed within the top two best matches, and only 18% of the 46 known matches were present in the top 20 of the hitlists. Therefore, there is a much larger chance of a non-match giving the better correlation results than the known matches, and as a result there are too many false positives using this method for it to be considered a forensic technique at this time.

Using $ACCF_{max}$ correlation results in an increase in correlation efficiency, resulting in 79.6% of the 44 known matches being placed within the top 20 best matches, and 86% of a possible 22 known matches were placed in the first position. Therefore, while the correlation efficiency has increased, there are still some instances in which known matches are not present within the top 20 of the hitlist. It is expected that as the Odyssey collection is comprised of cartridges manufactured by various different companies, a material difference in the primer caps may influence the topography of toolmarks. For instance, a harder material will not deform as much as a softer one when brought into contact with a tool. A discussion on the effect of cartridge material will follow in a later chapter.

As it can be shown that $ACCF_{max}$ will result in good correlation efficiency, the poor correlation results gained from Alicona acquired datasets using $ACCF_{max}$ indicate that the measurement technique itself is resulting in poor correlation. As the resolution of measurement techniques were kept as similar as possible, this difference in correlation efficiency is a direct result of the measurement technique, namely measurement artefacts present in datasets acquired by the Alicona system. As the Alicona relies on the calculation of surface height based on neighbouring points on the surface, the averaging can result in waviness being imparted as a measurement artefact (Leach, 2010). This is further supported in the fact that correlation completed using Alicona datasets resulted in a significant increase in results that were found to have too large a D_s value to be considered a true match. Table 4-6 and Figure 4-15 below show how often a D_s value of over 100 will occur in both Alias and Alicona acquired datasets:

Table 4-6: Comparison of D_s values between Alias and Alicona acquired datasets.

Test object	Alicona datasets Correlations with D_s over 100	Alias datasets Correlations with D_s over 100
1A	24	18
2A	22	23
3A	18	7
4A	367	366
5A	20	8
6A	8	9
7A	360	371
8A	28	12
9A	14	10
10A	30	12
11A	29	7
12A	20	20
13A	16	9
14A	30	28
15A	16	30
16A	52	25
17A	45	27
18A	23	11
19A	338	187
20A	59	23
21A	46	9
22A	79	48
Total	1,644 of 8,558 correlations	1,260 of 8,558 correlations

As can be seen in Table 4-6 and Figure 4-15, the general trend is that in each hitlist for a test object, more cartridge cases will be excluded for correlation using the Alicona measurement system than the Alias. As a higher D_s value can be attributed to larger scale differences between two surfaces, this implies that the Alicona measurement method imparts larger scale differences into the dataset, which consequently would make the technique less accurate, as seen in hitlist results. In other published results using $ACCF_{max}$ correlation, regardless of measurement technique, D_s values are always found to be under 100. In cartridge case correlation it is typical to find D_s scores under 20, and bullets under 100 (Cadevall & Schwarz, 2013; Song, 2015). Therefore, the author chose to disregard any D_s value over 100 to provide a unified approach for any evidence, which was corroborated by results gained in this study. There are three hitlists, those of cartridge cases 4A, 7A and 19A, where the majority of correlations have resulted in a D_s value too high (i.e. over 100) to be considered a true match. In these cases, the particular toolmark of the test object and its two matches are originally significantly different, as both Alicona and Alias hitlists have been significantly reduced. Due to the nature of the Odyssey collection, toolmarks could be affected by a number of variables including material differences, differences in primer cap seating depth/angle and variations in cartridge manufacturer.

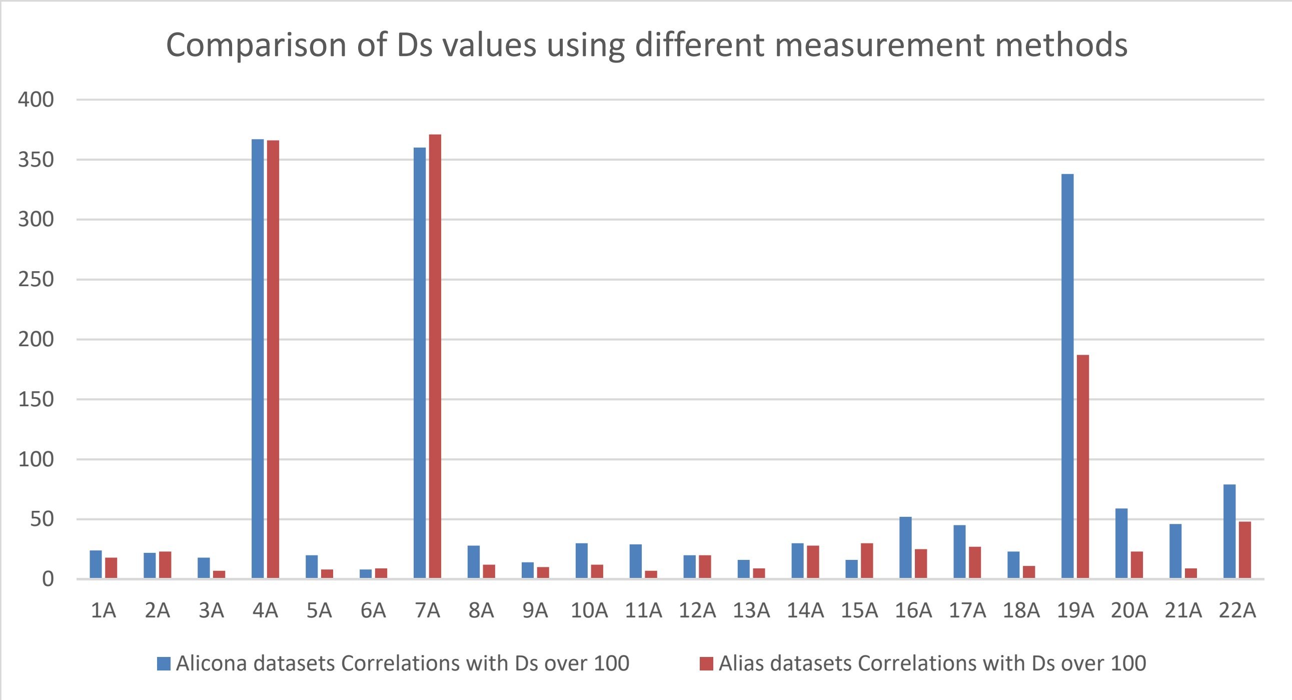


Figure 4-15: Comparative graph of Ds values using different measurement methods

4.6 Comparison to previous studies

In previous studies (Thomas J. , 2011), two commercially available 2D comparison systems were used to correlate the 22 test objects to all other cartridge cases within the Odyssey collection. In this case, each system gave one 'match score', which were ranked best to worst within the system software to create a hitlist. For commercial confidentiality, the two systems will be referred to as systems A and B. The hitlist results from the systems are as follows in Table 4-7:

Table 4-7: Correlation efficacy in 2D systems

	System A		System B	
Cartridge case number	KM B position	KM C position	KM B position	KM C position
1A	1	183	1	192
2A	3	57	1	166
3A	1	176	1	60
4A	9	1	4	1
5A	98	109	257	212
6A	1	6	1	2
7A	35	23	14	18
8A	74	3	1	48
9A	180	1	3	1
10A	49	1	13	1
11A	18	321	1	335
12A	1	315	1	345
13A	1	221	1	8
14A	1	287	1	359
15A	1	199	1	150
16A	1	2	1	2
17A	2	1	1	10
18A	1	2	8	11
19A	67	41	63	66
20A	1	2	1	2
21A	7	11	2	6
22A	12	82	32	176

The table above shows that the best efficiency in the two previous commercial systems is found in system B, which is further illustrated in the graph below (Figure 4-16):

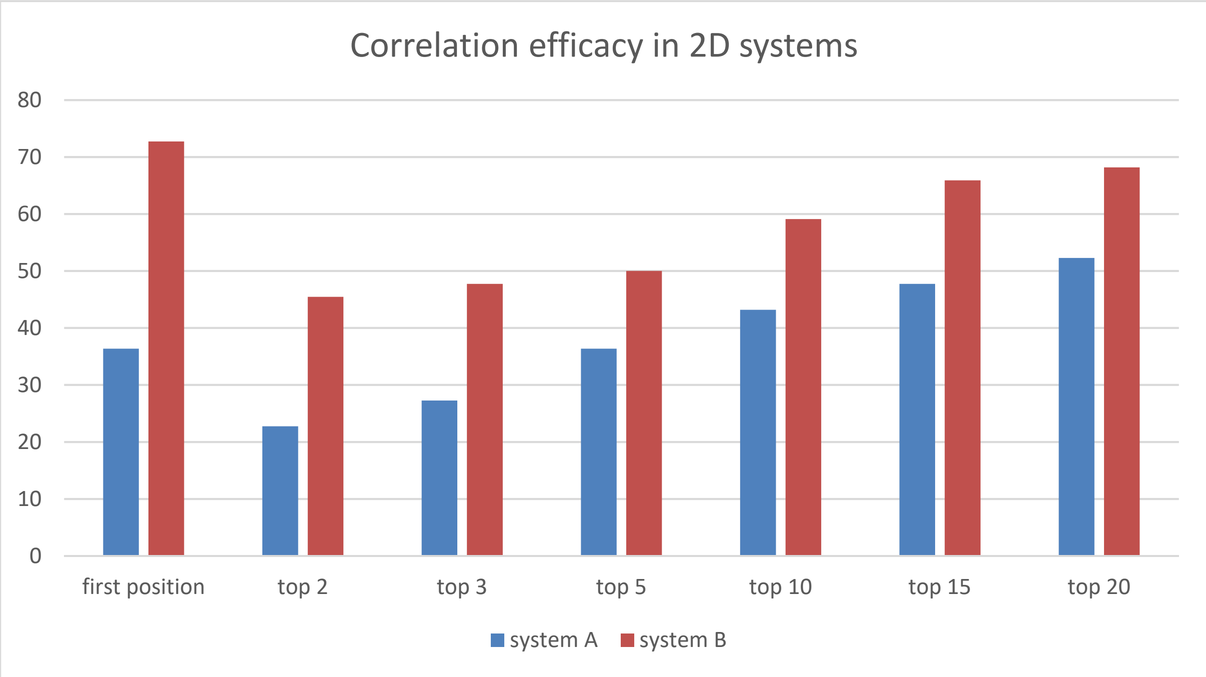


Figure 4-16: Correlation efficacy in 2D systems

In the previous Odyssey project, two commercially available 2D systems were used to acquire images of the Odyssey collection. This was accomplished by professional forensic examiners, who had previously been trained in the use of these instruments. In the acquisition of images, rotational variance is reduced by the examiner by using ejector marks as a registration tool, to ensure cartridge cases are measured in the same spatial location (Thomas J. , 2011).

As both systems used in the previous study relied upon the 2D imaging of the toolmarks, the discrepancy in correlation must originate in the pre-processing and correlation of the datasets, or variation in lighting of the surface. As is it possible that the two systems did rely on differing measurement techniques, it could be the case the variation was also introduced in the measurement quality. The results highlighted that the different systems would use different correlation and measurement techniques, and therefore interoperability between the two systems would be unlikely.

In comparing the results from studies conducted by the author and those in previous studies (Thomas J. , 2011; Yates, Akghar, Bates, Jopek, & Wilson, 2011), it can be seen that there are several factors in the accurate correlation of firing pin impressions, which must be taken into account for effective correlation.

The order of methods, from best to worse correlation results, is as follows:

- Alias acquired datasets with $ACCF_{max}$ correlation
- System B (previous study)
- System A (previous study)

- Alicona acquired datasets with $ACCF_{max}$ correlation
- Alias acquired datasets with Alias correlation software

The correlation methods are listed with regards to the percentage of known matches which are ranked within the top 20 of the hitlist. The most striking results can be seen in Alias acquired datasets. Using $ACCF_{max}$ correlation results in the highest correlation efficiency between all the systems, and using Alias correlation software gives the worst correlation results. As the exact same datasets are used between these two methods, the results serve to prove that while the use of areal datasets for acquiring toolmark information can improve correlation, it is vital that pre-processing and correlation methods are able to correctly differentiate between salient and non-salient data within the topography of the dataset. Both 2D imaging systems were also found to have a greater correlation efficiency than the use of the Alicona to acquire areal topography of the surface. As it has been shown using the Alias results, that good correlation results can be achieved using $ACCF_{max}$ correlation, the decrease in efficiency using the Alicona can be shown to be a result of the measurement method. Further work to determine whether pre-processing techniques are optimised for use of Alicona measurements is needed to corroborate these results. The Alicona focus variation method does not result in data of high enough quality to gain better correlation results than those gained in 2D methods, due to measurement artefacts being introduced into the acquired dataset.

Therefore, while advanced measurement and mathematical correlation can result in a more efficient analysis system, this is only the case when the data is treated correctly. As advanced measurement systems result in a significant increase in data being acquired compared to 2D systems, the differentiation between salient and non-salient data plays a vital role in further correlation, as shown in and Figure 4-17 and Table 4-8:

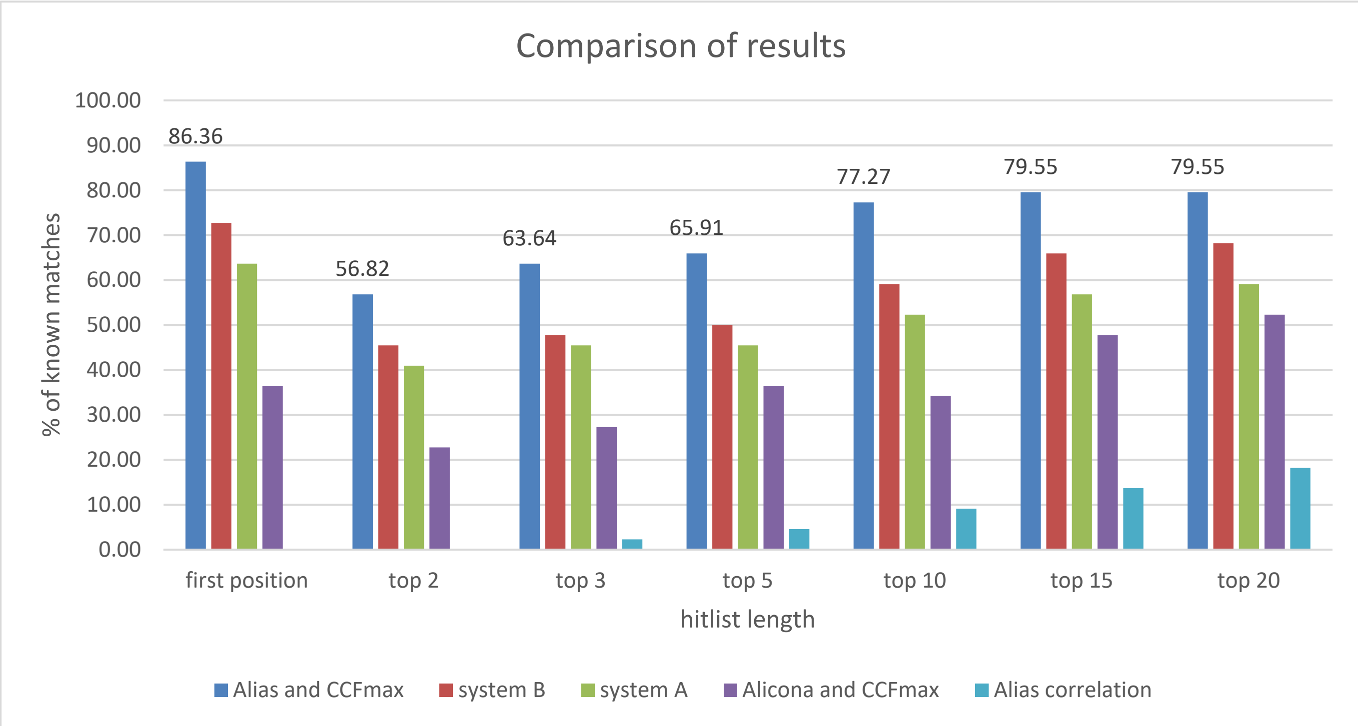


Figure 4-17: Comparative graph of correlation results across all methods

Table 4-8: Known match hitlist results across all methods

	First position (of 22)	Top 2 (of 44)	Top 3 (of 44)	Top 5 (of 44)	Top 10 (of 44)	Top 15 (of 44)	Top 20 (of 44)
Alias and $ACCF_{max}$	19	25	28	29	34	35	35
System B	16	20	21	22	26	29	30
System A	14	18	20	20	23	25	26
Alicona and $ACCF_{max}$	8	12	12	15	16	21	23
Alias correlation	0	0	1	2	4	6	8

4.7 Material composition

In the correlation of both firing impressions and bullets of the Odyssey collection, some erroneous results were identified, in which a known match was placed too low in a hitlist of correlation results. In forensic evidence, these results would be considered false negatives, and could harm a prosecution hypothesis.

Identified erroneous results, where known matches were placed outside of the top 20, were in the minority. Thus, pre-processing methods were determined to not be causing erroneous results, as most known matches resulted in successful correlation. Therefore, it is expected that a variation between the physical topography of the imparted toolmark may be causing issues in the correlation.

The Odyssey collection is comprised of cartridges from eight different manufacturers. In the original study it was noted that variations in manufacturer should be kept to a minimum to increase repeatability in the topography of the impression. However, it was found that using one manufacturer of cartridge would become very difficult in the timescale available to create the impression, and therefore it was decided to accept that eight different manufacturers would have to be used to complete the Odyssey collection.

In this study the aim is to ascertain whether the difference in materials used for cartridge manufacturer will influence the overall topography of the impression. It is expected that as a softer material will exhibit a greater degree of deformation under contact from a tool compared to a harder material, that some variation in toolmark topography may be seen. Comparison of correlation results based on material composition will be made using Alias acquired datasets only, to ensure variations in measurement method do not affect the findings.

4.7.1 Methods

In this study, the test objects and their known matches were used for material and topographical analysis. The total number of cartridge cases was 66.

The surface material of each primer cap was analysed using X-Ray Fluorescence (XRF) with the Bruker Tracer IV-SD handheld XRF instrument, using software driven voltage and current control to ensure measurement settings were correct for the material being analysed. XRF is based upon the differences in excitation of electrons leaving the outer shells under X-ray energy to determine the element, as every element has a different excitation level of the outer electrons.

The topographical datasets were exported as a coordinate system into the SURFSTAND software created at the University of Huddersfield, which was used for the pre-processing and parameter calculations of the overall topography of the firing pin impressions.

To be able to assess the overall difference in the form of firing pin impressions, it was decided to calculate the following areal volume parameters (Table 4-9). Further definition of these parameters is available in Appendix 2.

Table 4-9: Table of volume parameters used in the study

Areal Parameter	Description
Vmp	Peak material volume
Vmc	Core material volume
Vvc	Core void volume
Vvv	Valley void volume

Using volume parameters will give an indication as to the overall deformation of the surface due to the impression of the firing pin into the primer cap. A low pass filter was not used to ensure that all form remained in the surface. This would ultimately result in a better understanding in the differences in toolmarks departed due to the material properties of the primer cap.

4.7.2 Results: Material analysis of primer caps

Table 4-10 shows the results of material analysis for each of the primer caps of test objects and their known matches. In each case, the percentage of the element given is expressed as a percentage of the surface area being analysed. The study focuses on three most prevalent elements used in cartridge manufacture, Copper, Nickel and Zinc as they contribute the majority of material percentage. Figure 4-18-Figure 4-20 graphically represent the material composition results.

Table 4-10: Material analysis of test objects within the Odyssey collection

Object number	Element	%	Object number	Element	%	Object number	Element	%
1A	Cu	65.3	1B	Cu	65.7	1C	Cu	70.7
	Zn	27.6		Zn	27		Zn	28.4
	Ni	6.2		Ni	6.42		Ni	0
2A	Cu	68.4	2B	Cu	30.6	2C	Cu	68
	Zn	30.7		Zn	0		Zn	28.4
	Ni	0		NI	68.5		Ni	2.79
3A	Cu	67.9	3B	Cu	68.4	3C	Cu	70.1
	Zn	28.8		Zn	28.7		Zn	28.9
	Ni	2.4		Ni	1.96		Ni	2.4
4A	Cu	71.3	4B	Cu	71.5	4C	Cu	69.2
	Zn	27.2		Zn	27.5		Zn	30
	Ni	0.63		Ni	0		Ni	0
5A	Cu	71.1	5B	Cu	68.8	5C	Cu	71
	Zn	27.9		Zn	30.2		Zn	27.2
	Ni	0		Ni	0		Ni	0.95
6A	Cu	70.9	6B	Cu	69	6C	Cu	70.4
	Zn	27		Zn	30		Zn	28.7
	Ni	1.22		Ni	0		Ni	0
7A	Cu	70.7	7B	Cu	68.8	7C	Cu	71.4
	Zn	28.3		Zn	30.3		Zn	27.3
	Ni	0		Ni	0		Ni	0.38
8A	Cu	70.3	8B	Cu	69.5	8C	Cu	71.2
	Zn	28.7		Zn	29.6		Zn	28.7
	Ni	0		Ni	0		Ni	0.6
9A	Cu	70.8	9B	Cu	70.2	9C	Cu	69.9
	Zn	27		Zn	28.8		Zn	29.1
	Ni	1.37		Ni	0		Ni	0
10A	Cu	69.4	10B	Cu	71	10C	Cu	70.4
	Zn	29.5		Zn	28		Zn	27.1
	Ni	0		Ni	0		Ni	1.68
11A	Cu	71	11B	Cu	69	11C	Cu	70.3
	Zn	27.3		Zn	30		Zn	28.6
	Ni	0.77		Ni	0		Ni	0
12A	Cu	66	12B	Cu	66.8	12C	Cu	70.3
	Zn	28		Zn	27.4		Zn	28.7
	Ni	5.05		Ni	4.94		Ni	0
13A	Cu	66.3	13B	Cu	63.7	13C	Cu	70.6
	Zn	27.5		Zn	27		Zn	28.4
	Ni	5.31		Ni	8.48		Ni	0
14A	Cu	64.5	14B	Cu	65.7	14C	Cu	70.3
	Zn	27.3		Zn	27.8		Zn	28.7
	Ni	7.49		NI	5.66		Ni	0
15A	Cu	65.2	15B	Cu	66.5	15C	Cu	70.4
	Zn	27.5		Zn	28.2		Zn	28.7
	Ni	6.39		Ni	4.47		Ni	6.39

16A	Cu	74.3	16B	Cu	72.6	16C	Cu	72
	Zn	25.22		Zn	26.5		Zn	27.1
	Ni	0		Ni	0		Ni	0
17A	Cu	71	17B	Cu	70.3	17C	Cu	70
	Zn	28.2		Zn	28.8		Zn	29.1
	Ni	0		Ni	0		Ni	0
18A	Cu	72.1	18B	Cu	72.8	18C	Cu	66.8
	Zn	26.9		Zn	26.4		Zn	27
	Ni	0		Ni	0		Ni	5.45
19A	Cu	72.5	19B	Cu	72.7	19C	Cu	64.6
	Zn	26.6		Zn	26.5		Zn	26.6
	Ni	0		Ni	0		Ni	7.92
20A	Cu	67.7	20B	Cu	64.8	20C	Cu	69.4
	Zn	29.6		Zn	28.8		Zn	29.7
	Ni	1.77		Ni	5.43		Ni	0
21A	Cu	69.6	21B	Cu	69.8	21C	Cu	68
	Zn	29.5		Zn	29.4		Zn	28.2
	Ni	0		Ni	0		Ni	2.93
22A	Cu	69.6	22B	Cu	69.6	22C	Cu	67.4
	Zn	29.4		Zn	29.4		Zn	28
	Ni	0		Ni	0		Ni	0.19

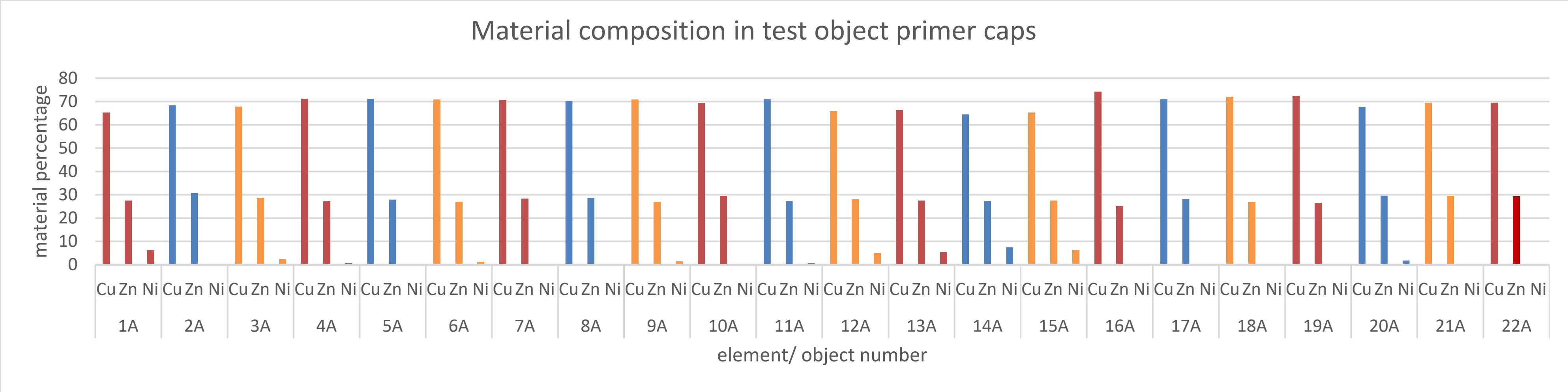


Figure 4-18: Material composition across test object (A) primer caps

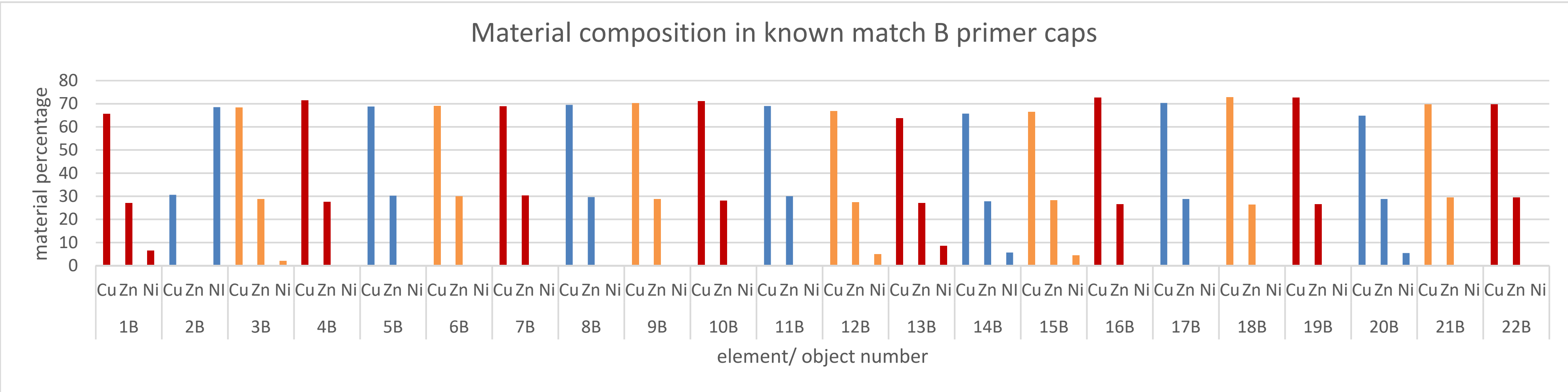


Figure 4-19: Material composition across known match B primer caps

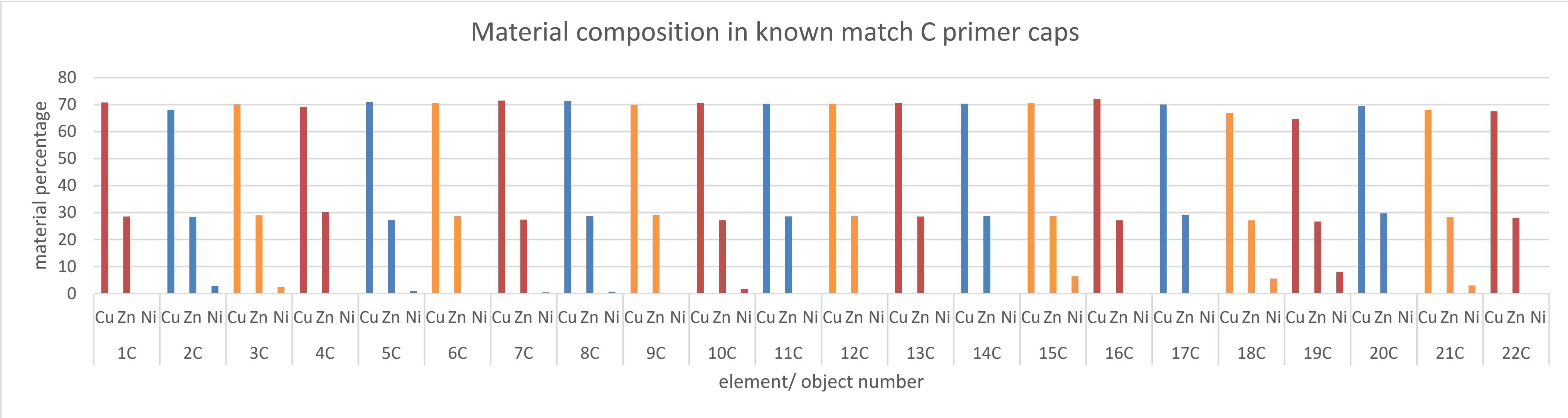


Figure 4-20: Material composition across known match C primer caps

In comparison of the elemental composition of the test object and known match primer caps, it can be seen that the elemental composition sometimes varies, but this is not true in all cases. To ascertain whether elemental composition does influence the correlation of imparted individual characteristics, correlation results will be compared in primer caps that have no change in composition and those that do. Table 4-11 below shows the difference in elemental composition and the $ACCF_{max}$ correlation result for each test object and its two known matches. Correlation was completed using the pre-processing methods as detailed in Chapter 4.4:

Table 4-11: Material composition difference and $ACCF_{max}$ value

Test object	Known match B $ACCF_{max}$ (%)	Difference in Cu (%)	Difference in Zn (%)	Difference in Ni (%)	Known match C $ACCF_{max}$ (%)	Difference in Cu (%)	Difference in Zn (%)	Difference in Ni (%)
1A	80.0374	0.4	0.6	0.22	80.5914	5.4	0.8	6.2
2A	76.4536	37.8	30.7	68.5	-17.6116	0.4	2.3	2.79
3A	95.5298	0.5	0.1	0.44	84.1454	2.2	0.1	0
4A	10.4123	0.2	0.3	0.63	70.7893	2.1	2.8	0.63
5A	22.4872	2.3	2.3	0	75.4182	0.1	0.7	0.95
6A	75.6504	1.9	3	1.22	35.3808	0.5	1.7	1.22
7A	56.5028	1.9	2	0	33.5901	0.7	1	0.38
8A	84.2626	0.8	0.9	0	76.5689	0.9	0	0.6
9A	91.6775	0.6	1.8	1.37	95.9432	0.9	2.1	1.37
10A	91.5566	1.6	1.5	0	93.1927	1	2.4	1.68
11A	91.6543	2	2.7	0.77	Damaged	0.7	1.3	0.77
12A	90.2047	0.8	0.6	0.11	86.5492	4.3	0.7	5.05
13A	81.0297	2.6	0.5	3.17	16.7658	4.3	0.9	5.31
14A	93.5607	1.2	0.5	1.83	78.4037	5.8	1.4	7.49
15A	90.7206	1.3	0.7	1.92	84.9622	5.2	1.2	0
16A	90.5817	1.7	1.28	0	70.521	2.3	1.88	0
17A	59.2021	0.7	0.6	0	54.023	1	0.9	0
18A	81.1848	0.7	0.5	0	46.8911	5.3	0.1	5.45
19A	67.517	0.2	0.1	0	69.2527	7.9	0	7.92
20A	80.0374	2.9	0.8	3.66	81.5914	1.7	0.1	1.77
21A	90.3661	0.2	0.1	0	83.1665	1.6	1.3	2.93
22A	89.9311	0	0	0	87.0085	2.2	1.4	0.19

In the graph below (Figure 4-21) the difference in elemental composition is plotted against $ACCF_{max}$ results gained.

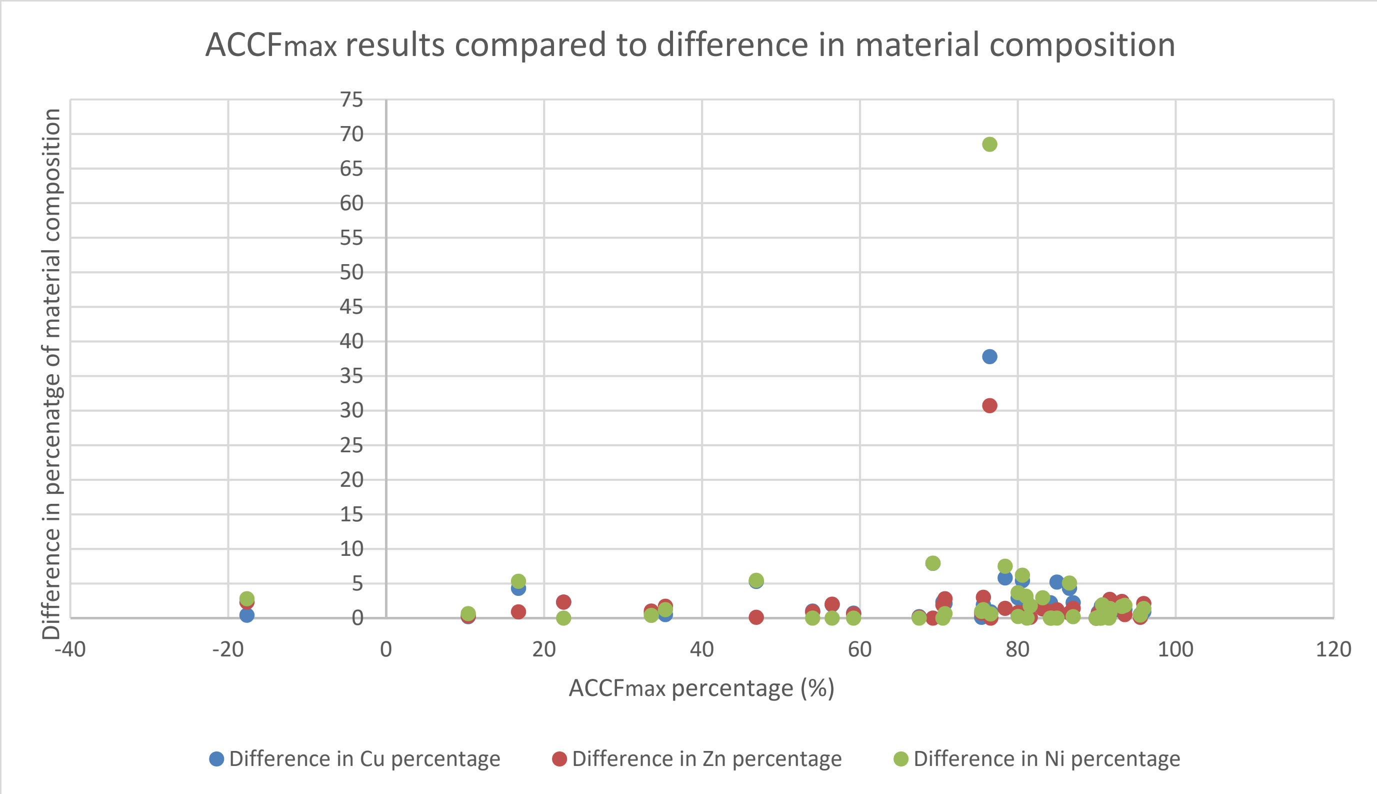


Figure 4-21: Graph of correlation against material composition

As can be seen in the above graph, where $ACCF_{max}$ correlation is over 90%, material composition does not vary by more than 5%. However, the same can be said for the lowest correlation scores (-17.61% and 10.41%). Where the highest difference in material composition is found, in correlation between test object 2A and known match 2B, a high correlation is still achieved. Therefore, a variation in material composition does not directly link to material composition of the primer cap.

As can be seen in Table 4-11 and Figure 4-21, using the percentage of material composition alone will not always differentiate between a good match and bad match in correlation, as there is no obvious pattern in the quality of the match and the composition. Therefore, further work will be needed to corroborate these findings. It is suggested by the author that hardness testing primer caps of various manufacturers should be completed to ascertain how the hardness values of the primer caps are affected by the material composition. During this study, this was not possible as the Odyssey collection must remain pristine and undamaged to allow for further research. As hardness testing is a destructive technique, such studies cannot be completed on the Odyssey collection.

4.7.3 Volume parameters

To corroborate the theory that the material of the primer cap will influence the topography of toolmarks imparted into the primer cap, volume parameters can be used to ascertain the difference in overall topography.

To fully ascertain the difference, test objects must be compared where they have a known match of similar material composition and a known match with a larger variation. For this purpose, test objects 1A, 2A and 14A will be compared to their known matches, as in each of these cases one known match has similar material composition while the other has variations.

The following graphs show the volume parameters found in each of the test objects and their known matches (Figure 4-22-Figure 4-25):

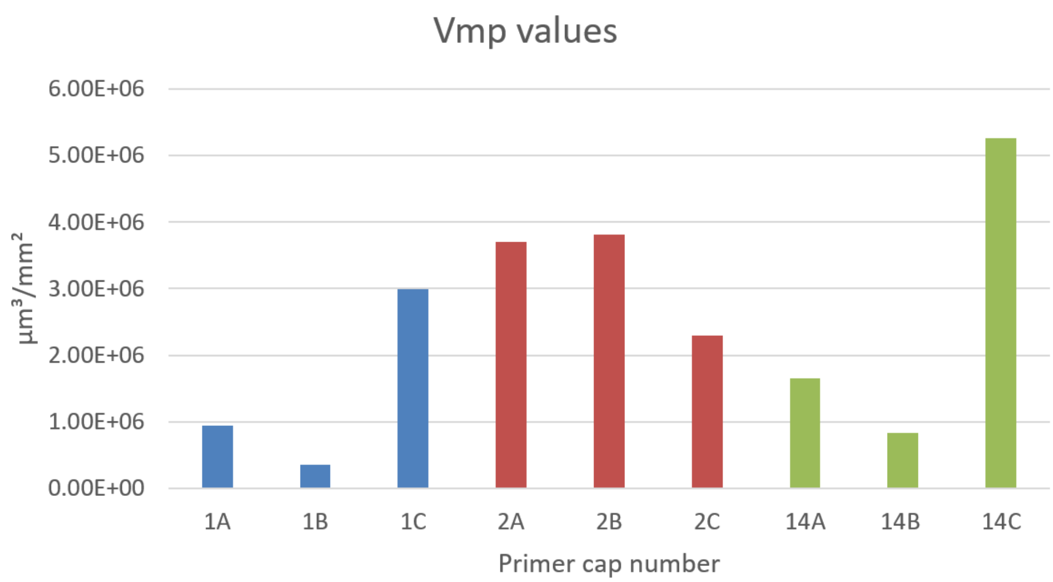


Figure 4-22: Vmp (peak material volume) values of each primer cap

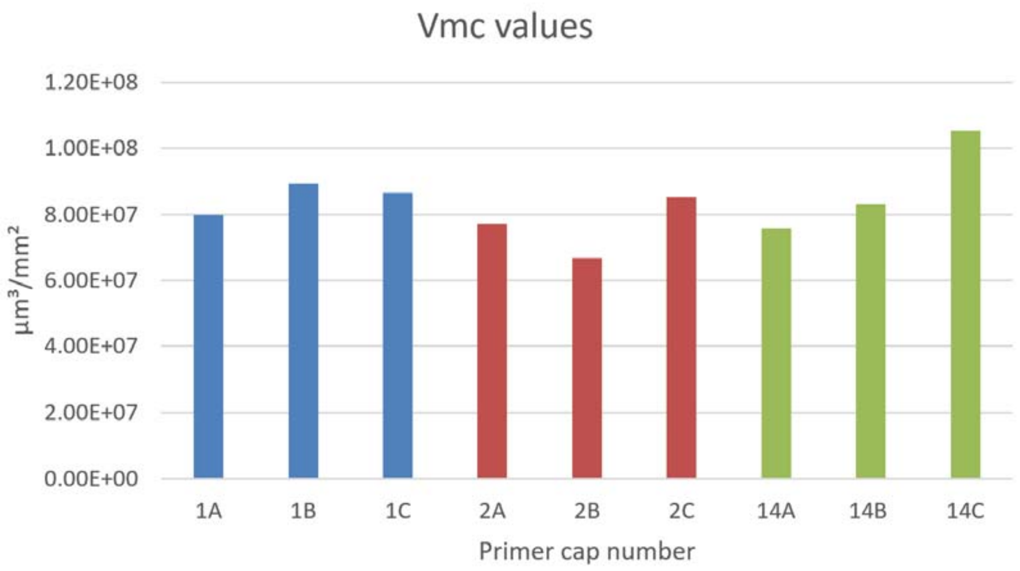


Figure 4-23: Vmc (core material volume) values of each primer cap

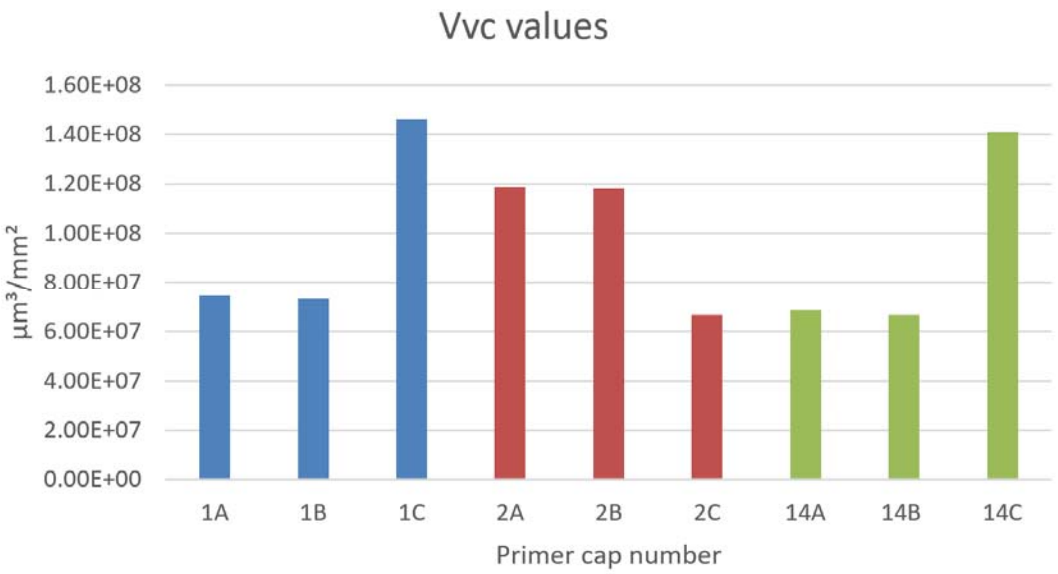


Figure 4-24: Vvc (core void volume) values of each primer cap

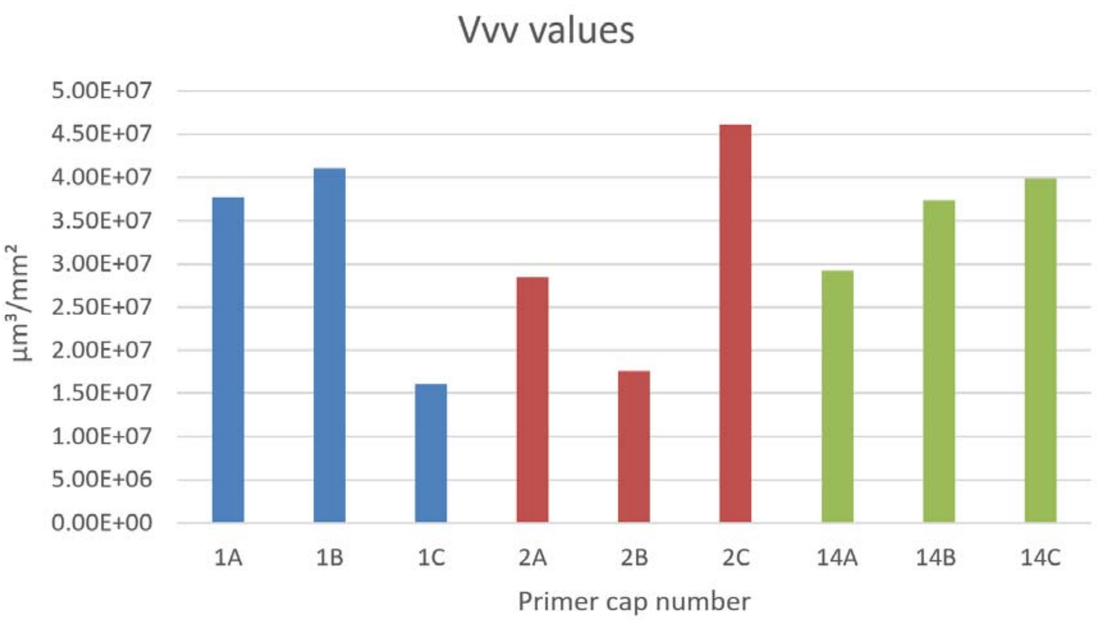


Figure 4-25: Vvv (valley void volume)

In the above results, it can be seen that in Vmp, Vmc and Vvv values gained there is not a pattern seen in the values with regards to the material of the primer cap. However, in Vvc values gained it appears that there is a stronger correlation in the values with regards to the material.

The cartridge cases 1A and 1B, 2A and 2B and 14A and 14B have similar material composition between the known match and the test object (A and B). In each of these cases it can be seen that the Vvc values are similar. Where the material composition does vary, between known matches A and test objects C, there is also an obvious difference in the Vvc value.

As the Vvc value relates to the core material void volume, it is expected that the changes in values are reflected in a difference in the depth of the impressions due to the material hardness. A softer material will deform more when hit by the firing pin, and therefore will have a larger core material void value.

4.8 Summary of findings

Within this chapter, it has been found that the use of a robust Gaussian filter with a high pass cut off of 75μm and a low pass cut off of 450μm would provide a pre-processing technique that is able to improve the correlation of firing pin impressions with the Odyssey collection. It was also shown that translating the filtered surface from Cartesian to polar coordinates would result both in a more accurate correlation and a more time efficient system. Further work would include increasing the number of non-matches within a database, to determine if the system will remain robust in increased sample sizes.

In previously published research, the pre-processing methods have varied from those used in this study. Gaussian filtration of the surface has been used with cut off values shown in Table 4-12 (Song, 2015; Zhang, Song, Tong, & Chu, 2016):

Table 4-12: Table of pre-processing used in previously published methods

High pass cut-off value (μm)	Low pass cut-off value (μm)
15	150
7.81	93.75
1.6	110

NIST currently employ the use of congruent matching cells (CMC) (Song, 2015) or congruent matching profiles (CMX) (Zhang, Song, Tong, & Chu, 2016) in firing pin impression correlation and have published the use of all the above filtering techniques. In the CMC method the firing pin impression is firstly segmented into a number of separate cells. Correlation efficacy is based on the number of congruent matching cells found, with a degree of dissimilarity in number of CMCs found in known matches and non-matches. To be considered a CMC, there must be a minimum $ACCF_{max}$ result of 25%, which alone would be considered a low correlation result, however in this case correlation is based on the number of matching cells rather than the $ACCF_{max}$ value. In previously published methods, results have shown a higher differentiation between known matches and non-matches. However, it should be noted that previous methods have included the use of firearms with random toolmarks gained from using bead and sand blasting in the manufacturing process, and fewer firing pin impressions were included in each database with a controlled sample of cartridges being used.

In the comparison of hitlist results using differing measurement and correlation techniques, it was found that while it is possible to increase the efficiency of identification systems based on advanced methods, this is only the case when using certain measurement techniques combined with the correct pre-processing methods. While using the Alicona to acquire datasets did not result in an increase in correlation efficiency compared to previous systems, datasets acquired using the Alias system did result in an increased efficiency. Finally, this study proves that it is possible to shift from subjective methodology into objective techniques for the identification of ballistic toolmarks, in which there is little user input and correlation is based upon a percentage match, rather than a 'match score', that has no evidence of statistical value. While this study did rely upon the author manually choosing the area in which the firing pin impression was present within the dataset, it is believed that with further work this could also be automated, thus resulting in an identification system in which user input is decreased in the identification of ballistic toolmark evidence, and thus satisfying the principles within the Daubert standard for the quality of evidence within a court of law. The author believes that the system would provide a means for forensic laboratories to effectively reduce the number of visual comparisons completed by the examiner. Instead, the system would be able to provide an examiner with a short hit list of potential matches to investigate, thus becoming more time and cost effective.

In previously published results, the correlation efficacy of the IBIS Heritage system has been discussed. Currently there is very little information published regarding the newer BrassTrax-3D system, which relies on confocal microscopy rather than greyscale imaging techniques.

One previous study (Tulleners, 2001), quoted hitlist results of known matches to be: 26% found in top position, 30% in the top three, 32% in the top five and 42% in the top ten. In this thesis, 77% were found in the top ten using Alias measurements and $ACCF_{max}$ correlation, indicating an increased efficacy in results gained in this study. In another study (George, 2004), 48% of known matches were found within the top ten of the hitlist, further corroborating these findings. However, in two other studies (De Kinder, Tulleners, & Thiebaut, 2004; Nennstiel & Rahm, 2006), 78.1% were found in the top ten, and 75%-95% were found in the top five. These results demonstrate a dependence on the test database used with regards to quoted efficacy. A direct comparison cannot be made until variables in the test database are minimised.

With regards to variability in correlation efficacy between known matches, there is no conclusive evidence that material composition has an effect in this study. However, Brinck (Brinck, 2008), found that correlation of copper jacketed bullets would be significantly more efficient than lead bullets using both the IBIS Heritage and Brass-Trax3D systems. No correlation between bullets with different material composition was achieved, which does occur in correlation of the Odyssey collection and has no clear effect on correlation results. It has been found in other studies that a difference in manufacturer will have an effect on toolmarks, but the larger effect can be found in breechface impressions. The transfer or breechface toolmarks to a primer cap depends on the chamber pressure created. Chamber pressure can be affected by manufacturer, due to differences in propellant used. While this pressure can influence the flowback on a firing pin impression i.e. the larger crater forced outward around the firing pin impression, it has a smaller effect on the contact between primer cap and firing pin (Cork, Rolph, & Meieran, 2008). Therefore, differences in void parameter are more likely due to variation in chamber pressure rather than material. Where level of flowback varies, there will be a difference seen in the volume of the firing pin impression.

Further work suggested would be to create more test fired cartridges, in which the same manufacturers cartridge is fired several times, rather than single instances of a cartridge manufacturer as seen in the Odyssey collection. The results would identify toolmark variability due to a difference in manufacturer used.

CHAPTER FIVE

BULLET CORRELATION STUDIES

In cartridge correlation studies, it was found that using a Robust Gaussian filter with cut off values of 75µm and 450µm along with translation to polar coordinates and $ACCF_{max}$ *correlation would allow for an increase in correlation efficacy compared to previous systems. However, the topography of bullet striations does vary from the individual characteristics found in firing pin impressions, and therefore pre-processing methods must also vary.*

This chapter presents the use of wavelet decomposition for separation of spatial frequencies present in the areal topography of bullets, along with $ACCF_{max}$ *correlation for the efficient identification of bullets. To the author's knowledge, this is the first instance in which the areal surface (rather than an average profile) has been used in correlation.*

Results from this study are compared against other correlation results gained using the Odyssey collection bullets, with an overall increase in correlation efficacy was observed. A study was also completed to ascertain whether or not a variation in bullet jacket material would have an effect on correlation efficacy.

5 Bullet Correlation Studies

5.1 Introduction

The comparison of toolmarks on bullet surfaces has until recently relied on visual side-by-side comparison of striations contained within Land Engraved Areas (LEAs). Research methods have determined methods in which both 2D images and areal datasets can be correlated mathematically.

There are two main methods in which correlation of bullet evidence occurs. Firstly, pattern matching can be used to ascertain the level of similarity between 2D patterns present on two bullet surfaces. This is used for the correlation of 2D images using microscopy and a CCD camera, and is used by commercial systems such as Balscan, Evofinder and Arsenal. Another current correlation method is to acquire the areal topography of a bullet and to then create an average profile for each of the Land Engraved Areas (LEAs). The averaged 2D profile it then used in correlation (Bachrach, 2002; Chu, Thompson, Song, & Vorburger, 2013).

In using 2D image acquisition, the images acquired can be affected by differences in lighting, in which areas may become shadowed and toolmarks obscured. The technique can be considered objective as the quality of the dataset depends on the user's measurement technique. As the technique cannot be considered reliable or repeatable, it does not satisfy the Daubert standard of evidence. Moving to areal data acquisition of the bullet surface would remove effects of lighting from measurements, as areal data is based on acquiring height point data rather than optical images.

While the more advanced systems such as the BulletTrax-3D system allow for acquisition of the full topography of the bullet surface, currently this is not being taken advantage of. Correlation using areal topography occurs only after the dataset has been averaged into a single profile measurement, which may result in outliers on the surface, for example deep scratches due to damage, skewing correlation results.

The aim of this study was to ascertain whether it would be possible to use the entire 360° circumference of a bullet acquired as areal topography to gain effective correlation between bullets. This would result in a technique that is truly subjective, with correlation results that are based in statistical reasoning.

To ascertain the differences in correlation in measurement techniques, two measurement systems were used, the Alias and the Alicona. It is expected that a more efficient correlation will be gained in using datasets acquired using the Alias system, as it has been shown that the Alicona may impart measurement artefacts into the measured surface data.

The results of this study will also be compared to the 2D systems used in a previous study to be able to ascertain whether the use of areal topography can increase the efficiency of correlation.

5.2 The Odyssey collection: bullets

The Odyssey collection consists of 390 bullets, all of which are 9mm Luger caliber with six LEAs and a right-hand twist. All bullets were fired and collected forensically, using either a water tank or cotton wool to capture the bullet without the risk of damage. Eight different bullet manufacturers were used, and therefore the outer material of the bullet does vary between nickel coated, copper coated and brass coated bullets with variation also in brass material composition. The class characteristics of each bullet within the Odyssey collection are the same, and therefore correlation will be needed to be able to differentiate between known matches and non-matches.

While it is acknowledged that Groove Engraved Areas (GEAs) do exist where the groove area of the barrel imparts toolmarks into the bullet surface, GEAs will not be used for correlation. It is known that there is a level of variability in the transfer of GEAs to bullets due to contact not always being present between a bullet and a GEA, dependant on the size and position on the bullet (Bolton-King, et al., 2010). In the case of the Odyssey collection, the vast majority of bullets did not have GEAs present.

Within the Odyssey collection, 23 of the bullets were designated as test objects. These 23 bullets each have two known matches within the collection, totalling 69 bullets within the collection that are known to have two matches, i.e. were fired from the same firearm.

The remaining 321 bullets were designated as the background sample, and had no matches within the collection, as described in Figure 5-1:

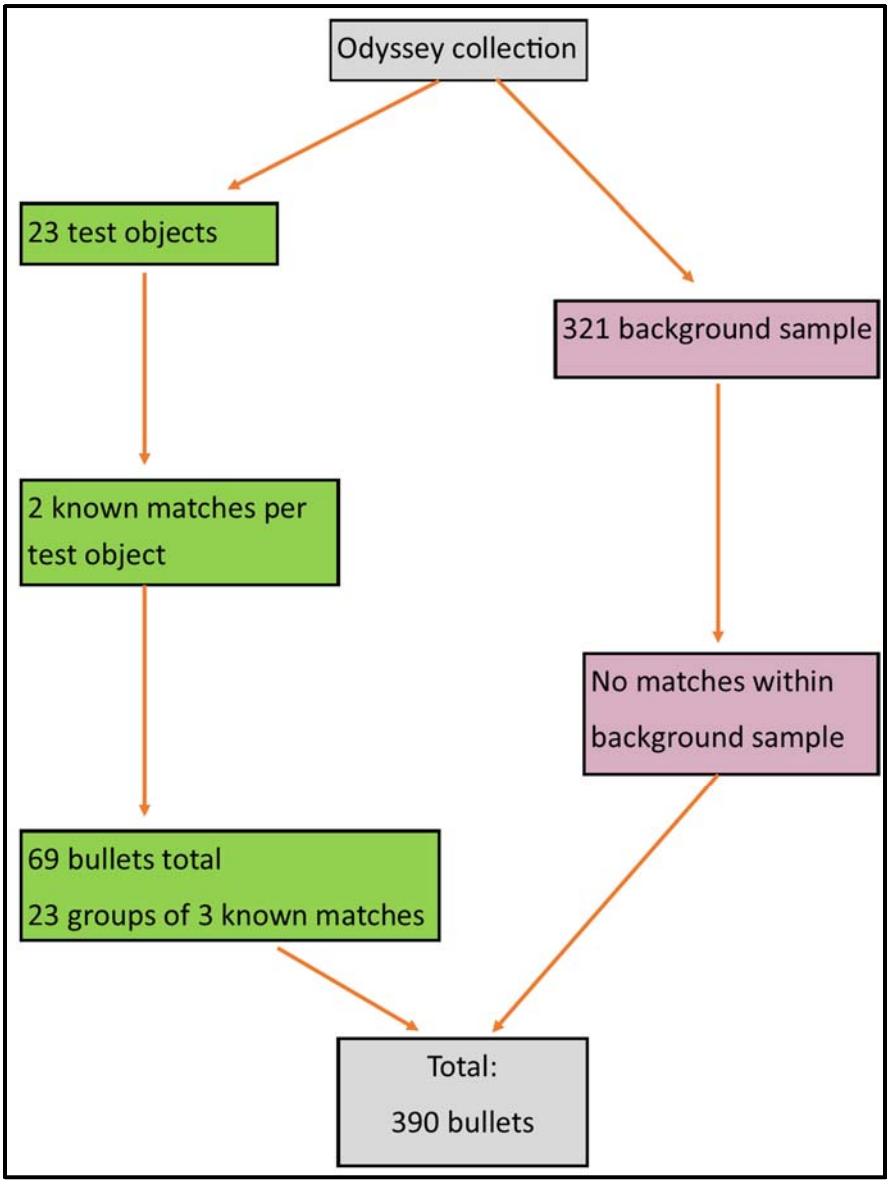


Figure 5-1: Design of the bullet Odyssey collection

Each bullet contained within the Odyssey collection was engraved with a four-digit identifying number to ensure the bullet could be traced back to the gun it was fired from. However, to ensure the Odyssey collection remains available for further blind testing this information will not be included.

5.3 Unwrapping of bullet data

The measurement of a bullet circumference means that a 360° measurement must be acquired. The cylindrical form of the measurement must be removed, resulting in a flat surface in which the length is the circumference of the bullet, so that pre-processing of the dataset can be achieved. In the two measurement systems used in this study, the ‘unwrapping’ of the dataset was achieved in different ways. A variation in the unwrapping method may results in a difference in the data fidelity, for example downsampling of data may occur where some height point data is lost during the unwrapping process.

5.3.1 Alias system

The unwrapping of a bullet measurement is controlled within the Alias software, and therefore there is no other user input needed. Based on the original information inputted by the user (number of LEAs, caliber, direction/angle of twist), the system will know the number of LEAs engraved on each bullet, and will acquire one dataset per LEA measurement is based on number of LEAs, as this is the class characteristic with less variation when compared to GEAs. For instance, if the user inputs that there are six LEAs on a bullet, the system software will ensure that six separate datasets are acquired across the circumference of a bullet, each encompassing 60° of the bullet surface. The starting point for the measurement is defined by the user, where a 2D camera is used to align the measurement start point to a LEA. Each dataset is then saved side-by-side in the software as one bullet measurement. Therefore, no cylindrical from is acquired during measurement and as such datasets can be pre-processed without the need for unwrapping, as pictured below in Figure 5-2:

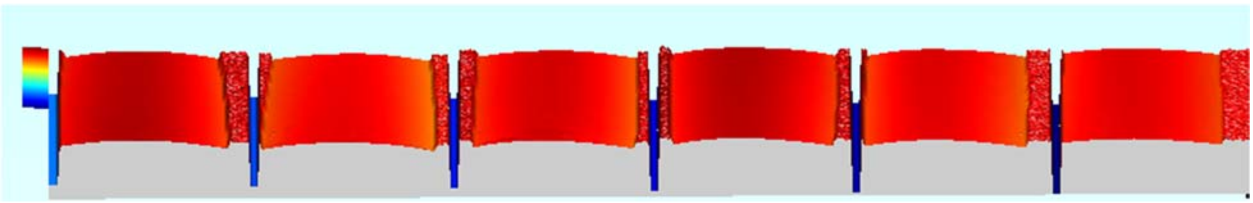


Figure 5-2: Alias acquired bullet dataset as represented in SURFSTAND software

5.3.2 Alicona measurements

As shown in Figure 5-3, to try and create a measurement method using the Alicona system that would need as little user input as possible, bullets were measured as a 360° cylindrical form. Therefore, in pre-processing of the acquired datasets it was necessary to remove the cylindrical form of the dataset, to move from an annular to areal dataset.

To achieve this, Alicona specific scripting was used to extract the data as a series of profiles, set to be 4µm apart to ensure the lateral resolution was not subsampled.

Firstly, the cylindrical dataset must be aligned so that it is perpendicular to the y axis. Some cropping of the dataset may be required to ensure that both edges of the dataset are perpendicular to the y axis. Then, profiles are extracted across the whole of the 360° at a point spacing of 4µm perpendicular to the y axis and are placed side by side to create an areal dataset (Walton, Addinall, Zeng, & Blunt, 2017). As the unwrapping of each bullet takes around 40 minutes, when combined with a measurement time of 40 minutes its becomes apparent that measuring the bullets using the Alicona system may be too time consuming in a forensic laboratory. However, due to the cheaper cost and ability to measure a wider variety of forensic evidence, the time expended could ultimately be less of a concern, dependant on efficacy of correlation.

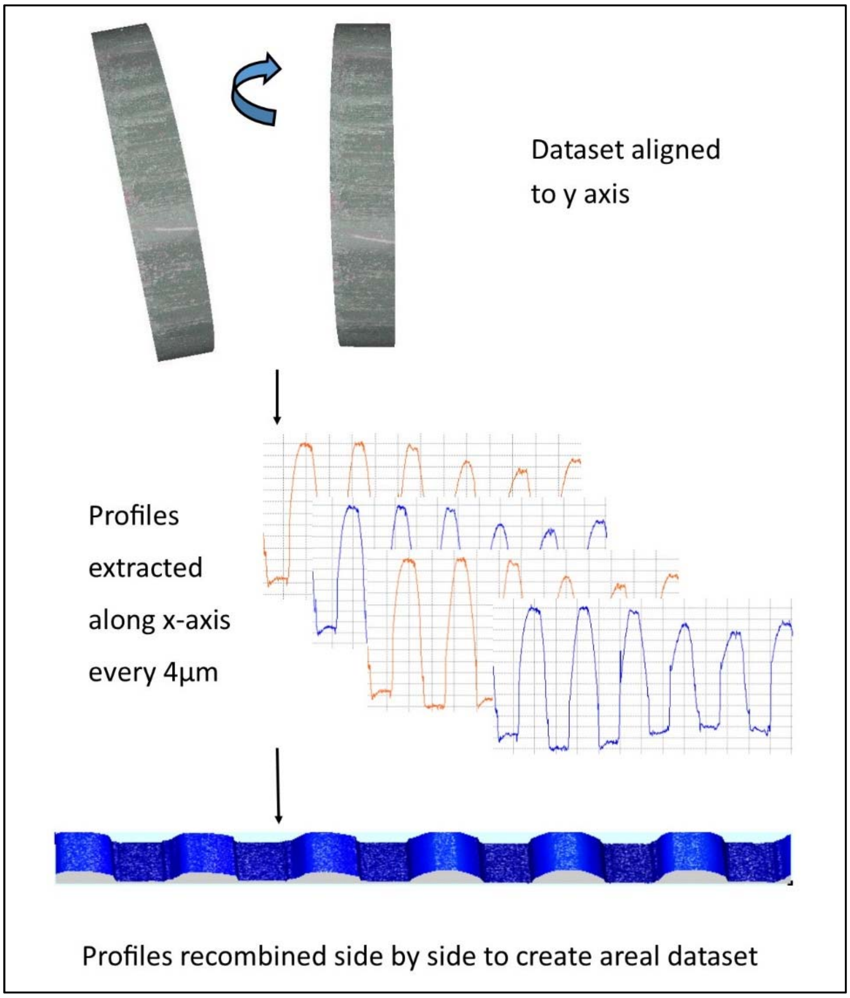


Figure 5-3: Schematic of unwrapping procedure for Alicona datasets

5.4 Pre-processing of data

This study aims to ascertain the correct pre-processing methods needed to be able to differentiate the salient (individual characteristics) from other non-salient data present within the dataset. The testing of pre-processing methods was limited to the Alias datasets, which would then be applied to Alicona datasets for a full comparison.

In previous research by NIST (Chu, et al., 2010), Gaussian filtering was used on the surface topography of a LEA acquired using confocal microscopy before the data was averaged into a 2D profile. This method was then used to test the Standard Reference material of bullets, to ensure each bullet was a similar as possible. The method showed that there was a high level of agreement in the SRM material. However, in tests carried out by the author it was found that using this method of filtration to then correlate the areal topography of the data would lead to poor results when using the Alias measurement instrument. $ACCF_{max}$ values gained were always high, meaning the method was unable to differentiate known matches and non-matches, Figure 5-4:

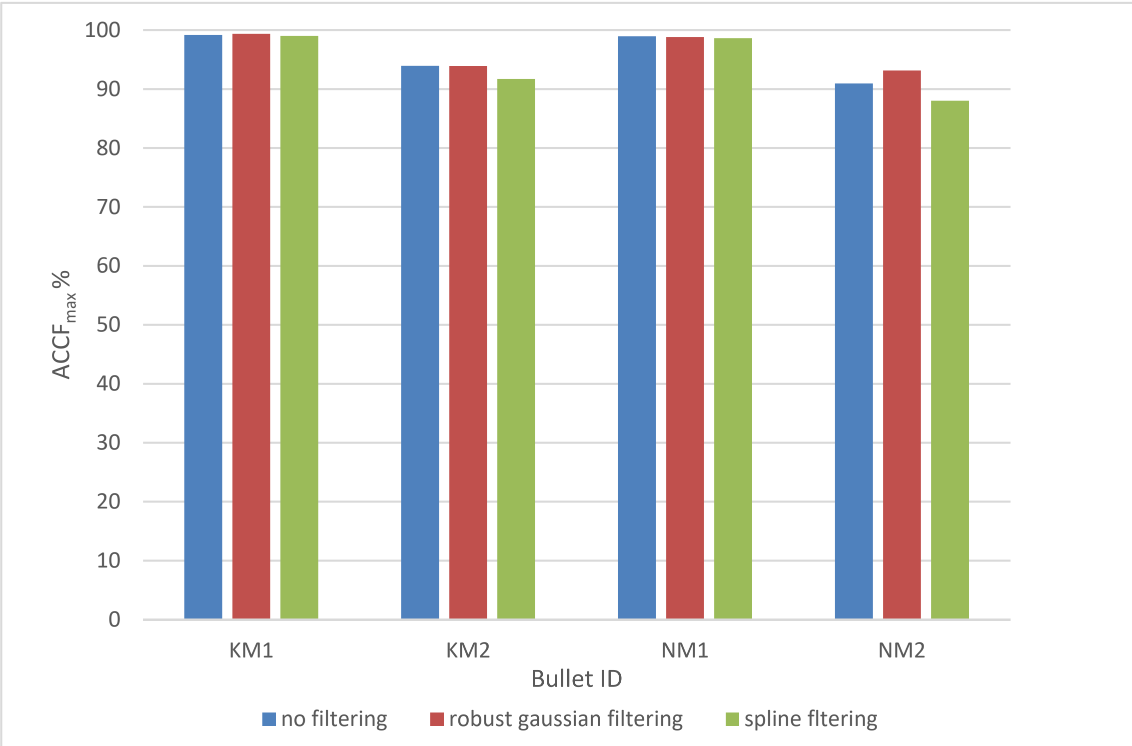


Figure 5-4: Graph of example $ACCF_{max}$ results using various filtering techniques

As can be seen in the graph above (Figure 5-4), using traditional filtering methods such as Robust Gaussian and Spline filtering, with ISO standard cut off values, has negligible effect on correlation values when compared with those gained from using an unfiltered dataset. In comparison between known matches and non-matches, it can be seen that there is no differentiation between the $ACCF_{max}$ values gained, and therefore the methods cannot be used in a forensic analysis system. Other results can be found in Appendix 5.
As traditional filtering methodology appeared unable to separate salient and non-salient data, it was hypothesised that decomposition of a surface into separate frequency bands present in the surface topography may deliver more fruitful correlation (EU Patent No. 0811985, 2008; Vorburger, Song, & Petraco, 2015). It is believed that in using such methods it will be possible to segregate the individual characteristics, such as the overall curvature of the bullet, optical noise and larger scale characteristics of the toolmark which are not individual, from other characteristics within the surface for further correlation.

Using wavelet decomposition, a wavelet is translated and dilated while multiplied with frequencies of the original surface, thus resulting in a different decomposed surface per frequency bank (Abdul-Rahman, Xiang, & Scott, 2013). The acquired dataset is decomposed into various bandwidths of frequency present in the full dataset. Decomposition starts with the highest frequency on the surface, and in bullet surfaces it results in ten decomposed surfaces, as in the case of bullet decomposition, 10 wavelet bands will encompass all frequency levels of the original surface. Table 5-1 describes the wavelet bands.

Table 5-1: Frequency bandwidths in each decomposed surface

Wavelet band	Frequency band	Corresponding spatial wavelengths (µm)
D1	Highest frequency (Fn) – Fn/2	3.94-7.88
D2	Fn/2- Fn/4	7.88- 15.76
D3	Fn/4- Fn/8	15.76-31.52
D4	Fn/8- Fn/16	31.52-63.04
D5	Fn/16- Fn/32	63.04-126.08
D6	Fn/32- Fn/64	126.08-252.16
D7	Fn/64- Fn/128	252.16-504.32
D8	Fn/128- Fn/256	504.32-1008.64
D9	Fn/256- Fn/512	1008.64-2017.28
D10	Fn/512- Fn/1024	2017.28-4034.56

The number of D levels is dependent on the original surface, and the shortest wavelength present (based on the lateral resolution of the measurement instrument) and the longest wavelength present, which in the case of bullets will be the curvature of the surface. 9mm Luger bullets are decomposed into 10 levels, each resulting in a surface containing features only within the frequency band of the D level.

It was necessary to ascertain which bandwidth(s) would relate to the individual characteristics. To do this, a small subset of the Odyssey collection was correlated, both known matches and non-matches, to determine if any of the bandwidths would show a differentiation in correlation between known matches and non-matches. Two test objects were chosen at random, and each were correlated against a known match and non-matching bullet. For each bullet pair, a total of six correlations were completed, representing each LEA of the test object bullet.

An example of correlation scores using different wavelet bandwidths between a known match pair and non-match pair can be seen below in Figure 5-5 and Figure 5-6.. Correlation was achieved by manually selecting one LEA of the test object bullet, which was then correlated against all other LEAs in the test bullet. This ensured there would be no error introduced into correlation through a rotational variance in the measurement of the bullets.

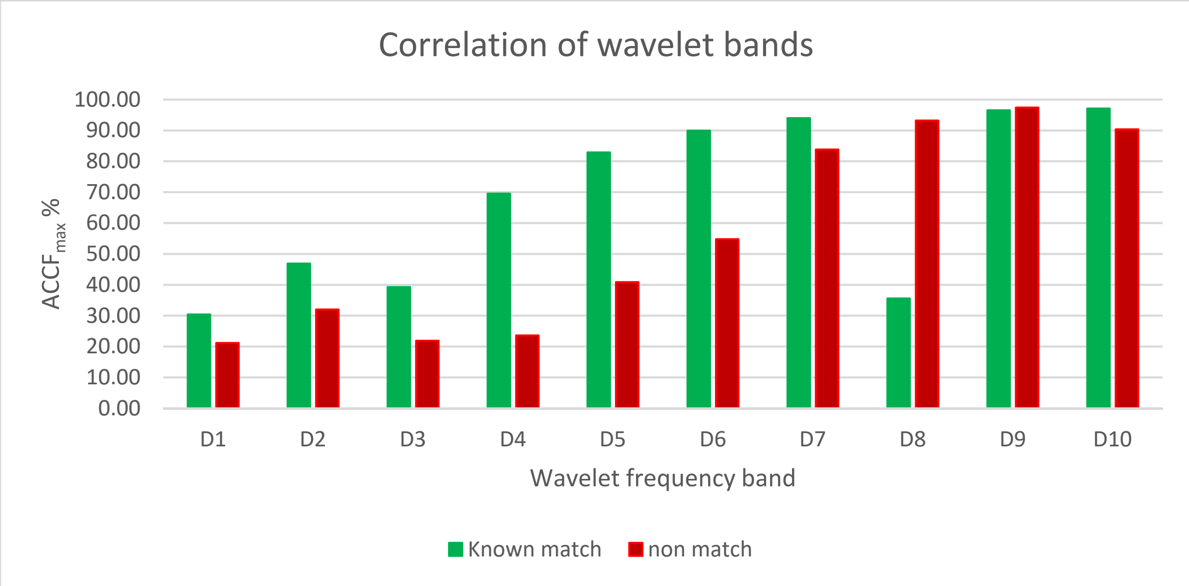


Figure 5-5: Graph of $ACCF_{max}$ results using different wavelet bands in a known match and non-match.

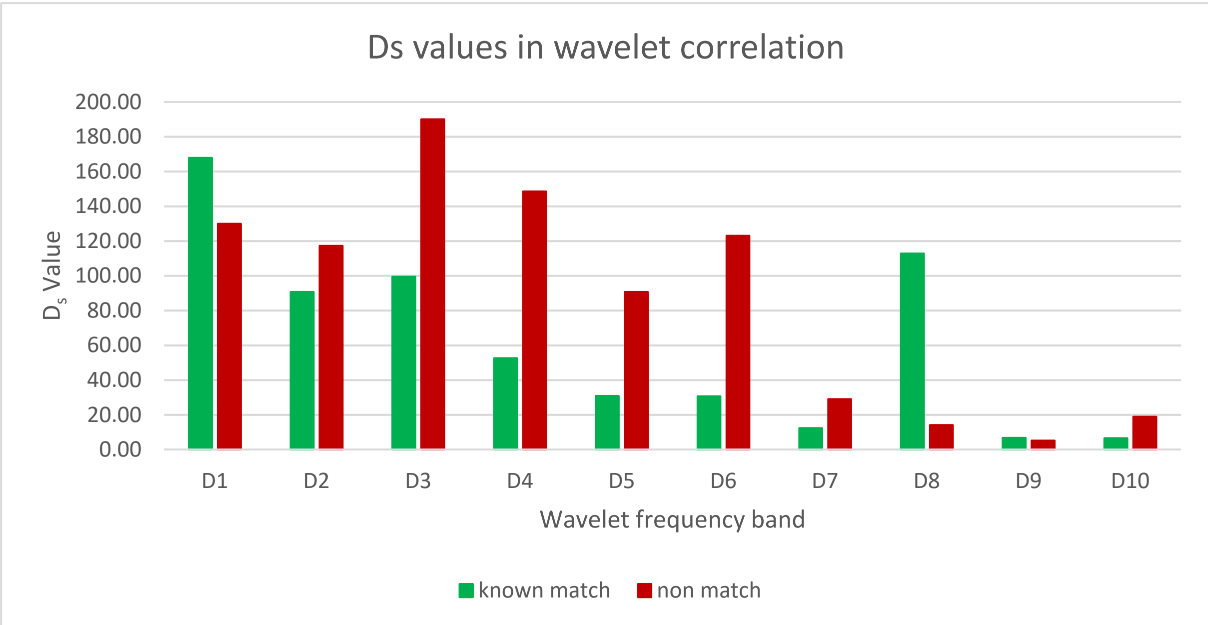


Figure 5-6: Graph of D_s values using different wavelet bands in a known match and non-match

As can be seen in the above graphs, the correlations with the least differentiation between a known match and a non-match can be found in the frequency bands D9 and D10, i.e. between spatial frequencies of 1008.64μm and 4034.56μm, where there is a high $ACCF_{max}$ percentage and low D_s value in both cases.

As the D9 (Figure 5-7) and D10 wavelet bands correspond to low frequency bandwidths contained within the original surface, it is to be expected that high correlation and low differentiation will be achieved within these bands. This is due to the fact that information contained within these bands corresponds to the overall form of the dataset, and as all bullets correlated are 9mm caliber, it is to be expected that form will be very similar in both known matches and non-matches.

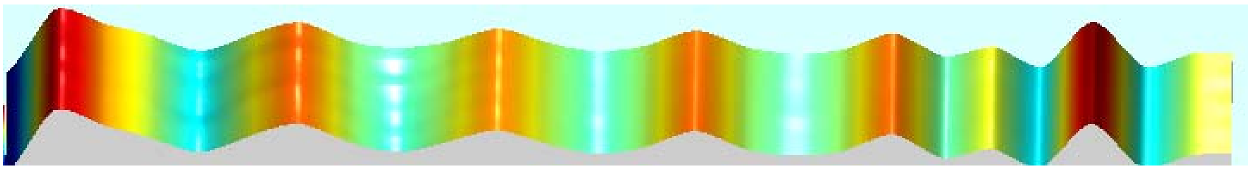


Figure 5-7: D9 surface of a bullet

The wavelet bands that show the largest differentiation between known matches and non-matches are the bands D4, 5 (Figure 5-8) and 6. In each of these cases the $ACCF_{max}$ percentage score was considerable higher in known matches than non-matches, and D_s lower in known matches. Therefore, using these wavelet bands it is possible to differentiate between known matches and non-matches, which indicates that individual characteristics have been separated from other surface characteristics.

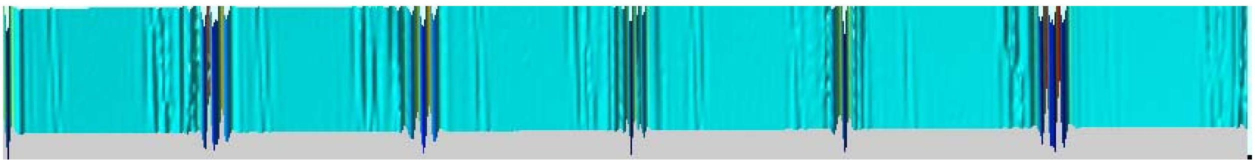


Figure 5-8: D5 surface of a bullet

In higher frequency bands, i.e. D1 (Figure 5-9), 2, and 3, it was found that there was less differentiation between known match and non-match $ACCF_{max}$ results, and D_s scores were too high to consider any of the correlations a true match due to scale differences between the two surfaces. As these wavelet bands correspond to high frequency components of the original surface, it is expected that high frequency noise and other surfaces components such as random damage are contained within these wavelet bands that do not correspond to any toolmark characteristics, thus resulting in erroneous correlation results.

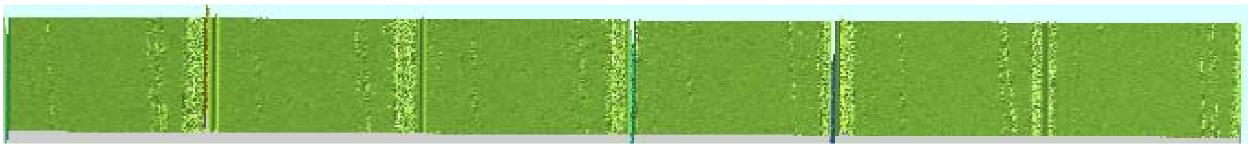


Figure 5-9: D1 surface of bullet

5.5 Overview of bullet correlation techniques

Correlation of the Odyssey collection within this study was gained using MATLAB code created by Dr Wenhan Zheng at the University of Huddersfield, which would allow for the cross-correlation of two bullet surfaces. However currently this is not fully automated, and as such the user must first process the surface and manually select the area of interest on one bullet surface. The MATLAB code will then compare this area to the full area of the second bullet surface and calculate the maximum cross-correlation result. The steps needed in correlation of bullet surfaces within this study are pictured in Figure 5-10.

It was decided that one LEA of the test object would need to be correlated against every other LEA within the compared bullet to reduce any effects in a rotational difference in measurement.

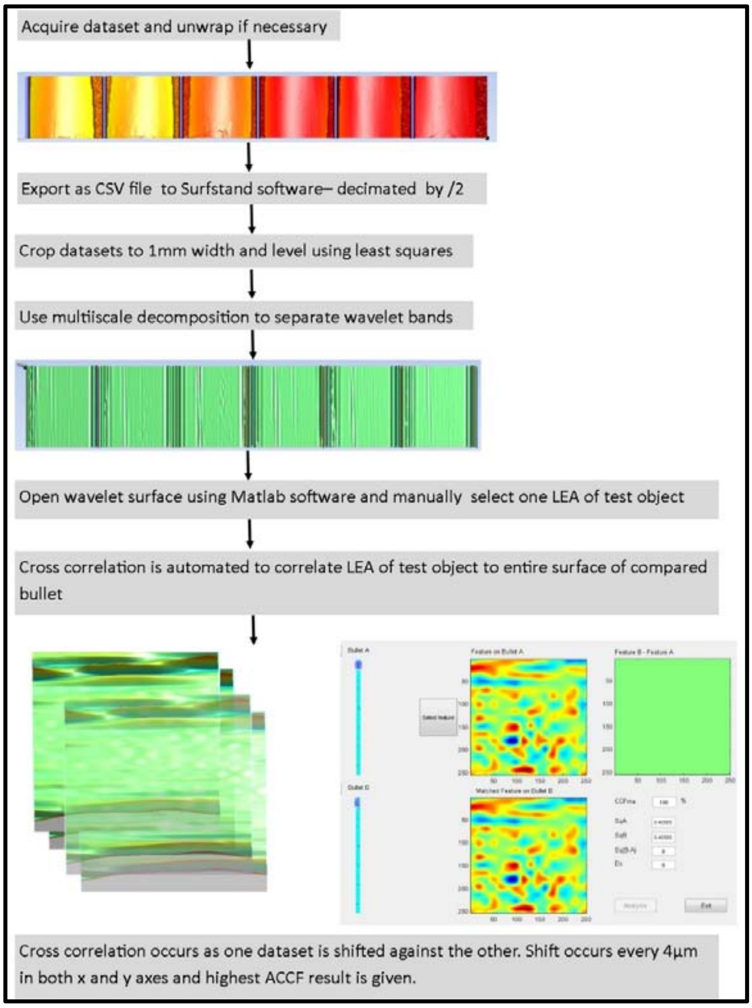


Figure 5-10: Flowchart of bullet correlation

Each LEA engraved into a bullet surface will have differences, due to each rifling pattern being created by a difference surface of the tool (Heard, 2013). As it is then virtually impossible to conclude which LEA came from which land in the barrel, there is a high chance that correlation may be adversely affected in a known match if an LEAs correlated in two bullets were not created by the same land in the barrel. To minimise such effects, it was decided to use one LEA of the test object to correlate against all over LEAs in the compared surface. A width of 1mm was used by the author, as the author's workstation would allow wavelet decomposition at a maximum of 1mm width.

Within this study, the following correlations of the complete Odyssey collection were completed (Table 5-2):

Table 5-2: Table of completed correlations

Instrument used	Wavelet band for correlation used	Total number of correlations
Alias	D5	8947
Alias	D6	8947
Alias	Combined surface of D5 and D6	8947
Alicona	D5	8947
Alicona	D6	8947
	Total correlations:	44,735

Each test object (23) is correlated against all other bullets within the Odyssey collection (389). This is repeated for each combination of instruments and wavelet band. The number of correlations detailed above was also used in a previous project with two commercially available 2D identification systems, therefore methodology was kept the same to ensure that a true comparison of hitlist efficacy between previous studies and this study could be achieved. While the D4 band also seems to offer differentiation in known matches and non-matches a full correlation test was not achieved due to time constraints, however the correlation of the D4 wavelet coefficient must be completed in further work to gain a wider perspective on the correlation of bullets using wavelet decomposition. As shown in Appendix 5, the D4 frequency band does have the potential to allow for differentiating correlations.

5.6 Correlation results: Alias measured datasets

The following table (Table 5-4) details the hitlist results gained in the use of each wavelet band separately, the combination of frequency band and the combination of the separate hitlist results. An LEA of the test object is chosen manually by the user, ensuring all data within the chosen area relates only to toolmarks present within an LEA. This area is then correlated against the entire surface of all other bullets within the Odyssey collection. As it was found through this study that a known match would in most cases result in a D_s value of under 100, any result with a D_s value higher than this would be excluded from the hitlist. It was also found that in most cases a correlation with a D_s value over 100 would have a low $ACCF_{max}$ value, corroborating the view that a D_s value over 100 indicated a non-match. Once these exclusions had been made the remaining results were ordered by decreasing $ACCF_{max}$ result, known as a hitlist. The hitlist position for each known match to the test object were then found and reported in Table 5-4.

As can be seen in Table 5-4, efficacy in correlation results between a test object and its known match can be variable. While 14 of the known matches have been placed as the best match when correlating in the D5 frequency band, ten have been excluded from correlation due to very high D_s values.

It is expected that this is the case as only the D5 wavelet band has been used for correlation, and those that do not match in the D5 band may match in the D6, as shown in Figure 5-11:

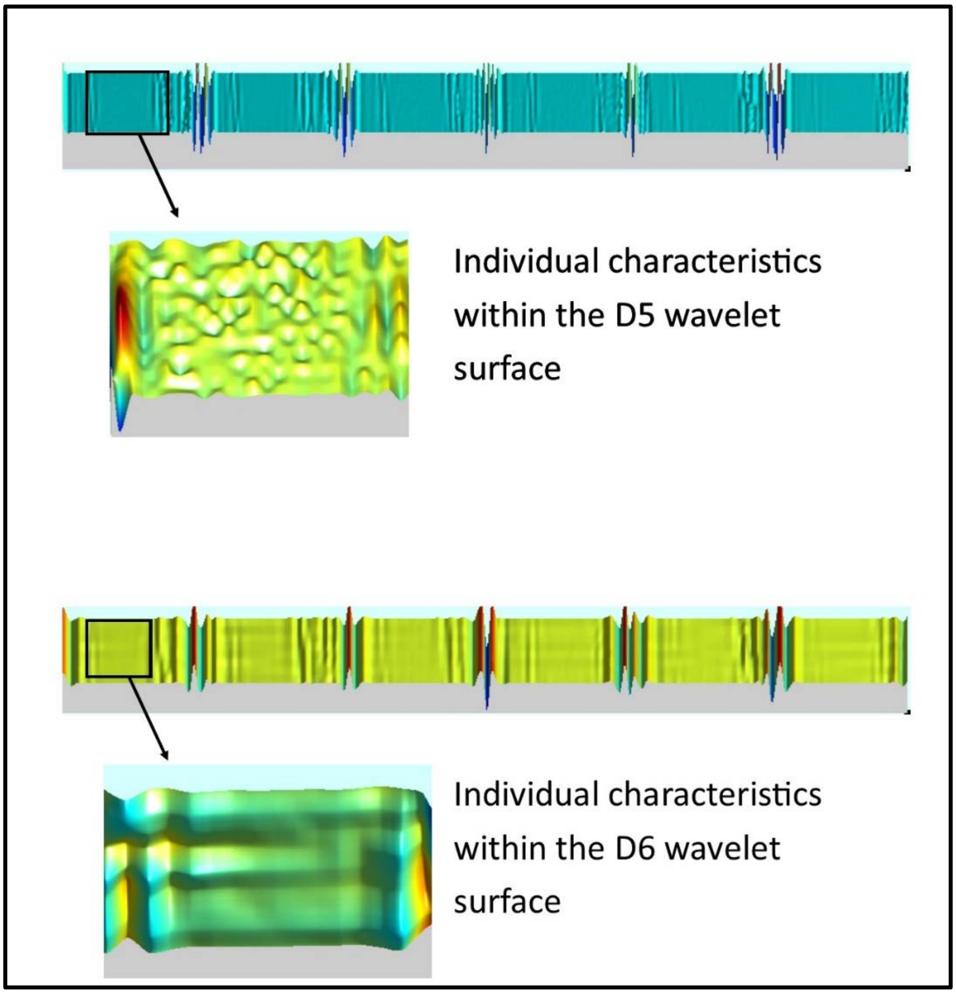


Figure 5-11: Individual characteristics within the D5 and D6 wavelet bands

To ascertain whether this was the case, test objects 3A, 5A and 11A were correlated against their known matches using the D6 wavelet band. The results of this correlation are as follows (Table 5-3):

Table 5-3: Table of correlation results using D6 wavelet band

Test object	Known match B $ACCF_{max}$	Known match B D_s	Known match C $ACCF_{max}$	KM2 D_s
3A	89.6933%	25.0077	62.7083%	61.6555
5A	89.8902%	19.6821	79.897%	61.4627
11A	93.5882%	47.3103	86.6768%	35.2216

As can be seen in the Table 5-3, the test objects which showed poor correlation from their known matches in using the D5 correlation have all given matching results using the correlation of the D6 band. $ACCF_{max}$ results are all a high percentage match, and one of the results can be excluded due to a high D_s score. Therefore, it is shown that in some cases while using the D5 wavelet band will result in poor correlation between known matches, correlation efficacy is increased by instead using the D6 band for correlation.
Based on the positive results in the above table, the correlation of the 23 test objects was completed again using the D6 band. The hitlist results of this correlation test are shown below (Table 5-4):

Table 5-4: Correlation results using various methods and Alias acquired datasets

	D5 band		D6 band		Combined frequency bands		Combined hitlist results	
Test object	Known match B Hitlist position #	Known match C Hitlist position #	Known match B Hitlist position #	Known match C Hitlist position #	Known match B Hitlist position	Known match C Hitlist position	Known match B Hitlist position	Known match C Hitlist position
1A	1	3	5	1	1	3	1	1
2A	1	No match	1	5	1	2	1	5
3A	No match	No match	4	2	55	95	4	2
4A	2	3	1	6	3	16	1	3
5A	No match	No match	9	5	No match	No match	9	5
6A	1	42	1	2	54	39	1	2
7A	5	26	161	166	24	51	5	26
8A	1	45	45	47	No match	19	1	45
9A	1	No match	No match	3	75	258	1	3
10A	2	23	6	13	4	15	2	13
11A	No match	No match	2	1	1	3	2	1
12A	46	2	7	2	1	2	7	2
13A	1	No match	No match	100	19	No match	1	100
14A	No match	2	15	40	9	1	15	2
15A	3	4	No match	3	2	20	3	3
16A	23	1	10	3	46	20	10	1
17A	2	1	No match	25	10	59	2	1
18A	1	15	No match	97	No match	9	1	15
19A	13	1	No match	12	18	2	13	1
20A	35	1	30	47	1	2	30	1
21A	1	2	21	37	1	18	1	2
22A	2	1	9	No match	2	No match	2	1
23A	1	92	No match	1	3	1	1	1

As can be seen in the comparison of hitlist results between the D5 and D6 wavelet bands, there are differences in results gained. For example, in the correlation of bullet 11A, no known matches were found in the hitlist when correlating the D5 wavelet band, however in use of the D6 band both known matches were in the top 2 results.

Between the two wavelet bands, the hitlists results are different for each test object bullet, with some correlations performing better in the D5 band and some in the D6, showing that between the two wavelet bands there are differences in the surface features being extracted.

As the wavelet bands will encompass different frequencies, it is possible that a material difference between test object bullets is causing shifts in the most efficient wavelet band for correlation. For instance, a harder bullet may deform less than a softer bullet, and as such the surface frequencies which contain individual characteristics may differ. Should the test object bullets have variations in material hardness, this may affect the frequency band in which the individual characteristics are contained.

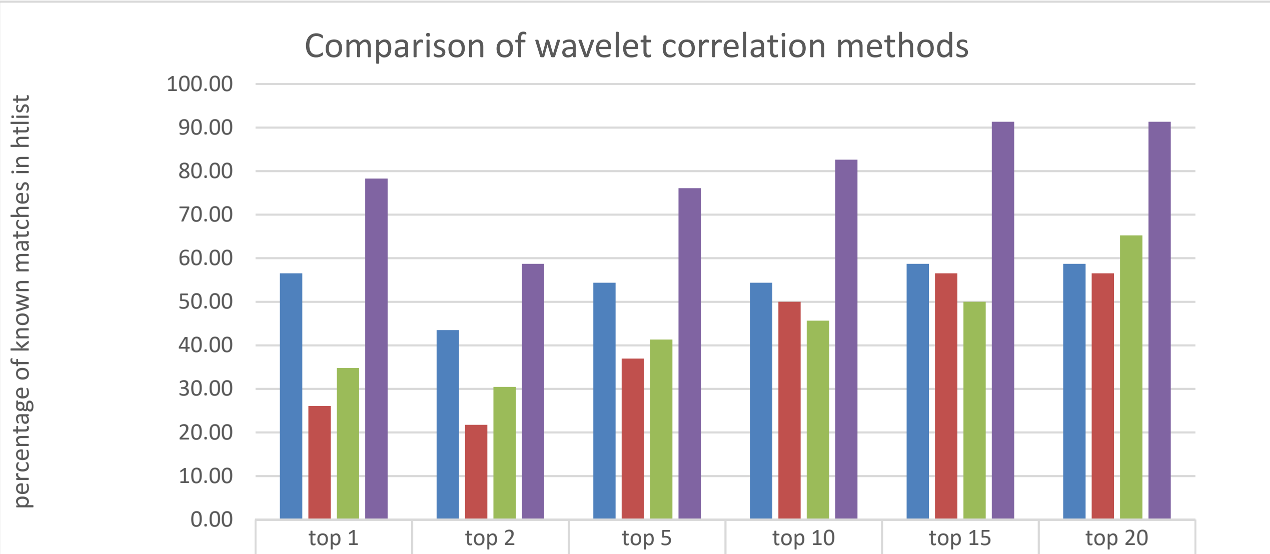
It was found that the effective wavelet band for correlation would vary between the D5 and D6 bands, it was decided to combine the two wavelet bands into one surface to determine whether it would result in a filtered surface that would allow for effective correlation regardless of the highest frequency of the original surface. The results of this correlation test are shown in Table 5-4.

As can be seen in the above table, combining the D5 and D6 wavelet bands into one filtered surface for correlation does slightly improve correlation gained from using the D6 band only, however efficacy is less than using the D5 wavelet band. In all cases, there are some known matches which are not included in the hitlists, and therefore from a forensic investigation point of view it cannot be proved that any of the above techniques would offer correlation sensitive enough for use as forensic evidence.

The combination of the D5 and D6 wavelet bands into a single filtered dataset encompassing all frequencies of both bands did not offer increased efficacy, and it is expected that this is due to the fact that combining the wavelet bands will introduce non-salient data into all surfaces, thus skewing correlation. When a bullet shows good correlation results in the one band, this is due to the salient data being contained within that band. In introducing another wavelet band data that does not relate to individual characteristics, non-salient data is introduced into the surface which will result in a skewed correlation. Therefore, it was decided to combine the hitlist results of the separate D5 and D6 correlations into one hitlist containing the better correlation result from either separate band. In this case, the better correlations were chosen from correlations using separate D5 and D6 wavelet bands, to determine whether correlation should be completed using two separate bands.

Table 5-4 shows that the best overall correlation efficacy can be found in combining the hitlists gained from correlation D5 and D6. This shows that in some instances individual characteristics will be contained within the D5 band, and in others they will be contained in the D6 band. It is accepted however, that in correlating unknown bullets as would be the case in a forensic laboratory, it would not be viable to use this approach, as there would be no information regarding which frequency band gave the correct results. Due to this, further work is suggested in which the frequency band can be determined manually, so that a singular band can be created which has differentiation between known match and non-match bullets. For this research, the frequency bands were determined by the lateral resolution of the instrument.

Figure 5-12 details the percentage of known matches found in hitlists of varying sizes:



	top 1	top 2	top 5	top 10	top 15	top 20
D5	56.52	43.48	54.35	54.35	58.70	58.70
D6	26.09	21.74	36.96	50.00	56.52	56.52
combined surfaces	34.78	30.43	41.30	45.65	50.00	65.22
combined hitlists	78.26	58.70	76.09	82.61	91.30	91.30

Figure 5-12: Graph of hitlist efficacy using various wavelet techniques

It can be seen in Figure 5-12 that the best correlation efficacy is gained from using combined hitlist results from separate correlation of the D5 and D6 wavelet bands, which indicates that salient information lies in a frequency band that encompasses part of both the D5 and D6 bands. To determine whether this is true, it would be necessary to create a system in which the frequency band can be determined by the user rather than the resolution of the instrument.

5.7 Correlation using Alias software

Within the Alias system algorithms exist that allow the correlation of bullet datasets. Correlation in the Alias system firstly relies upon the examiner's input with regards to the class characteristics of the bullet. This ensures that bullets with only the same or very similar calibers (e.g. 9mm Luger and 9mm Parabellum), the same direction of twist and number of LEAs are correlated against one another.

Once the possible datasets have been reduced dependent on the class characteristics, the software will automatically deduce the LEAs of the bullet for further correlation. The transition step between the LEA and bullet surface is found using seed growing, in which the surface of the LEA is grown from a seed, which will cut off when a large step in z height is detected. Once the LEAs have been separated from the rest of the bullet surface, LEAs are transformed into their component Zernike moments, which are a series of polynomials present on the surface. The Zernike moment that corresponds to the frequency containing the individual characteristics of the toolmark is then used for correlation.

In this study, the 23 test objects were correlated against all other bullets in the Odyssey collection. However, the correlation software failed to include any of the known matches within the hit list, meaning all results were negative, and as such the technique was deemed unsuitable.

Upon further investigation, it was found that the main reason for negative results was in the separation of LEAs from the unmarked bullet surface. It was found that in all cases the LEA was not separated efficiently, meaning that the unmarked surface (Groove Engraved Area) was included in the area chosen for correlation. In correlation of surfaces that contain no salient information, results will be skewed, and in this case, it resulted in poor correlation results. Previously published results show that the transitional area between LEA and GEA can be identified efficiently (Bolton-King, et al., 2012), and therefore Alias algorithms could be improved. Using 2D binary images of an acquired dataset principal component analysis of a fast Fourier transform can successfully define the transitional area. Also, a canny edge detection of the binary image, where edge detection is set to define the transition rather than striations, has proven to successfully define the LEA area.

5.8 Correlation using Alicona acquired datasets

Based on correlation results gained using bullet datasets acquired by the Alias system, it was decided to use the separate D5 and D6 bands for correlation of the Alicona datasets. As detailed previously (Figure 5-3), Alicona datasets are acquired in a cylindrical format, and are unwrapped into an areal data format. The areal data is then decomposed into the frequency bands. The D5 and D6 bands are then imported into the MATLAB toolbox, where an LEA of the test object is manually selected and correlated against all other LEAs of the 389 bullets.

In the correlation of bullets using Alicona datasets and the D5 wavelet band, there were very few accurate results gained. It was found that either the scale differences between known matches excluded the correlation result from being considered a true match, or the $ACCF_{max}$ percentage values were so low (<20%) that correlation could not be considered to have been successful. Upon further examination of the filtered datasets, it was found that this is most likely due to the fact that there is some instrument noise visible in datasets when compared to Alias measurements. While there have been previous results that show correlation can be achieved using Alicona D5 datasets, previous studies were based upon a very small database of bullets. In increasing the database size to include the entirety of the Odyssey collection, issues were found in a larger amount of the datasets acquired(Walton, Addinall, Zeng, & Blunt, 2017).

The two datasets shown in Figure 5-13 and Figure 5-14 above are of the same bullet acquired using different measurement techniques and decomposed to the D5 wavelet band. In visual comparison, it can be seen that the surface texture does vary significantly between measurement systems. While there is a high level of noise in the Alias acquired dataset (Figure 5-14) between the six measurements taken, on the actual LEA surface there is no obvious presence of noise. Therefore, in correlation of the LEAs on the surface, the surface noise can be separated from actual surface texture, reducing skewing of correlation results (Figure 5-15).

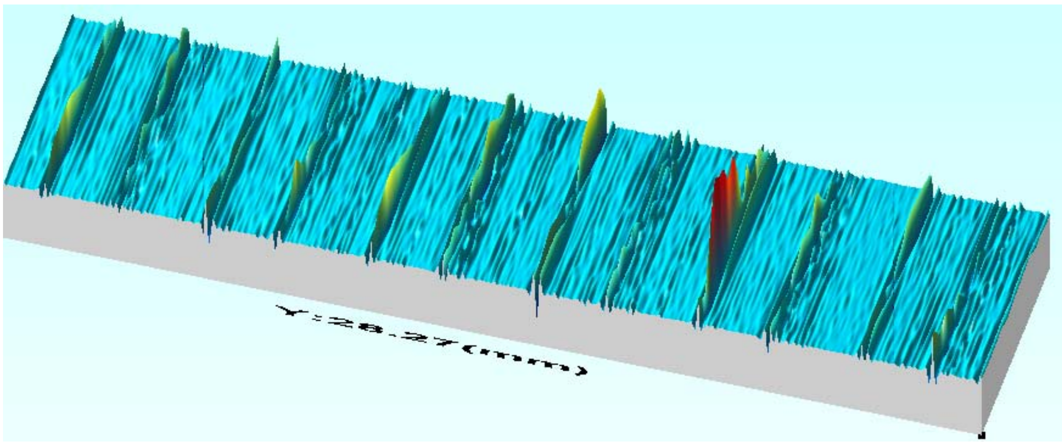


Figure 5-13: Picture of D5 bullet surface using Alicona acquisition

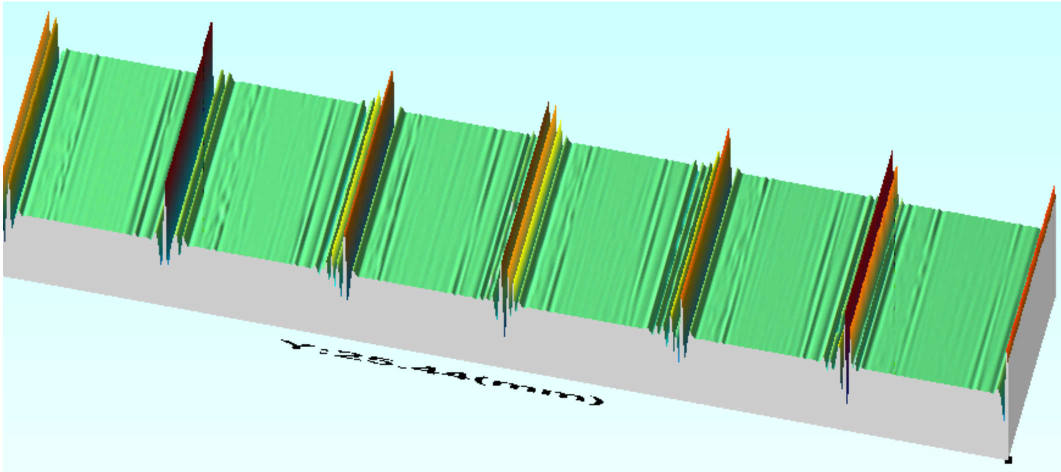


Figure 5-14: Picture of D5 bullet surface using Alias acquisition

LEA

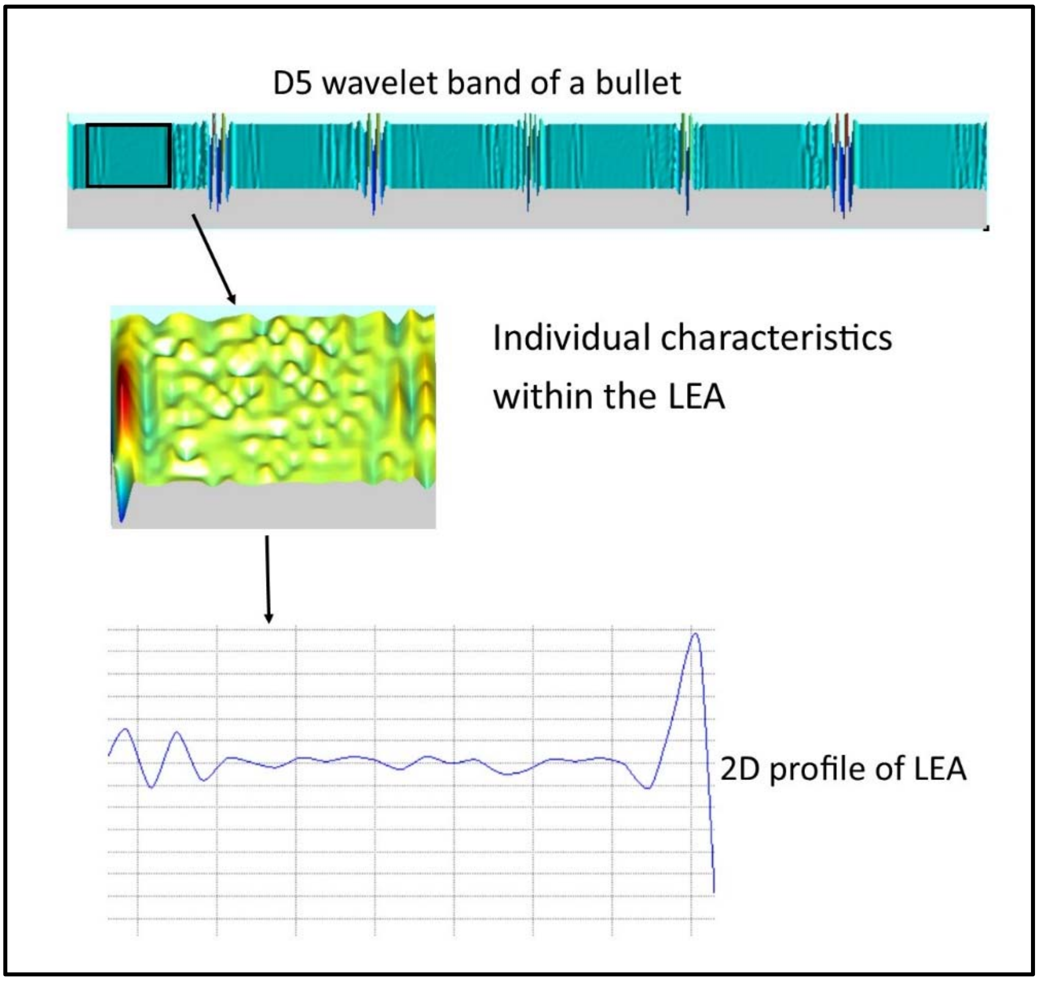


Figure 5-15: Surface texture of individual characteristics using Alias acquired D5 bullet surface

In the case of Alicona acquired datasets (Figure 5-13), it can be seen that there is a significant amount of noise present throughout the entirety of the dataset, thus obscuring actual surface texture relating to the individual characteristics. In the correlation of surfaces acquired using the Alicona system, this noise component of the surface will skew the results, hence correlation results have been found to be poor. It would appear that most of the instrument noise is contained within the D5 wavelet band, as visual comparison of the D6 wavelet band shows a dataset more comparable to those gained using the Alias system, in that while there are some areas which contain a large amount of noise, they can be segregated from the LEA for further correlation:

As can be seen in Figure 5-16, the levels of noise present in the D6 wavelet band are less than those seen in the D5 band, however in comparison to the Alias acquired surface there is still some level of waviness present. Both the Alicona software and the unwrapping scripts created remove coaxial features from acquired datasets, and the mounting of bullets in the rotational stage was achieved using a tight fit collar around the bullet. Due to this, it is unlikely that measurement errors were introduced through coaxial errors.

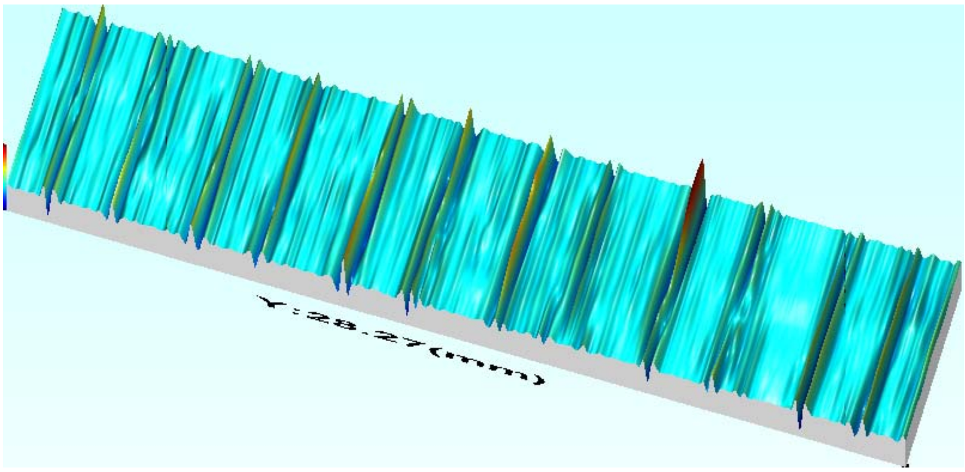


Figure 5-16: Picture of D6 wavelet band of an Alicona acquired bullet.

It is expected that the acquisition issues are due to the large z range in acquisition due to the overall form of the bullet, and the reflectance issues caused by this. As at any one time the measurement area of the Alicona will include measurement across the curve of a bullet surface, the difference in height values can cause lighting to vary across the surface. As the Alicona relies on neighbouring points to infer intensity/height information for each point, the variation in height differences can affect the acquired dataset, thus resulting in noise being imparted into the acquired surface. It is suggested that in future work, each LEA is measured separately, rather than as a 360° measurement to minimise these issues.
As described in chapter three, measurement artefacts are included in Alicona acquired datasets due to discrepancies in the standard deviation of light intensity in neighbouring points. Therefore, a difference in data quality can be seen within the two measurement systems.

Table 5-5 shows the hitlist results gained in correlation bullet datasets acquired using the Alicona, coupled with wavelet decomposition to gain the D6 frequency band:

Table 5-5: Hitlist results using the D6 wavelet band and Alicona datasets

Test object	Known match 1 Hitlist position #	Known match 2 Hitlist position #
1A	No match	3
2A	7	45
3A	122	127
4A	73	No match
5A	16	No match
6A	95	29
7A	203	13
8A	21	No match
9A	91	270
10A	No match	No match
11A	69	No match
12A0	67	139
13A	7	No match
14A	57	155
15A	No match	56
16A	No match	45
17A	11	No match
18A	11	142
19A	No match	72
20A	1	11
21A	1	11
22A	32	17
23A	No match	3

The hitlist results shown in Table 5-5 show that there is poor efficacy in using Alicona acquired datasets for the correlation of bullets. Only two known matches were placed as best match to the corresponding test object, and 13 of the 46 known matches were placed in the top 20. Since from a likelihood perspective this indicates that there is more chance of a known match appearing outside of the top 20 hitlist than within it, it is unlikely that this particular technique would be accepted as a form of evidence within a court of law. However, with optimisation of the technique this could change.

5.8.1 Comparison of Alias and Alicona results

Figure 5-17 shows a comparison of efficacy between the best results gained in bullet correlation using the Alias and Alicona systems. In the case of Alias correlation, it was found that the better results would differ between the D5 and D6 wavelet bands per each test object. As unreliable results were gained in the D5 band using the Alicona acquisitions, the better results were found using the D6 wavelet band.

While other frequency bands were not tested in use of Alicona acquired datasets, it is believed that the testing represents a fair comparison of correlation efficacy. This is due to the fact that the original lateral resolution of the measurement techniques was set to be the same, thus ensuring frequency bands would also be the same for both measurement techniques. As the only variable is then the measurement instrument, a comparison on data quality can be achieved.

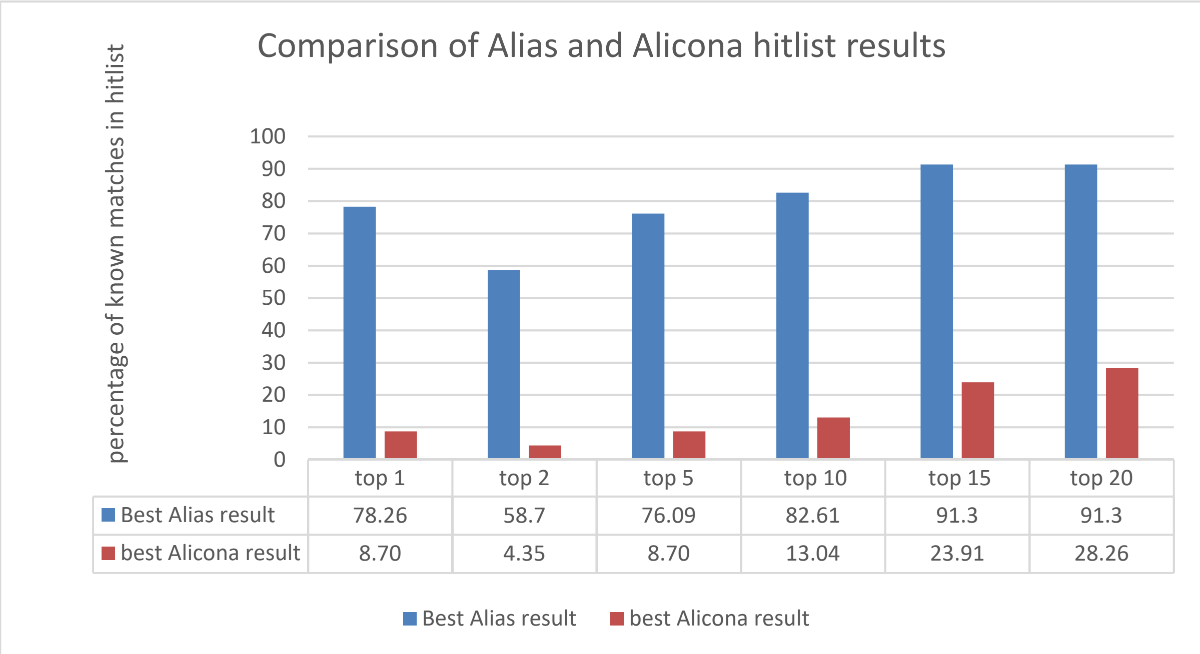


Figure 5-17: Comparison of hitlists using different measurement methods

It is apparent that there is a significant difference between the correlation hitlist results gained from using the two different measurement techniques, in which datasets acquired using the Alias system have performed significantly better. As there are some issues regarding the presence of measurement artefacts present on Alicona acquired datasets, it was expected that correlation would be less efficient than correlation using Alias datasets.

5.9 Comparison to previous systems

In previous studies conducted by Thomas (Thomas J. , 2011), two commercially available 2D identification systems were also used to gain hitlist information for the Odyssey collection. 2D measurements of all the Odyssey collection were acquired using both systems, and the same 23 test objects as previously described were correlated using each systems correlation software. The correlation hitlist results are as follows (Table 5-6):

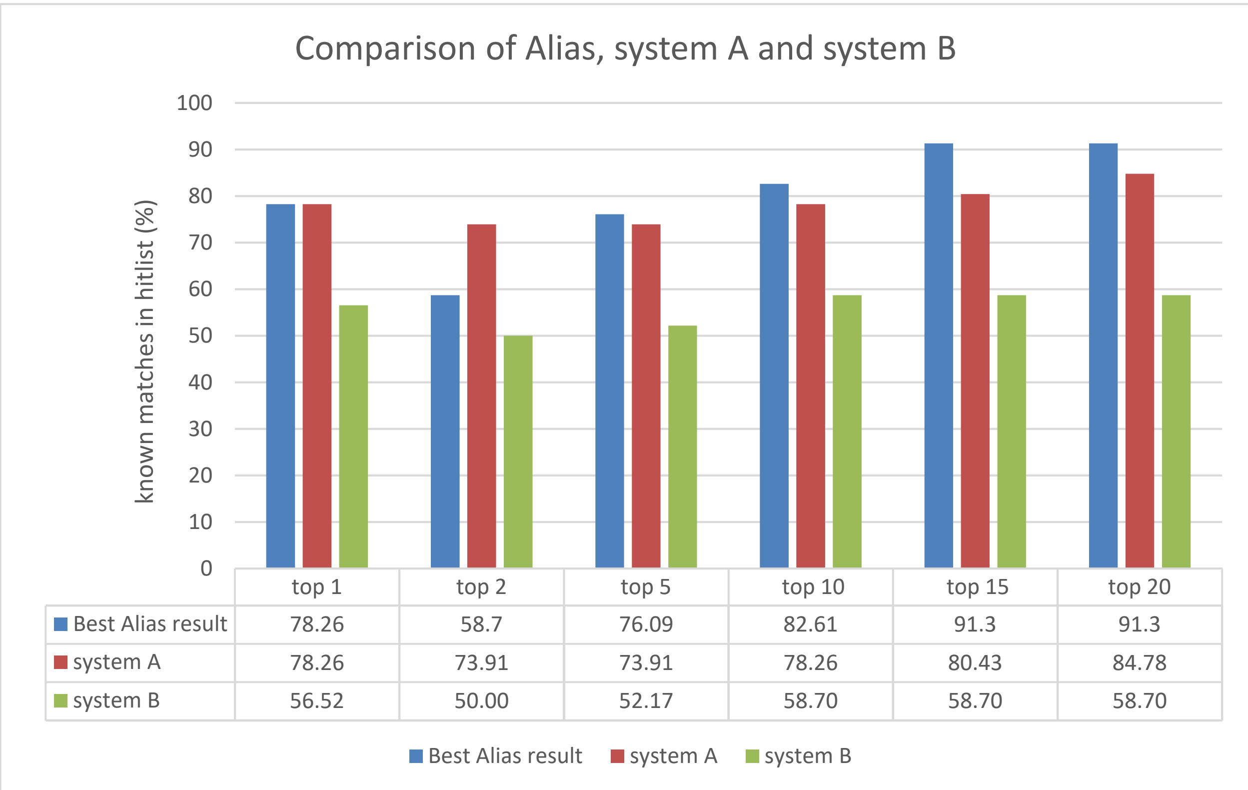
Table 5-6: Hitlist results gained using system A and system B

	system A		system B	
Bullet test object	Known match B Hitlist position	Known match C Hitlist position	Known match B Hitlist position	Known match C Hitlist position
1A	1	2	1	NO MATCH
2A	2	1	1	8
3A	1	2	2	1
4A	2	1	1	2
5A	1	2	1	2
6A	2	1	2	1
7A	1	2	1	NO MATCH
8A	2	1	2	1
9A	1	2	1	2
10A	NO MATCH	NO MATCH	NO MATCH	NO MATCH
11A	1	2	1	2
12A	1	17	NO MATCH	NO MATCH
13A	1	2	NO MATCH	3
14A	1	2	NO MATCH	8
15A	2	20	NO MATCH	NO MATCH
16A	1	2	1	2
17A	57	107	NO MATCH	NO MATCH
18A	1	2	1	2
19A	1	6	NO MATCH	NO MATCH
20A	78	NO MATCH	NO MATCH	NO MATCH
21A	6	11	NO MATCH	10
22A	2	1	2	1
23A	1	37	NO MATCH	NO MATCH

The results in Table 5-6 show that in most cases a good correlation efficacy is achieved, 18 of a possible 23 known matches have been placed as the top match to its test object, which is the same number of known matches in first position using the Alias system. However, in this case three of the known matches were not placed in the hitlist, and a further four were placed outside of the top 20.

In correlation using system B, the known match has been successfully placed in the hitlist, correlation results are positive, with all results in the top 10, however there are a lot of instances in which a known match is not being placed within the hitlist. As the results of correlation are unpredictable, it is unlikely that correlation results using this system would be used as forensic evidence, as a non-match result in this case does not necessarily mean that the bullet did not match, as it is also likely to be due to correlation issues.

The graph below in Figure 5-18 shows the comparison of efficacy in the 2D systems along with the Alias system. System B performed the worst, with the least number of known matches in each of the hitlist lengths.



	top 1	top 2	top 5	top 10	top 15	top 20
Best Alias result	78.26	58.7	76.09	82.61	91.3	91.3
system A	78.26	73.91	73.91	78.26	80.43	84.78
system B	56.52	50.00	52.17	58.70	58.70	58.70

Figure 5-18: Comparative graph of hitlist results gained from the Alias system along with system A and system B

In the case of system A, it was found to have the same number of known matches in the first position as the Alias system combined with wavelet correlation, and system A also surpassed Alias results with regards to number of known matches within the top two positions of the hitlist. However, the overall better efficacy of correlation can be found in the Alias/wavelet correlation methods, as the highest percentage of known matches within the top 10, top 15 and top 20 of the hit lists. These results show that the most consistent correlation results can be gained by using the Alias/wavelet technique. This is further shown in Table 5-7:

Table 5-7: number of known matches in varying hitlist lengths

	First position (of 23)	Top 2 (of 46)	Top 5 (of 46)	Top 10 (of 46)	Top 15 (of 46)	Top 20 (of 46)
Alias	18	27	35	38	42	42
System A	18	34	34	36	37	39
System B	13	23	24	27	27	27

5.10 Material composition

In the correlation of both firing impressions and bullets of the Odyssey collection, in some cases erroneous results were gained, in which a known match was placed too low in a hitlist of correlation results. In forensic evidence, these results would be considered false negatives, and could harm the prosecution hypothesis.

As a minority of the results were found to be erroneous, in which known matches were placed outside of the top 20 of the hitlist, it was determined that the pre-processing methods would not be the cause, as in most cases correlation was successful, therefore it is expected that a variation between the physical topography of the imparted toolmark may be causing issues in the correlation.

The Odyssey collection is comprised of bullets from seven eight different manufacturers. As already stated, in the original study it was found that using one manufacturer of cartridges would become very difficult in the timescale available to create the impression, and as such it was decided to accept that eight different manufacturers would have to be used to complete the Odyssey collection.

Therefore, in this study the aim is to ascertain whether the difference in materials used for cartridge manufacturer will influence the overall correlation efficacy in bullets. It is expected that as a softer material will deform more under contact from a tool compared to a harder material, that some variation in toolmark topography may be seen. Comparison of correlation results based on material composition will be made using Alias acquired datasets only, to ensure variations in measurement method do not affect the findings.

In Table 5-8 the material composition of each bullet test object and its known matches are shown. Copper, Zinc and Nickel percentages are given as they relate to the majority of the material composition.

Table 5-8: Material composition results of bullet surface using XRF

Object number	Element	%	Object number	Element	%	Object number	Element	%
1A	Cu	68.1	1B	Cu	68.8	1C	Cu	83.5
	Zn	30.8		Zn	30.1		Zn	9.35
	Ni	0		Ni	0		Ni	
2A	Cu	74.61	2B	Cu	75	2C	Cu	69.4
	Zn	24.95		Zn	24.95		Zn	29.5
	Ni	0		NI	0		Ni	0
3A	Cu	90.13	3B	Cu	90.61	3C	Cu	89.2
	Zn	9.25		Zn	9.18		Zn	9.72
	Ni	0		Ni	0		Ni	0
4A	Cu	89.16	4B	Cu	83	4C	Cu	84.1
	Zn	10.56		Zn	8.04		Zn	0
	Ni	0		Ni	0		Ni	0
5A	Cu	75.27	5B	Cu	77.27	5C	Cu	90.27
	Zn	24.19		Zn	22.55		Zn	9.57
	Ni	0		Ni	0		Ni	0
6A	Cu	90.75	6B	Cu	91.75	6C	Cu	68.5
	Zn	9.12		Zn	8.79		Zn	30.4
	Ni	0		Ni	0		Ni	0
7A	Cu	90.24	7B	Cu	90.84	7C	Cu	69.1
	Zn	9.37		Zn	9.09		Zn	29.7
	Ni	0		Ni	0		Ni	0
8A	Cu	99.6	8B	Cu	99.74	8C	Cu	86.3
	Zn	0		Zn	0		Zn	9.51
	Ni	0		Ni	0		Ni	0
9A	Cu	82.2	9B	Cu	83.8	9C	Cu	90.95
	Zn	10		Zn	0		Zn	8.96
	Ni	9.09		Ni	9.09		Ni	0
10A	Cu	79.3	10B	Cu	87.34	10C	Cu	91.53
	Zn	9.52		Zn	7.87		Zn	8.35
	Ni	10.2		Ni	4		Ni	0
11A	Cu	80.8	11B	Cu	84.4	11C	Cu	91.13
	Zn	9.79		Zn	10.5		Zn	8.13
	Ni	8.28		Ni	4.12		Ni	
12A	Cu	83.5	12B	Cu	82.8	12C	Cu	91.21
	Zn	10.2		Zn	10.1		Zn	8.63
	Ni	5.25		Ni	6.15		Ni	
13A	Cu	76.6	13B	Cu	74.4	13C	Cu	83
	Zn	23.2		Zn	24.97		Zn	8.67
	Ni			Ni			Ni	
14A	Cu	86.5	14B	Cu	87.01	14C	Cu	75.08
	Zn	6.72		Zn	8.72		Zn	24.07
	Ni			NI			Ni	
15A	Cu	70.1	15B	Cu	72.1	15C	Cu	90.93
	Zn	28.1		Zn	26.7		Zn	8.97

	Ni			Ni			Ni	
16A	Cu	68.9	16B	Cu	68.8	16C	Cu	75.6
	Zn	30		Zn	30		Zn	24.2
	Ni			Ni			Ni	
17A	Cu	69.3	17B	Cu	69.2	17C	Cu	74.23
	Zn	29.6		Zn	29.8		Zn	25.04
	Ni			Ni			Ni	
18A	Cu	85.5	18B	Cu	85.8	18C	Cu	68.5
	Zn	9.5		Zn	9.47		Zn	30.2
	Ni			Ni			Ni	
19A	Cu	80	19B	Cu	86.6	19C	Cu	69.1
	Zn	8.79		Zn	9.69		Zn	29.7
	Ni			Ni			Ni	
20A	Cu	71.2	20B	Cu	72.8	20C	Cu	80.5
	Zn	27.6		Zn	26		Zn	7.88
	Ni			Ni			Ni	
21A	Cu	69.3	21B	Cu	70	21C	Cu	76.86
	Zn	29.8		Zn	29		Zn	22.96
	Ni			Ni			Ni	
22A	Cu	69.2	22B	Cu	69	22C	Cu	76.18
	Zn	29.8		Zn	30		Zn	23.49
	Ni			Ni			Ni	
23A	Cu	90.33	23B	Cu	90.02	23C	Cu	89.95
	Zn	9.42		Zn	8.91		Zn	9.5
	Ni			Ni	8.91		Ni	

As can be seen in Table 5-8, there is a variation in material composition between some test objects and known matches, while others have a similar composition. This is further demonstrated in the graphs below (Figure 5-19-Figure 5-21).

To ascertain whether the differences in material composition will affect the efficacy of correlation, correlation results must be compared between known matches with a similar material composition and those with variations within the composition.

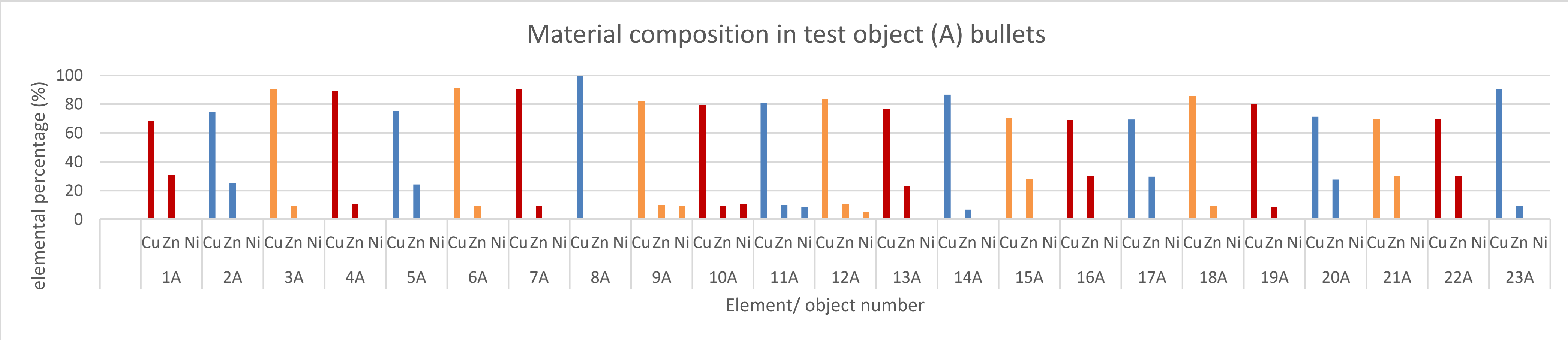


Figure 5-19: Material composition in test object bullets

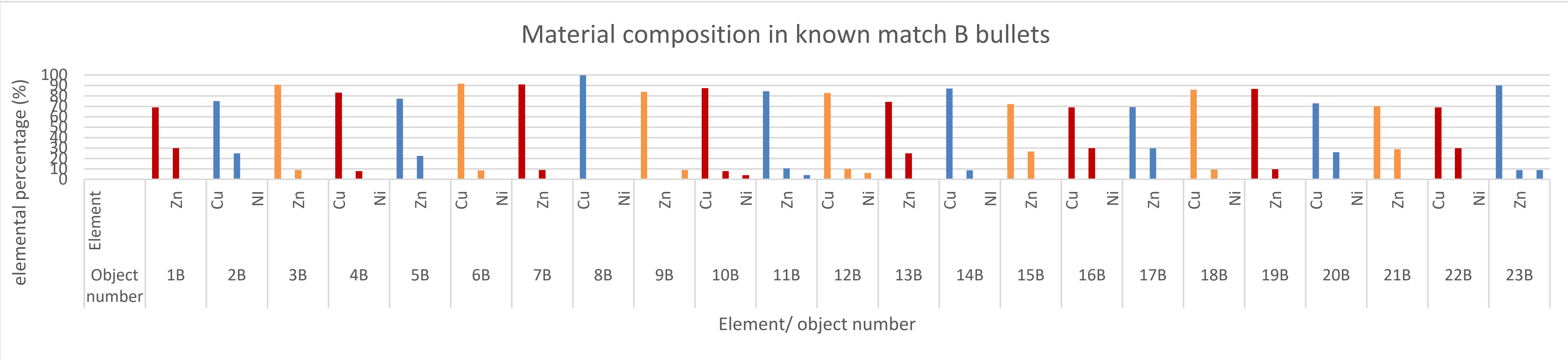


Figure 5-20: Material composition in known match B bullets

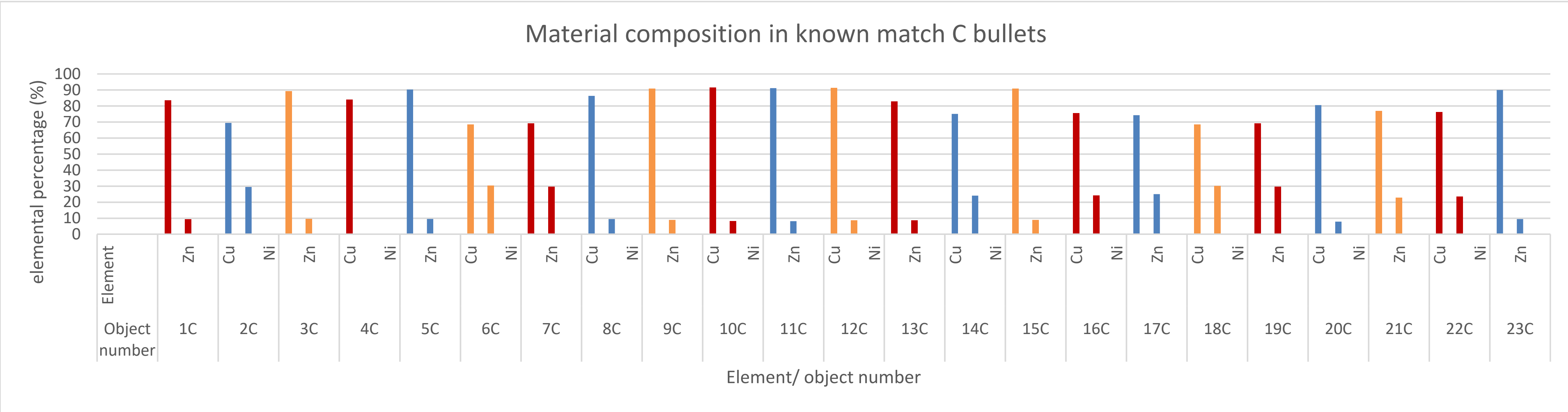


Figure 5-21: Material composition in known match C bullets

In Table 5-9, correlation results are compared with the frequency band in which the best correlation results were gained, along with the difference in the percentage of elements present.

Table 5-9: Correlation results compared to material difference

Test object	Known match B %	band	Difference in Cu %	Difference in Zn %	Difference in Ni %	Known match C %	band	Difference in Cu %	Difference in Zn %	Differe in Ni %
1A	97.3843	5	0.7	0.7	0	96.708	5	15.4	21.45	0
2A	98.3843	5	0.39	0	0	62.2644	6	5.21	4.55	0
3A	62.7466	6	0.48	0.07	0	54.0282	6	0.93	0.47	0
4A	96.1137	6	6.16	2.52	0	88.9108	6	5.06	10.56	0
5A	65.4914	6	2	1.64	0	87.62	6	15	14.62	0
6A	85.219	6	1	0.33	0	82.0352	5	22.25	21.28	0
7A	89.0228	5	0.6	0.28	0	76.3522	5	21.14	20.33	0
8A	81.4197	5	0.14	0	0	57.6963	5	13.3	9.51	0
9A	73.4887	5	1.6	10	0	58.5142	6	8.75	1.04	9.09
10A	99.2979	5	8.04	1.65	6.2	98.3948	5	12.23	1.17	10.2
11A	68.1352	6	3.6	0.71	4.16	58.31	6	10.33	1.66	8.28
12A	67.5115	5	0.7	0.1	0.9	88.6563	5	7.71	1.57	5.25
13A	77.386	5	2.2	1.77	0	82.9294	5	6.4	14.53	0
14A	83.3865	6	0.51	2	0	68.0916	5	11.42	17.35	0
15A	74.0408	5	2	1.4	0	76.2943	5	20.83	19.13	0
16A	74.1488	6	0.1	0	0	94.4828	5	6.7	5.8	0
17A	68.2893	5	0.1	0.2	0	71.4562	5	4.93	4.56	0
18A	60.2936	5	0.3	0.03	0	100	5	17	20.7	0
19A	53.7236	5	6.6	0.9	0	82.5621	6	10.9	20.91	0
20A	64.691	6	1.6	1.6	0	95.9262	5	9.3	19.72	0
21A	95.9097	5	0.7	0.8	0	93.8334	5	7.56	6.84	0
22A	92.042	5	0.2	0.2	0	97.0634	5	6.98	6.31	0
23A	28.2023	5	0.31	0.51	8.91	95.825	5	0.38	0.08	0

In comparison of correlation results compared to the material differences between known matches, it can be seen that there is no effect of material hardness on the correlation results gained. While it has been published in various articles that bullet material will have an effect on correlation, it would appear that in the case of the Odyssey collection, another variable is having a larger effect. Where a large difference in material composition can be seen, for example between 1A and 1C, 6A and 6C and 15A and 15C, good correlation results are still achieved. The largest variations in material difference are always observed between a test object and known match bullet C. However, as the Odyssey collection bullets were assigned at random, this can only be attributed to random chance. In Figure 5-22 below, a comparison of material difference and correlation result is further demonstrated:

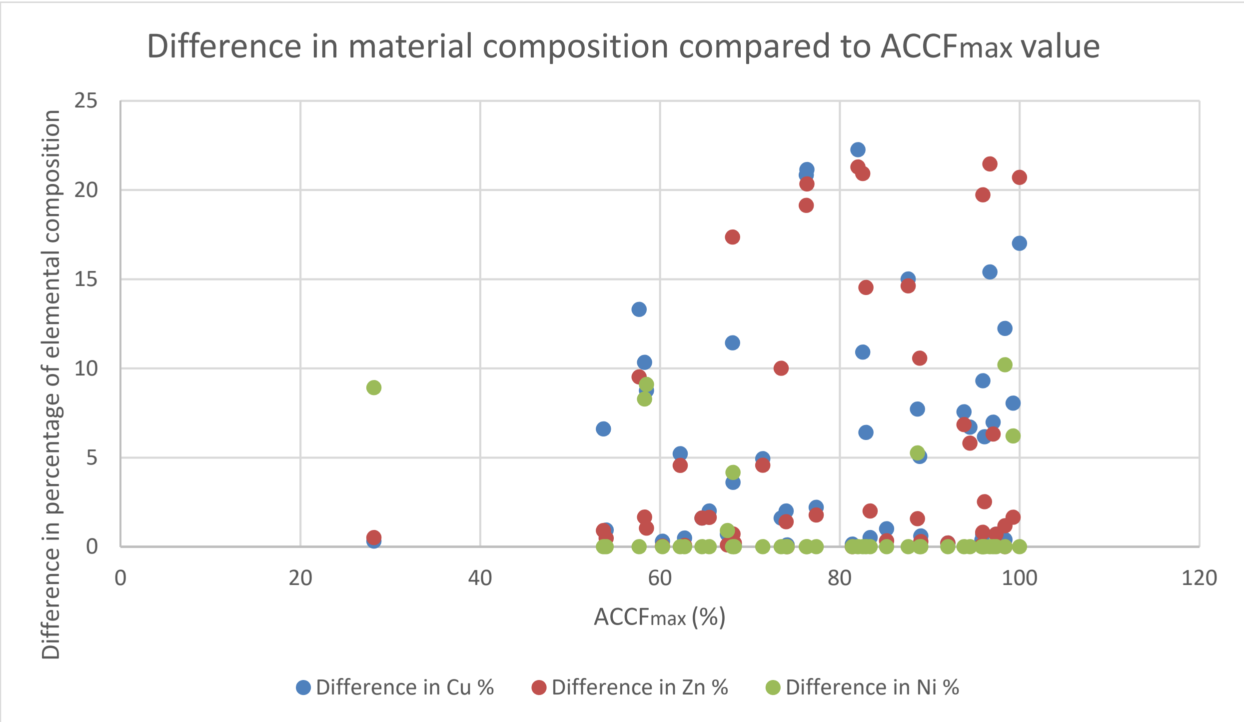


Figure 5-22: Graph of $ACCF_{max}$ results compared to difference in material composition

5.10.1 Material conclusions

The results gained in the material study show no link between the material of the bullet surface and correlation results achieved. In reviewing the frequency band in which the best correlation achieved compared to material differences (Table 5-9), there is no relationship between the frequency band and material difference. High correlation results with a low difference in material composition appear both in the D5 and D6 frequency bands.

As material difference does not appear to be the variable in bullets accounting for the most difference in correlation, other variables should be studied to understand the differences in correlation. As the fit of the bullet affects contact areas to the barrel, for example the possibility of GEA transfer to bullets, it is possible that the fit or angle of the bullet could result in small differences in striation transfer. Where there is a 'loose fit' between bullet and barrel, it is possible the pressure within the barrel is decreased and the transfer of toolmarks is affected.

5.11 Summary of findings

Firstly, a novel contribution to knowledge was gained in this study, as it was shown the correlation efficacy can be increased using areal datasets for correlation of bullets. Currently, there is no other published work that relies upon areal datasets for correlation, and all other published work relies either on 2D pattern matching or an averaged profile of the bullet topography for correlation.

It has been found in this study that using wavelet decomposition to segregate the salient information from other surface texture in bullet datasets would result in more successful correlation than using ISO standard Gaussian filtering techniques, currently used by NIST to confirm correlation (Song J. , et al., 2009). As correlation results gained are always high, regardless of whether the two bullets match, it is expected that salient data is not thoroughly segregated from other surface texture, thus resulting in skewed correlation results. In one paper (Ma, et al., 2004), NIST quoted a mean 2D CCF_{max} value of 99.29%, with a standard deviation of 0.26%. In this case, a singular profile was used for 2D cross - correlation between a Standard Reference Material bullet and the digital profile used for machining SRM. Considering the added variables in the Odyssey collection, and the correlation of areal surface rather than 2D profiles, it is understandable that a decreased efficacy has been found.

As Biasotti's research (Biasotti, 1959) on the use of Consecutively Matching Striae is widely accepted in the forensic field, NIST have also published research regarding the use of CMS with areal topography, which has proven to be successful (Chu, Thompson, Song, & Vorburger, 2013). In this research, NIST used a standard Gaussian filter with cut off values of 15μm and 250μm before applying edge detection techniques to identify striae present. The advantage of this research is that striations can be separated from random marks such as skid marks through angle detection, and therefore salient information can be easily separated. It is suggested that edge detection could be applied to the frequency bands used in this study, as making differentiation of striations more efficient may resolve issues with correlation occurring in more than one frequency band.

Secondly, in the use of wavelet decomposition of the originally acquired datasets, it was found that the use of datasets acquired by the Alias system would surpass correlation results gained in previous studies, and correlation results of Alicona acquired datasets. It is expected that in the case of the Alicona results, this is due to the data quality being reduced through measurement artefacts being present in the decomposed surface. As previous systems relied on 2D imaging and pattern matching, it is expected that unreliable results are due to various issues known in 2D imaging, including the shadowing of information on the surface, obscuring individual characteristics within the toolmark.

Finally, the success of Alias dataset correlation currently relies upon the correlation of LEA surfaces both in the D5 and D6 wavelet band. It is expected that this could be due to either the mother wavelet or the material of the bullet surface.

It is possible that the material hardness of the bullet surface will have an effect on the frequency of the individual toolmarks, for example a softer material will deform more, and may have some individual characteristics that relate to a lower frequency than those imparted into a harder material. Therefore, it is suggested that further work would include correlation using the D4 frequency band, and correlation based in a frequency band defined by the user. This would also ascertain whether or not defining the frequency band by the mother wavelet (i.e. the lateral resolution of the instrument) is affecting correlation results.

In previously published papers, both the IBIS Heritage and BulletTrax-3D systems were used to determine correlation efficacy in bullets. In one study (Brinck, 2008), the correlation efficacy of copper jacketed bullets and lead jacketed bullets were compared. In correlation of copper jacketed bullets, 90% of matches were found to be in the in top position with the remainder between position 11 and 20 using the IBIS Heritage system. Using the BulletTrax-3D system, 100% of copper jacketed bullets were found in the top position. In lead jacketed bullets, 30% of known matches were found in the top 10 positions with the remainder outside of the top 20 when using the IBIS Heritage system. In using the BulletTrax-3D system, 70% of lead jacketed bullets were found in the top position, while the remainder were within the top 10.

Based on the above research, there is an obvious discrepancy in the correlation efficacy based on the materials used. Even when correlating within the same material type, the variation in efficacy is large. When considering the Odyssey collection, the random allocation in material to known match groups will echo these findings.

As it has been reported that the IBIS Heritage and IBIS Trax-3D systems are interoperable in terms of correlation, the research also highlights a variation in correlation efficacy based on the measurement method. Where the same test group of bullets were measured on the different systems, similar correlation results would be expected when based in the same correlation algorithms, however it can be seen in this research that this is not the case. This corroborates the findings in this research that with the only variable being the measurement system used, correlation efficacy can vary substantially.

In another previously published paper (Nennstiel & Rahm, 2006), the IBIS Heritage system was used in the correlation of bullets, and a success rate of 50-75% of known matches within the top five was given. As these findings differ from those previously mentioned, it can be shown that correlation efficacy is determined based on the test set of bullets chosen.

Due to this, it can be determined that true correlation efficacy must be quoted on a large-scale test sample with variables expected within forensic casework. Therefore, further work suggested would be to increase the current database of measurements.

As the material study proved inconclusive with regards to effects on correlation, further work suggested would be to investigate the diameters of both fired and unfired bullets of various manufacturers to determine whether there is an overall deformation in bullets due to tightness of fit. As such a study would involve contact techniques (i.e. profilometry) that can be considered destructive, such studies cannot be completed using the Odyssey collection.

It is known that in the collection of Odyssey bullets, two methods of bullet capture were used, both water tank and cotton wool. It was not recorded which bullet capture method was used for each bullet. As differences in capture method can result in striations being damaged, either through contact with the walls of the water tank or the cotton fibres, it cannot be ruled out that obstruction of striations may have occurred during bullet capture. Further work is suggested where variables in capture method are kept to a minimum, i.e. only one is used, to ensure variables are not introduced after the bullet has been fired.

CHAPTER SIX

APPLYING THE BAYSIAN FRAMEWORK

As described in chapter one, there is currently a trend towards using the Bayesian framework to describe the likelihood of evidence within a court of law. While other evidence groups, such as DNA, rely on the use of likelihood ratios to present evidence in court, the same cannot be said for ballistic toolmark evidence.

Likelihood ratios allow for standardisation of evidence presentation, with defined verbal associations based on the strength of the evidence. Therefore, it would be in the best interest to create a path forward to ensure in future all evidence is based around the same presentation method. This would ultimately reduce misrepresentation of evidence in court.

The following chapter presents the likelihood ratios gained in each of the systems used for analysis within this thesis, along with a discussion on how the Bayesian Framework could be applied in the future.

6 Applying the Bayesian Framework

6.1 Introduction

Current standard practices in the presentation of forensic evidence within a court rely on applying the Bayesian framework of posterior odds based on the likelihood of the evidence being present in a crime scene, or the results gained in forensic examination of the evidence.

Multiplying the likelihood ratio with hypotheses presented by both the defence and prosecution results in values known as the posterior odds, with values indicating support of either the defence or prosecution hypotheses and the extent of the support.

The use of posterior odds in court means the laypersons of the jury gain a better understanding with regards to the value of the evidence being presented. However, studies have shown (Martire, Kemp, Sayle, & Newell, 2014) that laypersons on a jury can experience difficulty in understanding evidence when presented using probabilities and statistics. To this effect, it has been suggested using both a verbal and numerical scale when presenting evidence to minimise the under-valuing of evidence presented in a probabilistic format.

While the Bayesian framework is becoming more accepted in other forensic disciplines, the same cannot be said for firearm toolmark examination. It is common for a firearm toolmark examiner to use categorical or inconclusive statements, thus not taking into account prior odds of the evidence. It is believed the cause of this is a worldwide lack of training in the Bayesian approach in evidence, and research has suggested that both training and a standardisation in verbal scales of evidence would move firearm toolmark evidence to a Bayesian approach (Bunch & Wevers, 2013). In this chapter, the strength of the results gained using methods presented in this thesis will be discussed with regards to the Bayesian framework, to assess the strength of the evidence should they be presented in a court of law. This will indicate whether the methods are feasible for the presentation of forensic evidence. The results are shown for comparative purposes across the various systems only and would require further research to ensure the odds were applicable in a court of law.

To determine the posterior odds, and therefore the strength of the evidence, both the prior odds and the likelihood ratio must first be stated as per equation 1:

$$\frac{\Pr(Hp|E)}{\Pr(Hd|E)} = \frac{L\,(E|Hp)}{L\,(E|Hd)} \times \frac{\Pr(Hp)}{\Pr(Hd)}$$

$$posterior\ odds = likelihood\ ratio\ \times prior\ odds$$

(Equation 1)

Where assigning likelihoods in the case of both the likelihood ratio and the prior odds is a scale from 0-1. 0 represents an impossibility and 1 represents a certainty.

6.1.1 Prior odds

The prior odds in the case of forensic toolmark examination can be kept simple, either the tool did cause the toolmark or it did not cause the toolmark –i.e. another tool was used. Therefore, the prior odds in the case of a shooting can be described as follows:
Prosecution hypothesis (Hp): The questioned firearm was used to fire the cartridges found at the crime scene.
Defence hypothesis (Hd): The questioned firearm was not used to fire the cartridges found at the crime scene.
As prior odds (Equation 9) do not take into account any other information regarding the specifics of each case, at this point it can be said that the Hp and Hd are as equally likely. Therefore, they are each assigned the probability level of 0.5.

Thus, prior odds equal:

$$\frac{\Pr(Hp)}{\Pr(Hd)} = \frac{0.5}{0.5} = 1$$

(Equation 9)

6.1.2 The likelihood ratio

The likelihood ratio takes into account the evidence gained during forensic investigations. Therefore, the likelihood value assigned relates to the probability of the evidence found, given the stated hypothesis.
In a forensic examination, test firing will be completed using the firearm that is suspected to have been involved in the crime. The test fired bullets or cartridge cases will then be compared to evidence found at the crime scene to ascertain whether the firearm was used in the crime.
If forensic examinations conclude that it is beyond reasonable doubt[7] that the test fired rounds and crime scene evidence were fired using the same firearm, then the following would apply to the likelihood ratio, described in research conducted by Bunch et al (Bunch & Wevers, 2013), as follows: Equation 11:

$$LR = \frac{800}{1} \times \frac{0.5}{0.1} = 4000$$

(Equation 10)

---

[7] The strongest support that can be assigned is "beyond reasonable doubt". Absolute certainty in a hypothesis should not be assigned as it is not possible, and misleading.

As the calculation of the likelihood ratio results in a large positive integer, this shows that the evidence supports the prosecution hypothesis, Hp. A small positive integer would show support for the defence hypothesis, Hd, with the integers being a sliding scale of support.

6.1.3 Posterior odds

Posterior odds are simply calculated from the multiplication of the prior odds and the likelihood ratio, thus in this example posterior odds would equal 800x5 = 4000. Using verbal association, an expert witness in court would describe these results as "strong support" for the prosecution hypothesis, i.e. the firearm was used to fire the cartridges found on the crime scene. The verbal association is in accordance with guidelines issued by the European Network of Forensic Science Institutes (ENFSI) (European Network of Forensic Science Institutes, 2010).

6.2 Calculation of the likelihood ratio using the Alias system

The main factor in the calculation of posterior odds is the likelihood ratio, as the efficacy of the testing methods used to gain the evidence will affect the likelihood value calculated. For example, a method with high efficacy will offer stronger support than one that is prone to errors. Therefore, to apply the Bayesian framework approach to the results illustrated within the thesis, the efficacy must be discussed.

When correlating firing pin impressions, it was found that the most efficient method used in this thesis was the acquisition of data using the Alias system, followed by pre-processing steps of least squares levelling and Robust Gaussian filtration with cut off values of 75µm and 450µm. Correlation was then achieved using the Areal Cross-correlation Function.

In the testing of the above firing pin correlation method, 8,558 correlations were completed, 44 of which were known to be true matches, contained within a total of 22 separate hitlists. For this study, the author decided to define the likelihood with regards to a known match appearing within the top 10 of a hitlist, as the author believes for an identification system to be considered accurate, known matches should not appear outside the top 10 of a hitlist.

Of the 44 known matches (two known matches per each of the 22 test objects), 34 have been placed within the top 10, giving 77.27% of known matches. Therefore, in terms of likelihood, a value of 0.77 can be assigned to the chance of a known match appearing within the top 10 results of a hitlist.

Should the prosecution hypothesis be stated as 'the firearm was used to fire cartridges found at the crime scene, then a 0.77 likelihood value can be applied to the hypothesis. Therefore, the counter hypothesis must have a likelihood value of 0.23.

Using these values, posterior odds can be calculated using (Equation 1) to gain a value of 3.35. While this shows that using the Alias system will support the hypothesis in the case of evidence being placed in the top 10 of a hitlist, the support is lower than expected for a forensic technique (European Network of Forensic Science Institutes, 2010).

Using the same hypotheses and hitlists created in this study, the evidential value of bullet correlation can also be calculated. Using the better results gained (combined D5 and D6 hitlists of Alias acquired datasets) 38 of a possible 46 known matches were placed in the top 10. This gives an 82.6% chance of a known match appearing within the top 10 Therefore, the prosecution hypothesis can be assigned a likelihood value of 0.83, with the opposite hypothesis therefore being assigned a value of 0.17. Using the calculation for posterior odds will therefore result in a value of 4.88. Once again, while the value is supportive, it is not as high as would be expected in a forensic technique.

Based on the ENFSI guidelines for verbal association of the likelihood ratio (European Network of Forensic Science Institutes, 2010) (Table 6-1) results gained in this study can provide weak support:

Table 6-1: Verbal association of likelihood ratio

Likelihood ratio	Verbal associations (two recommended options)
1	Does not support one hypothesis over the other Evidence provides no assistance in addressing the issue
2-10	Provides weak support for the first proposition (prosecution hypothesis) relative to the defence hypothesis Findings are slightly more probable given one proposition relative to the other
10-100	Provides moderate support for Hp Evidence more probable given Hp than Hd
100-1000	Provides moderately strong support Evidence is appreciably more probable
1000-10,000	Provides strong support Evidence is much more probable
10,000-1,000,000	Provides very strong support Evidence is far more probable
Over 1,000,000	Provides extremely strong support Evidence is exceedingly more probable

6.3 Calculation of posterior odds using 2D systems

As previously stated, the Odyssey collection was acquired and compared in a previous doctoral thesis using two commercially available 2D comparison systems. To corroborate the findings of increased efficacy using the Alias measurement system combined with $ACCF_{max}$ correlation, the posterior odds of the two 2D systems will also be calculated for comparison.

6.3.1 System A

Using the same hypotheses as detailed above, a prior odds value of 0.5/0.5= 1 is gained. The efficacy of system A for placing known match cartridge cases in the top 10 was found to be 23 of 44 possible known match cartridge cases, thus 52%. Therefore, using 0.52 as the value for Hp, and 0.48 as the Hd value, a posterior odds value of 1.08 is gained.
In terms of the bullet correlation efficacy, 36 of a possible 46 known matches were placed in the top 10 of the hitlist, thus 78%. Therefore, the likelihood ratio in this case will be 0.78/0.22, and the posterior odds is calculated as 3.54.

6.3.2 System B

Once again using the same hypotheses and prior odds (0.5/0.5 = 1) the efficacy of system B can be calculated.
Cartridge case correlation resulted in 26 of a possible 44 known matches being placed in the top 10 of the hitlist, or 59%. Therefore, using likelihood values of 0.59/0.41, a posterior odds value of 1.43 is achieved.
In bullet correlation, 27 of a possible 46 known matches were placed in the top 10, which equates to 59% efficacy. Thus, a likelihood ratio of 0.59/0.41 is achieved, and posterior odds equals 1.43.

6.4 Conclusions

As it can be seen in the above results, the Bayesian approach shows that using correlation methods set out in this thesis will result in correct support from the posterior odds, i.e. when the evidence supports the prosecution hypothesis, the posterior odds value reflects this. However, the strength of this support should be increased before the methods are applied as a true forensic technique, as values of 3.55 and 4.88 indicate only weak support (European Network of Forensic Science Institutes, 2010).
There are a number of reasons that the posterior odds value is showing weak support. Firstly, as discussed in a previous chapter, a difference in cartridge manufacturer could be affecting the correlation results. As it would be expected for a forensic examiner to acquire test fires based on the manufacturer of cartridges found at the crime scene (Riva &

Champod, 2014), it is expected that erroneous results due to cartridge manufacture will be minimised in such scenarios. Therefore, it would be a sensible approach for future work to use only cartridges of the same manufacturer to assess the strength of the evidence and help determine variation and repeatability in toolmarks without variables introduced by using several cartridge manufacturers.

Secondly, for a true statistical evaluation of the strength of the evidence, a large sample size should be used in calculations. In the Odyssey collection, there are 44 known matches within the cartridge case collection and 46 within the bullet collection. The more tests carried out the more relevant statistical calculations will be, and therefore it is recommended that the research is expanded.

Finally, each hitlist contained 2 known matches and 388 non-matches, and therefore any results gained are only relevant to a database of the same size. It would be recommended to increase the database size so that more non-matches are included in each hitlist, to ascertain how posterior odds are affected.

In comparing the correlation results achieved using the 2D systems, it can be seen that there has been an increase in the strength of the evidence, brought on by the increased efficacy of the more advanced Alias system. Therefore, it can be shown that a positive step has been achieved with correlation efficacy, and with further work the strength of evidence can be increased.

CONCLUSIONS

Throughout the thesis, methods have been presented for the measurement and correlation of ballistic toolmark evidence, along with discussion of toolmark variability and the application of the Bayesian Framework.

The following chapter details the conclusions gained in each study and allows for a discussion of findings compared to previous research.

Firing pin impression correlation

In firing pin correlation, firstly it was found that using ISO standardised filtration for correlation resulted in correlation values that were unable to distinguish between a known match a non-match. ISO standard robust Gaussian filtration has previously been used in studies by NIST to determine the level of agreement in Standard Reference Material, in which a set of standard cartridge cases were CNC machined so that all toolmarks would be exactly the same (Vorburger T. V., et al., 2011). In this study, it was found that a high level of agreement was found in correlation, which is to be expected due to the exact machining of each cartridge case. When the same filtration techniques were applied to the Odyssey collection, the correlation results showed an inability to differentiate between known matches and non-matches, and therefore it was concluded that the filtration techniques could not be used in a true forensic identification system.

In previously published papers, variations in cut off values using Robust Gaussian filtration were used as shown in Table A. Quoted correlation efficacy was better using these methods, however it is important to note that the test database included less variables than found in the Odyssey collection. Current research at NIST is based around the Congruent Matching Cell technique (Song, 2015), and quoted differentiation between known matches and non-matches is high. However, in these studies, a cell must have an $ACCF_{max}$ value of only 25% to be considered a congruent cell, which can be considered a low correlation value. Test firearms used in this study were machined using bead and sand blasting techniques, thus ensuring transferred toolmarks would be random in nature. Once again, test databases of firing pin impressions contained less variables than those found in the Odyssey collection. Through the testing of various filtration techniques, it was found that applying least squares levelling to the primer cap surface and using a Robust Gaussian filter with cut off values of 75μm and 450μm would result in successful correlation, in which it was possible to differentiate between known matches and non-matches in most cases. Table A below details the difference in hitlist position of the known matches between each system. It can be shown that in each hitlist length, using Alias acquired datasets with $ACCF_{max}$ correlation increases the number of known matches found. 2D systems used in the previous study (Thomas J. , 2011) outperformed both Alicona acquired datasets with $ACCF_{max}$ correlation and the Alias correlation systems. This highlights that the quality of dataset and efficacy of correlation method is vital in areal correlation techniques.

Table A: Differences in known match accuracy in each system

	First position (of 22)	Top 2 (of 44)	Top 3 (of 44)	Top 5 (of 44)	Top 10 (of 44)	Top 15 (of 44)	Top 20 (of 44)
Alias and $ACCF_{max}$	19	25	28	29	34	35	35
System B	16	20	21	22	26	29	30
System A	14	18	20	20	23	25	26
Alicona and CCF max	8	12	12	15	16	21	23
Alias correlation	0	0	1	2	4	6	8

Using the successful pre-processing techniques, a full correlation test was carried out in which each of the test objects within the Odyssey collection were correlated against 387 cartridge cases that were fired from another firearm (non-matches), and two cartridge cases that were fired from the same gun (known matches). Using these testing methods allowed a direct comparison of efficacy to systems used in a previous study. It was found in comparison of hitlist results from the two older systems that use 2D pattern matching, that correlation using areal surface topography can increase correlation efficacy, in cases where high fidelity data acquisition has been used. Therefore, it can be shown that a shift from 2D to areal acquisition and correlation of ballistic toolmarks is achievable and will result in systems with higher efficacy. It has also been shown that objectivity can be potentially removed from ballistic toolmark identification with the use of mathematical correlation, which removes user input to ascertain the degree of similarity and allows comparison to be based on mathematical correlation, based on objective standards.

In comparison of areal data acquisition systems, it was found that there is a large variation in correlation results dependant on the acquisition system being used. While datasets acquired using the Alias pOCT interferometer allowed for accurate correlation of the Odyssey collection, it was found that using focus variation techniques resulted in poor correlation. It is expected that this is due to the fact that the Alicona focus variation technique is more likely to impart measurement artefacts into the measured surface that are within the same frequency bandwidth of the individual characteristics, thus pre-processing techniques are not able to perform successful filtration of all measurement artefacts. In terms of the correlation study, the term accurate is used to describe the presence of a known match within the top ten best matches of the hitlist, with more accurate results meaning more known matches appearing within the top 10.

In previously published results, the correlation efficacy of the IBIS Heritage system has been discussed.

As focus variation relies on the standard deviation of illumination between the point being measured and its neighbouring points (Leach, 2010), differences in reflective properties across the surface can result in measurement artefacts being imparted into the measured dataset. It was also found that measurement using the Alias system is considerably quicker, with measurements taking around five minutes per primer cap compared to 15 minutes when using the Alicona system. Taking all these factors into account it can be concluded that the most accurate correlation method can be achieved by using the Alias system for the acquisition of datasets, combined with least squares levelling and a Robust Gaussian filter with cut off lengths of 75µm and 450µm. The Areal Cross-Correlation Function was then used to confirm that the pre-processing techniques were accurate and an objective hitlist of each test object within the Odyssey collection could be created.

In comparison to previous studies using the IBIS Heritage system, in some cases methods used in this study have outperformed previously published results, while in others this is not the case. One previous study (Tulleners, 2001), quoted hitlist results of known matches to be: 26% found in top position, 30% in the top three, 32% in the top five and 42% in the top ten. In this study, 77% were found in the top ten using Alias measurements and $ACCF_{max}$ correlation, indicating an increased efficacy in results gained. In another study (George, 2004), 48% of known matches were found within the top ten of the hitlist, further corroborating these findings. However, in two other studies (De Kinder, Tulleners, & Thiebaut, 2004; Nennstiel & Rahm, 2006), 78.1% were found in the top ten, and 75%-95% were found in the top five. These results demonstrate a dependence on the test database used with regards to quoted efficacy. A direct comparison cannot be made until variables in the test database are minimised.

The study of material variance in primer caps resulted in no relationship being found between the correlation efficacy and material difference. Previous studies have alluded to the fact that there will be a variance in correlation when cartridges by different manufacturers are used in test fires, however a material study has not been completed. It was found that a difference in manufacturer may change the chamber pressure and potential flowback of the firing pin impression (Cork, Rolph, & Meieran, Ballistic Imaging, 2008). The differences in volume parameters do corroborate with these findings. Further work suggested would be to create more test fires, in which the same manufacturer is fired several times, rather than singular instances of cartridge manufacturer as seen in the Odyssey collection. The results would ascertain toolmark variability due to a difference in manufacturer used.

Due to the fact that manual detection of the firing pin impression was used for this study, not all subjectivity in the techniques has been eradicated, as any user input can result in observer bias or errors in judgement of the regions of interest. However, it is believed that this can be overcome in further research, in which automatic detection of the firing pin impression could be applied to methods used in this study.

It is expected that this would result in a measurement system that could reduce subjectivity in forensic techniques, and minimise time needed for toolmark examiners. Due to the ability to minimise the list of potential matches to an unknown firing pin impression, toolmark examiners would be able to concentrate efforts in a smaller pool of potential matches, thus encouraging a more time and cost -efficient system.

Bullet correlation

Firstly, within this research it was found that the areal topography of a fired bullet can be utilised for the correlation of ballistic toolmarks. Current techniques rely on either 2D pattern matching or the correlation of an averaged 2D profile of the areal topography (Chu, Song, Vorburger, & Ballou, 2010; Chu, Thompson, Song, & Vorburger, 2013). It is therefore believed that this research is the first example of true areal topography of toolmarks imparted into bullets being used for correlation.

During testing of the pre-processing techniques, it was found that the use of wavelet decomposition would lead to the best separation of salient data from other surface features. In most cases, it was found that the D5 wavelet frequency band would effectively separate the topography, however in some cases it was found the D6 band would perform more effectively. As the highest frequency of the mother wavelet is determined by the lateral resolution of the measurement system, it was not possible that the lateral resolution would vary between measurements using the Alias system, due to the fact that the Alias system is set to use a consistent lateral resolution. Therefore, the shift in wavelet band cannot be attributed to a change in the highest frequency of the mother wavelet. It was then expected that a change in material composition in some of the bullet casing materials were causing a shift in the effective correlation band. For example, a softer material may deform more as it passes through the barrel, resulting in striations becoming wider and therefore being present in a different wavelet band. The material composition of each test object bullet and its known matches was investigated to determine whether it could have an effect on the efficient correlation wavelet band.

Further work would include correlation using the D4 frequency band, and correlation based in a frequency band defined by the user. This would ascertain whether or not defining the frequency band by the mother wavelet (i.e. the lateral resolution of the instrument) is affecting correlation results. It is possible that being able to define a frequency band which overlaps some frequencies contained within both the D5 and D6 wavelet band would allow for good correlation efficacy without the added variable of correlation in two different frequency bands. NIST research has shown that the use of automatic CMS identification is successful in areal datasets of bullets (Chu, Thompson, Song, & Vorburger, 2013), where edge detection is used to identify striae. The advantage of this research is that striations can be separated from random marks such as skid marks through the use of angle detection. Therefore, salient information can be easily separated. It is suggested that edge detection could be applied to the frequency bands used in this study, as making differentiation of striations more efficient may resolve issues with correlation occurring in more than one frequency band.

It was found that the material composition was not the cause of the shift in the effective correlation band. Where there was a higher variance in material composition between a test object and its known match, it was found that correlating in different wavelet bands made little difference to the accuracy. Where there is a variation in material composition, the correlation does become very varied, with some performing well and others unable to confirm the match. This suggests that further work needs to be conducted into the material hardness variance and the variance of unfired bullet surfaces across different material and manufacturers. An investigation of both fired and unfired bullets of various manufacturers should be completed, to determine whether there is an overall change in deformation due to differences in contact between the bullet and the barrel.

It is expected that the shift in correlation efficacy between the D5 and D6 wavelet band must therefore be due to a physical property of the barrel and the rifling marks within it. The known match test groups 3 (i.e. test object 3A and known matches 3B and 3C), 5 and 11 all gave a non-match result in correlation using the D5 wavelet band and matched each test object within the top 10 using the D6 wavelet band for correlation. In each case, the firearm used to fire the bullets was manufactured by Beretta. However, as 109 of 196 of the firearms used were manufactured by Beretta, it cannot be concluded that the wavelet band is specified by the manufacturer. It is possible that a difference in manufacturer in a subset of barrels, for example a different technique or tool used in manufacturer, would result in a change of frequency in individual characteristics imparted into bullets, hence a shift in the effective correlation band. As the manufacturing process of each firearm used in the Odyssey collection was not recorded, it is suggested that further research is carried out on this hypothesis using firearms selected on how they were manufactured.

In comparison of correlation results gained in using older systems (Thomas J. , 2011) based on 2D pattern matching and areal surface correlation, it can be seen the efficacy has been increased in the use of areal correlation. The number of non-matches has been decreased, and more of the known matches have been placed in the top five of the hitlists, as can be seen in Table B. Therefore, through this direct comparison of the correlation result of the Odyssey collection it has been demonstrated that correlation efficacy can be increased in using advanced measurement methods and mathematical correlation. Combined with the ability to remove objectivity from the process, it is believed that such methods can be validated for use in forensic casework.

Table B: Differences in known match bullet accuracy in each system

	First position (of 23)	Top 2 (of 46)	Top 5 (of 46)	Top 10 (of 46)	Top 15 (of 46)	Top 20 (of 46)
Alias	18	27	35	38	42	42
System A	18	34	34	36	37	39
System B	13	23	24	27	27	27

The comparison of advanced measurement systems showed that there was a large variation in the quality of correlation results between pOCT interferometry and focus variation techniques. Firstly, there was a significant difference in scan time between the two systems. A 360° axial measurement using the Alias system will take around five minutes, while using the Alicona system will take around 45 minutes per scan. Combined with the need to unwrap data from a point cloud into an areal dataset, the overall time taken for one bullet measurement will be around 1 hour 45 minutes, without development of a measurement specific software comparable to that of the Alias instrument. When considering the high load of casework in some forensic laboratories, it would be inefficient to utilise a system that takes a large amount of time to complete one scan. Therefore, it is believed for those reasons the Alicona measurement method may not be suitable for routine measurement of bullets.

In the comparison of correlation efficacy between the two systems, it can be seen that the Alicona system is lacking. Not only did the correlation results using the Alias system outperform Alicona results, the results gained using older 2D systems were also better. Using the Alias system, 91.3% of the known matches were placed in the top 20 of the hitlists, 28.26% were placed in the top 20 using Alicona acquired datasets, and systems A and B resulted in 84.78% and 58.7% in the top 20 respectively.

Previously published results (Brinck, 2008) used both the IBIS Heritage and BulletTrax-3D systems to determine correlation efficacy in bullets. In one study, the correlation efficacy of copper jacketed bullets and lead bullets were compared. In correlation of copper jacketed bullets, 90% of matches were found to be in the in top position within the remainder between position 11 and 20 using the IBIS Heritage system. Using the BulletTrax-3D system, 100% of copper jacketed bullets were found in the top position. In lead jacketed bullets, 30% of known matches were found in the top 10 positions with the remainder outside of the top 20 when using the IBIS Heritage system. In using the BulletTrax-3D system, 70% of lead jacketed bullets were found in the top position, while the remainder were within the top 10.

In a paper published by NIST (Ma, et al., 2004),a mean CCF_{max} value of 99.29%, with a standard deviation of 0.26% was reported. In this case, a single profile was used for 2D cross -correlation between a Standard Reference Material bullet and the digital profile used for machining SRM. As it has been reported that the IBIS Heritage and IBIS Trax-3D systems are interoperable in terms of correlation, the research also highlights a variation in correlation efficacy based on the measurement method, similar to the findings in these studies where Alias and Alicona acquired datasets will result in different correlation outcomes.

In visual comparison of datasets acquired using the Alias and Alicona systems, it could be seen that measurement artefacts were present in Alicona datasets, even in decomposed wavelet bands which contained surface texture relating to individual characteristics of the toolmarks. Because surface artefacts are included in correlation of Alicona datasets, it is unavoidable that they will cause a skewing of the correlation results. This shows that the Alicona system is prone to noise/artefact generation issues while the Alias system is not. As with cartridge case correlation, this is due to the Alicona relying on relative intensity difference of neighbouring points to ascertain height information.

Overall correlation conclusions

Overall, it can be concluded that in correlation of the Odyssey collection efficacy can be increased in the use of advanced correlation techniques, where data acquisition is of high fidelity. Shifting from 2D to areal correlation results in higher density datasets, as information on the x, y and z placement of each data point is recorded. Due to this there is a large increase in surface information that must be filtered before correlation of individual characteristics can be completed without being skewed by other surface textures. It has been shown that individual characteristics can be separated efficiently from other surface information, however measurement artefacts imparted into the dataset cause issues which pre-processing of data cannot fully remove. Therefore, it is advised that measurement systems that rely on interferometry rather than focus variation/microscopy techniques will result in increased efficacy in a correlation system. Consequently, data quality and sample consistency are critical in assessing advanced algorithms.

Using advanced measurement systems with high data fidelity, it has been shown through direct comparison of hitlists of the Odyssey collection that correlation efficacy is increased with regards to 2D comparison systems. The shift from 2D to areal will ultimately remove subjectivity from the system and eradicates lighting issues causing such effects as canopy shadowing from the measurement process.

In the use of Areal cross-correlation, it has been shown that the hitlist length can be decreased by using the scale difference, D_s algorithm, to determine bad matches due to a significant scale difference between the two surfaces. It was found in these studies that a D_s value of over 100 would be effective for the removal of bad matches, which was found to corroborate with other studies (Cadevall & Schwarz, 2013). $ACCF_{max}$ is then used to rank matches from best to worst, which differs from traditional rank order approach as a mathematically derived percentage match is given, rather than a rank score with little relative meaning.

Bayesian likelihood

Using the Bayesian framework of likelihood, it was found that using the data acquired using the Alias system combined with $ACCF_{max}$ correlation would result in a system that can only weakly support the evidence. Using the verbal association for posterior odds, it can be seen that in the worked example within this thesis, limited support of the prosecution hypothesis would be gained should a match be found in the system. As other forensic techniques such as DNA and glass evidence currently gain strong support, it would be expected that ballistic toolmark identification would also carry the same support. Therefore, as this current stage, while it has been shown the methods have potential for use in forensic casework the strength of evidence must be increased before they can be considered a truly validated method in forensic investigation.

Novel contributions

Novel contributions within the thesis firstly include a direct comparison of commercially available systems based on 2D imaging and matching, commercial systems based on areal measurement and correlation, and multi-purpose instruments capable of areal measurement of a wide range of surfaces.
Secondly, pre-processing techniques were put forward that have not been applied to the correlation of ballistic toolmark identification previously. The filtering techniques used on firing pin impressions are unique to this study, in which a Robust Gaussian filtration with cut off values of 75µm and 450µm were applied and have proved successful. The correlation of bullet toolmark impressions were completed using wavelet decomposition and allowed for the areal surface to be correlated, using both the D5 and D6 wavelet bands of a Discrete Wavelet Transform using a spline wavelet technique. This is to the author's knowledge the first instance in which correlation of bullet LEAs has been successfully completed using the areal dataset, rather than an averaged profile.
Finally, this is the first instance in which a system has been created which would result in a transparent and completely objective technique. While most current commercial systems rely on a 'match score', it means that the user is unable to ascertain how good a match has been found, as a match score tends to be an arbitrary value. However, in this case a percentage match is achieved, meaning the user gains a better understanding on the success of the correlation.

Further work

It is accepted that due to the manual selection of the region of interest for correlation, the techniques used in this study may still result in user bias being introduced into the results. As it has been shown in the results that areal correlation can result in a system with higher efficacy than current commercial systems, further work would include the automation of the ROI methods. To gain full automation, software would need to be created that is able to select the region of interest through examination of the class characteristics of the toolmarks. In the case of firing pin impressions, it is expected that this can be achieved in examination of the heights of each point to determine the beginning of the impression, and in bullet toolmarks it is expected that software will be able to differentiate the transitional area of the LEA and thus extract the region of interest from the surface. Once the region of interest has been extracted from the surface, automated pre-processing and correlation steps should be introduced. This would result in a system with high efficacy based upon objective methods that do not depend on the user or measurement variables such as lighting conditions and differences in measurement position.

As pre-processing methods were also determined based on Alias acquired measurements, optimising techniques for Alicona acquired datasets is also suggested to determine whether efficacy could be increased. While the size of individual characteristics of the toolmark will not change dependant on measurement method, and therefore wavelet bands and Gaussian cut-off values would remain the same, it is possible measurement artefacts within Alicona acquired datasets could be decreased.

Within the Odyssey collection there is a lack of information regarding manufacturing processes of the firearms and forensic firing techniques used for the firing of each test object. During forensic firing of the Odyssey collection, two methods were used, both water capture of the bullet and cotton wool capture. It has been generally accepted within the forensic ballistic identification field that variations in bullet capture may affect the imparted toolmark differently.

Cartridge manufacturers were also randomly applied to each firearm, which also affect the topography of imparted toolmarks. Therefore, it is likely that there may be variables introduced into the correlation methods that can be avoided. Therefore, it is advised that further work should be completed to corroborate the findings in this thesis, using evidence samples where such variables have been controlled to gain a better understanding of their effects on successful correlation.

To be able to maximise the support gained in the likelihood ratio, it is vital to create an automatic system based on the methods detailed. This would minimise any errors in the manual detection of the region of interest. It is also suggested that a set of posterior odds are created based around variables such are the differences in material composition, to ascertain the effect of variables on the likelihood ratio. This information can then be taken into account and may result in standard operating procedure that would warn against using a different manufacturer in test firing.

APPENDICES

Appendix 1: Publications

Work completed within this thesis was presented at various conferences, with the details as follows:

"The use of additive manufacturing for the presentation of ballistic toolmark evidence in court". Katie Foster, Paul Bills, Liam Blunt. Conference proceeding, 7th European Academy of Forensic Sciences conference, Prague 2015

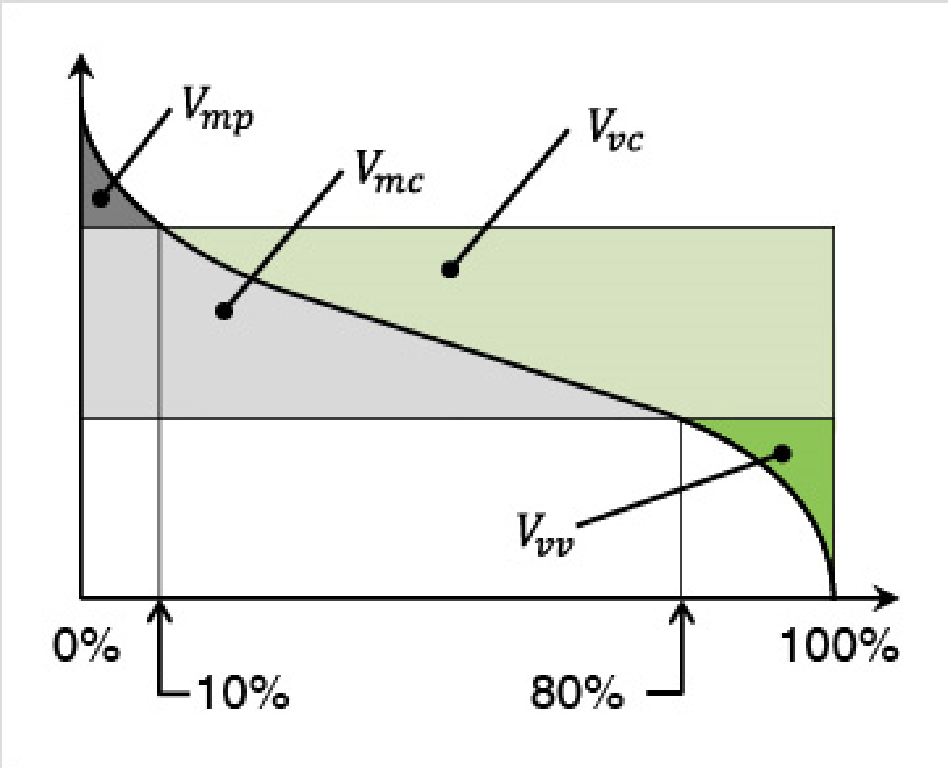
"Bridging the gap in ballistic toolmark identification" Katie Addinall (nee Foster), Wenhan Zheng, Paul Bills Liam Blunt. Conference proceeding at Met and Props 2017.

"Focus variation with integrated form removal for forensic comparison of bullet striations" Karl Walton, Katie Addinall (nee Foster), Wenhan Zeng, Liam Blunt, Poster presentation at the 21st triennial meeting of the IAFS International Association of Forensic Sciences 2017

There are also two journal papers currently in the process of being published, due to the commercial confidentiality of the experiments, it was necessary to complete patent applications before results were published.

Appendix 2: Definitions

Before defining volume parameters for a surface, material ratios must be defined to be able to separate peaks and valleys from the core surface. The standard used for this is the top 10% of the surface volume being defined as a peak, and the bottom 20% being defined as valleys, as detailed in the following graph (Appendix 2-1). The graph shows the range of volumes that are then used to define the volume parameters of the surface (Digital Surf, 2017).



Appendix 2-1: Definition of volume parameters

These ratios were used as is SURFSTAND software to calculate volume parameters for fired primer caps.

- Vmp: Peak material volume of the surface
- Vmc: Core material volume of the surface
- Vvc: Core void volume of the surface
- Vvv: Valley void volume of the surface

Appendix 3: Firing pin correlation

The following tables show the correlation results of each known match for each method tested in the thesis.

Appendix 3-1: Correlation score results of known matches within the Alias software

Test object	Known match B score	Known match C score
1A	34.76	802.76
2A	716.40	86.49
3A	6.28	66.74
4A	32.79	15.02
5A	99.25	99.99
6A	6.37	34.04
7A	40.07	2.96
8A	45.63	15.79
9A	16.99	15.67
10A	84.24	45.07
11A	60.69	47.78
12A	1.05	1793.03
13A	1.1	28.29
14A	51.63	891.08
15A	110.80	1691.33
16A	No result given	32.39
17A	1.55	47.28
18A	66.44	46.78
19A	169.02	181.63
20A	111.72	207.05
21A	1653.82	898.47
22A	1079.47	1147.74

Appendix 3-2: ACCF max results of firing pin correlation using Alias acquired datasets

Test object	Known match B score	Known match C score
1A	80.0374	80.5914
2A	76.4536	-17.6116
3A	95.5298	84.1454
4A	10.4123	70.7893
5A	22.4872	75.4182
6A	75.6504	35.3808
7A	56.5028	33.5901
8A	84.2626	76.5689
9A	91.6775	95.9432
10A	91.5566	93.1927
11A	91.6543	Damaged CC
12A	90.2047	86.5492
13A	81.0297	16.7658
14A	93.5607	78.4037
15A	90.7206	84.9622
16A	90.5817	70.521
17A	59.2021	54.023
18A	81.1848	46.8911
19A	67.517	69.2527
20A	80.0374	81.5914
21A	90.3661	83.1665
22A	89.9311	87.0085

Appendix 3-3: $ACCF_{max}$ results gained using Alicona acquired datasets.

Test object	Known match B score	Known match C score
1A	90.9141	64.7384
2A	76.5297	3.6096
3A	56.1192	92.3977
4A	0.70921	42.6632
5A	69.0974	57.5362
6A	1.1984	-22.3437
7A	74.1512	58.338
8A	78.2321	58.1768
9A	88.9252	54.6478
10A	91.9749	No match
11A	21.9788	No match
12A	87.3343	70.0864
13A	91.2244	44.9739
14A	89.5541	42.1357
15A	98.9151	87.203
16A	93.688	81.2328
17A	79.832	1.6638
18A	34.7844	No match
19A	68.4934	90.5045
20A	56.2992	93.5604
21A	87.8355	87.4336
22A	87.2965	76.6419

Appendix 4: Bullet correlation

The following tables show the correlation results gained using Alias and Alicona acquired datasets

Alias dataset correlations

Appendix 4-1: Correlation results using Alias acquired datasets and the D5 wavelet band

Test object	Known match B score	Known match C score
1A	97.3843	96.708
2A	98.3843	No match
3A	No match	No match
4A	82.2166	78.3249
5A	No match	No match
6A	69.3862	82.0352
7A	89.0228	76.3522
8A	81.4197	57.6963
9A	73.4887	No match
10A	99.2979	98.3948
11A	No match	No match
12A	67.5115	88.6563
13A	77.386	82.9294
14A	No match	68.0916
15A	74.0408	76.2943
16A	19.1647	94.4828
17A	68.2893	71.4562
18A	60.2936	100
19A	53.7236	81.1204
20A	64.3462	95.9262
21A	95.9097	93.8334
22A	92.042	97.0634
23A	28.2023	95.825

Appendix 4-2: Results of bullet correlation using the D6 wavelet band

Test object	Known match B score	Known match C score
1A	85.8946	66.2211
2A	84.8115	62.2644
3A	62.7466	54.0282
4A	96.1137	88.9108
5A	65.4914	87.62
6A	85.219	78.5244
7A	60.5487	57.2063
8A	54.5564	54.3794
9A	No match	58.5142
10A	71.6944	66.3144
11A	68.1352	58.31
12A	63.0324	43.3281
13A	No match	18.2398
14A	83.3865	60.6261
15A	No match	58.9031
16A	74.1488	90.2003
17A	No match	64.5088
18A	No match	48.8599
19A	No match	82.5621
20A	64.691	39.7261
21A	58.282	48.0568
22A	60.2491	No match
23A	No match	90.1335

Appendix 4-3: Table of correlation results using the combined D5 and D6 wavelet bands

Test object	Known match B score	Known match C score
1A	71.6184	68.1782
2A	91.9202	79.523
3A	56.1592	44.6348
4A	90.2284	79.0612
5A	No match	No match
6A	74.3906	77.8614
7A	59.3472	52.1388
8A	No match	59.4852
9A	51.1551	22.5299
10A	72.1433	63.8603
11A	81.5379	45.5725
12A	70.5608	61.4725
13A	42.6741	No match
14A	57.3633	81.3823
15A	80.6895	70.9216
16A	67.4385	78.0854
17A	85.4589	68.4175
18A	No match	68.1103
19A	73.5423	79.7938
20A	80.9706	78.9938
21A	83.4732	62.7044
22A	54.9273	No match
23A	50.5932	87.4916

Appendix 4-4: Results of correlation using Alicona acquired datasets using the D6 wavelet band.

Test object	Known match B score	Known match C score
1A	No match	76.7427
2A	82.47	73.8372
3A	69.0707	68.1093
4A	52.285	No match
5A	75.1336	No match
6A	64.8691	83.4518
7A	76.4717	51.5269
8A	77.3048	No match
9A	59.4366	42.6249
10A	No match	No match
11A	51.8932	No match
12A	58.9696	48.78
13A	76.2686	No match
14A	61.0041	39.0165
15A	No match	86.5922
16A	No match	60.5883
17A	55.4361	No match
18A	71.4632	55.1466
19A	No match	85.7354
20A	55.6909	47.4688
21A	81.5515	69.2889
22A	65.7415	68.7454
23A	No match	72.5369

Appendix 4-5: Results of correlation using Alicona acquired datasets in the D5 wavelet band

Test object	Known match B score	Known match B D_s score	Known match C score	Known match C D_s score
1A	61.0807	2126.9312	58.985	246.7635
2A	50.1091	231.3795	56.2171	554.9986
3A	83.3866	184.4565	51.9236	3562.0913
4A	58.6798	7503.1132	67.1874	7756.7334
5A	81.4191	195.5441	84.2417	3265.2198
6A	67.6073	230.267	61.1775	503.7006
7A	50.8971	104.7427	64.0359	486.1586
8A	49.8355	1233.6558	-2.8375	4037.3362
9A	50.4814	348.9704	57.4512	134.9191
10A	55.4665	596.4338	46.5444	100.9061
11A	49.5524	149.8677	7.8524	292.2345
12A	39.3734	141.0475	44.6169	80.7805
13A	83.7503	302.9034	60.937	2409.8868
14A	33.6801	2180.2988	39.0165	87.9632
15A	63.2483	3987.6269	64.4041	31825.636
16A	61.2692	114.3852	62.7036	212.2202
17A	49.8819	75.3636	84.8989	2723.52
18A	39.7186	89.2288	31.3856	696.7024
19A	2.6547	103.5956	41.2962	155.732
20A	57.6449	895.7922	62.9191	979.0721
21A	52.0152	111.3491	44.4024	1205.834
22A	74.6201	144.7367	64.4509	262.8049
23A	62.808	3507.8053	72.0846	378.0025

Appendix 5: Pre-processing tests

Appendix 5-1: Example firing pin correlation results using various pre-processing methods

Pre-processing method	KM1 $ACCF_{max}$	KM1 D_s	KM2 $ACCF_{max}$	KM2 D_s	NM1 $ACCF_{max}$	NM1 D_s	NM2 $ACCF_{max}$	NM2 D_s
Least squares levelling Robust Gaussian 0.025-2.5mm	97.9606	20.3063	98.1733	10.8634	89.9069	56.9115	88.9259	70.9324
Robust polynomial second order levelling Robust Gaussian 0.025-2.5mm	98.2371	16.1235	97.9999	11.2908	89.6542	54.8441	89.6959	65.731
No levelling Robust Gaussian 0.025-2.5mm	14.2633	210.4705	0.88321	128.8412	20.6569	325.6929	2.8626	125.6963
Robust polynomial third order levelling Robust Gaussian 0.025-2.5mm	98.18	16.1499	97.8638	11.4337	91.3252	48.5009	89.1264	65.3658
Robust polynomial fourth order levelling Robust Gaussian 0.025-2.5mm	97.6317	28.0946	96.7809	19.0009	89.827	69.747	88.8136	83.9665
Robust polynomial second order levelling Robust Gaussian 0.025-0.8mm	62.7712	226.8862	66.1323	83.0405	36.023	7285.843	59.1378	1895.377
Least squares levelling Robust Gaussian 0.050- 0.45mm	84.7691	0.9631	86.3345	3.5293	67.8369	3.3518	57.9431	94.1872
Least squares levelling Robust Gaussian 0.075-0.45mm	86.1369	0.865	89.9887	4.8069	71.448	2.9189	60.9687	58.5846
Least squares levelling Robust Gaussian 0.1-0.45mm	89.506	0.64664	89.1201	4.3139	73.4994	2.5907	59.0439	117.3679
Least squares levelling	93.6893	0.39149	82.1227	5.2908	79.1284	2.2778	29.0047	174.7453

Robust Gaussian 0.150-0.45mm								
Least squares levelling Robust Spline 0.025-0.45mm	47.3978	69.6721	22.4642	117.734	17.6971	212.1845	42.7295	388.0956
Least squares levelling Robust Spline 0.05-0.45mm	58.6144	43.8956	37.6406	90.1572	24.0234	162.5061	17.2467	1113251.4
Least squares levelling Robust Spline 0.075-0.45mm	68.7853	30.5509	23.6566	119.2214	37.7225	157.9934	34.9331	489.3097
Least squares levelling Robust Spline 0.1-0.45mm	69.4637	26.0122	29.2473	119.2462	34.4337	165.6698	25.9987	682.3949
Least squares levelling Robust Spline 0.15-0.45mm	45.6788	163.4937	29.3781	97.5405	24.8902	568.3205	53.1785	456.6519

Appendix 5-2: Correlation results using unfiltered bullet surfaces

Striations	KM1 $ACCF_{max}$	KM1 D_s	KM2 $ACCF_{max}$	KM2 D_s	NM1 $ACCF_{max}$	NM1 D_s	NM2 $ACCF_{max}$	NM2 D_s
All	83.6791	3.6128	70.605	6.0771	71.8574	10.6716	51.3457	15.0144
S1	99.1743	0.0974	93.9141	0.7171	98.945	0.53583	90.9291	2.1616
S2	91.9744	1.195	87.4328	1.6019	92.3329	6.2399	59.1121	6.4985
S3	78.6217	3.3284	59.4189	7.7351	79.5251	21.7542	7.1173	21.3656
S4	82.2132	10.8513	63.8694	17.4195	81.2349	33.5317	35.9989	43.5779
S5	83.7856	7.0492	52.727	16.7472	57.6894	18.4082	19.2081	30.8351
S6	67.5993	2.6122	67.7818	3.7705	76.4986	3.0336	98.5455	0.88029

Appendix 5-3: Correlation results in bullets using least squares levelling and Robust Gaussian filtering 0.025-2.5mm

Striations	KM1 $ACCF_{max}$	KM1 D_s	KM2 $ACCF_{max}$	KM2 D_s	NM1 $ACCF_{max}$	NM1 D_s	NM2 $ACCF_{max}$	NM2 D_s
All	84.6923	30.7757	70.7652	55.88	79.9931	45.8752	54.0347	97.275
S1	99.3641	1.2753	93.8861	11.6133	98.8139	2.27784	93.15	12.1141
S2	75.3669	41.6285	86.5708	23.2958	54.9903	75.4261	67.5512	53.9597
S3	93.4518	13.3349	92.3906	17.5681	44.5872	79.4534	92.5374	25.0323
S4	99.0751	9.3556	79.0615	44.7241	54.5428	159.1731	45.0771	81.0267
S5	94.6766	11.1774	95.8907	8.9159	60.4827	77.728	57.9491	550.0477
S6	38.4465	107.1928	86.1667	36.483	54.5214	92.0916	40.6183	89.2026

Appendix 5-4: Correlation results in bullets using least squares levelling and spline filtering 0.025-2.5mm

Striations	KM1 $ACCF_{max}$	KM1 D_s	KM2 $ACCF_{max}$	KM2 D_s	NM1 $ACCF_{max}$	NM1 D_s	NM2 $ACCF_{max}$	NM2 D_s
All	79.9189	41.4927	63.0219	75.8801	75.6389	60.8195	44.1032	162.5959
S1	98.9931	2.9582	91.7091	17.2743	98.6152	3.2154	88.0022	42.779
S2	94.0287	11.6123	96.8673	6.5556	98.5491	5.8491	68.282	84.5136
S3	98.9724	4.281	89.7427	30.3644	61.3578	73.3326	88.7214	39.1596
S4	88.249	28.723	89.8922	32.9574	71.9943	158.0278	92.5426	83.8342
S5	96.8085	6.4612	98.308	9.7079	91.6067	35.3105	91.6143	89.226
S6	97.1243	6.065	94.5896	16.4609	90.3907	25.2724	81.4568	60.4251

Appendix 5-5: $ACCF_{max}$ results of bullet correlation in wavelet decomposition between a test object and KM1

$ACCF_{max}$ results	Test object									
KM1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	30.3761	24.9712	20.2419	16.7556	17.6779	12.7343	10.009	8.5555	8.8215	4.2757
D2	32.7885	46.8883	29.8965	20.2594	22.6222	17.5189	14.5451	11.8149	12.8508	6.5075
D3	22.2643	20.7463	39.2396	24.2915	15.9902	2.022	20.119	15.8011	17.4983	9.2558
D4	18.4915	21.0107	46.3374	69.5198	36.7391	20.9968	20.8584	21.7245	19.3772	10.8726
D5	10.3082	21.6183	21.6183	38.4738	82.8849	30.5239	25.2911	20.7648	30.8159	14.2046
D6	13.9514	16.6488	16.6488	10.8315	32.0165	89.9134	29.6734	20.4356	26.2842	17.8094
D7	11.0968	13.6894	17.8583	26.2773	26.5527	24.5371	93.9671	33.6722	29.6519	14.6453
D8	11.2534	16.6413	19.6829	28.2314	32.4119	30.9934	13.7485	35.5631	2.2182	7.4704
D9	8.6454	12.3861	15.8487	22.7836	27.2447	22.6103	33.1371	0.42259	96.5384	9.9047
D10	3.9521	5.6842	7.8019	12.2999	13.1153	13.3709	17.3568	-4.4958	10.3087	97.0876

Appendix 5-6: D_s results of bullet correlation in wavelet decomposition between a test object and KM1

D_s score	Test object									
KM1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	167.8378	110.7379	102.7047	100.0067	98.554	98.1991	98.7877	99.4353	99.2635	105.1616
D2	134.3686	90.7209	98.7542	100.4518	97.1794	96.6508	97.4231	98.7685	98.3324	107.3886
D3	121.1571	177.528	99.4144	104.2762	110.2301	102.7013	2308.9738	2153.3745	2691.0555	113.9674
D4	371.9003	268.8456	123.0008	52.6107	101.4347	95.1835	94.9184	95.9521	96.8288	130.0237
D5	108.359	102.6213	102.6213	122.1044	30.8505	747.4481	797.3361	778.3336	897.202	140.7486
D6	97.9564	96.9668	96.9668	107.1568	382.6244	30.6781	175.8095	198.409	211.6715	355.3098
D7	3116.6683	2082.3544	1032.6212	539.9042	464.8375	126.1976	12.3793	118.5034	114.9419	412.1307
D8	3001.6587	2067.8856	1032.9163	542.1634	453.3581	119.537	146.0767	112.8311	157.8224	421.0018
D9	4583.4468	3127.5587	1554.0876	828.0856	691.7782	161.3716	137.3229	211.2514	6.8151	605.3936
D10	747.689	517.6871	303.3657	197.159	180.9334	106.5358	722.6791	768.1014	821.7134	6.5907

Appendix 5-7: $ACCF_{max}$ results of bullet correlation in wavelet decomposition between a test object and NM1

$ACCF_{max}$ results	Test object									
NM1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	21.1428	21.6209	7.7277	13.96	11.4441	11.6414	10.154	9.9408	8.3843	4.856
D2	20.9503	31.9855	12.6807	21.1173	14.5567	16.1558	13.6315	14.0822	12.3415	7.2881
D3	13.4124	13.1133	21.9054	22.8481	16.5646	4.299	18.226	19.0771	16.717	9.6536
D4	21.3238	36.2826	15.9717	23.5919	37.6837	17.2616	21.7705	23.3594	17.6918	12.6855
D5	18.0929	26.5992	20.7941	28.3134	40.8868	28.3768	24.6925	28.4187	26.6209	15.0646
D6	13.0175	20.1549	26.0135	14.6573	34.3709	54.7383	36.9926	27.9824	25.1558	18.0269
D7	10.5191	16.9905	15.2148	18.5879	26.1076	30.5798	83.7977	12.1702	30.0889	15.8096
D8	10.7986	15.7869	19.5161	21.3859	26.3143	25.4379	18.099	93.2181	1.0616	10.6018
D9	8.7041	13.7612	14.8497	19.0203	21.5722	22.4504	24.654	-1.2518	97.4082	6.3124
D10	4.0242	5.8249	7.5546	8.2039	9.0217	11.5457	6.6949	2.7366	9.3562	90.349

Appendix 5-8: D_s results of bullet correlation in wavelet decomposition between a test object and NM1

D_s results	Test object									
NM1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	129.9029	116.5288	115.9326	104.9549	100.3549	98.4815	98.8168	99.0512	99.3662	104.7444
D2	157.4897	117.2577	121.7898	104.2564	100.3108	97.1161	97.8258	98.0177	98.4867	106.7209
D3	243.6279	199.4263	190.005	244.8619	102.881	2535.2551	96.1186	96.3653	97.2356	112.3333
D4	386.8405	250.0601	198.465	148.4896	89.1125	97.1438	94.6514	94.9369	97.4257	123.9676
D5	102.7374	97.9979	120.8478	115.4421	90.7161	819.284	844.332	796.3499	91.4382	263.6102
D6	98.3509	95.4294	92.6497	100.2461	260.247	123.0086	166.0126	190.7769	117.0233	118.1139
D7	3030.0872	2147.7782	1113.3798	773.3813	312.7393	113.2754	29.0513	149.8342	114.4119	382.2076
D8	3024.2564	2161.1634	1101.7461	773.7403	320.0071	125.0853	136.6296	14.1613	159.4521	394.492
D9	4549.4502	3253.0556	98.6226	1143.6419	471.9997	154.4244	150.1322	207.2802	5.1677	562.002
D10	709.9475	542.3021	133.1931	139.4949	154.7031	107.0021	110.5262	112.691	106.0182	18.8938

Appendix 6: Calibration certificates

Certificate of calibration

Certificate number: CAL.17.8970 **Date:** 30th November 2017

Model type: Alicona IFM G4 Vg **Serial number:** 017104701811

Optical Axis:

	X	Result	Y	Result
Pre Optical Axis	N	-0.12°	Y	0.05°
Post Optical Axis	Y	-0.01°	Y	0.04°

Sensor Rotation:

Pre Sensor Rotation	Y	0.01°
Post Sensor Rotation	Y	0.01°

Lateral Accuracy:

Lens	Adjusted	Calibrated	Nominal size	Deviation	Uncertainty
x2.5	N/A	N/A	120.00µm	N/A	480nm
x5	Y	Y	49.982µm	0nm	200nm
x10	Y	Y	24.004µm	0nm	200nm
x20	Y	Y	24.004µm	1nm	200nm
x50	Y	Y	24.004µm	4nm	200nm
x100	Y	Y	24.004µm	5nm	200nm

Curvature of Field:

Lens	Adjusted	Calibrated	Max Required Sz	Measured Sz	Uncertainty
x2.5	N/A	N/A	4.600µm	N/A	480nm
x5	Y	Y	820nm	99.9317nm	200nm
x10	Y	Y	200nm	34.8123nm	200nm
x20	Y	Y	100nm	27.6787nm	200nm
x50	Y	Y	40nm	12.7736nm	200nm
x100	Y	Y	20nm	9.3024nm	200nm

Curvature of Field Polarization:

Lens	Adjusted	Calibrated	Max Required Sz	Measured Sz	Uncertainty
x2.5	N/A	N/A	4.600µm	N/A	480nm
x5	Y	Y	820nm	471.5663nm	200nm
x10	Y	Y	200nm	41.1308nm	200nm
x20	Y	Y	100nm	27.2705nm	200nm
x50	Y	Y	40nm	13.7562nm	200nm
x100	Y	Y	20nm	11.8258nm	200nm

Optimax Imaging Inspection & Measurement Ltd
Michael Francis House • 3 Trimbush Way
Market Harborough • Leicestershire • LE16 7XY
T +44 [0]1858 436940
F +44 [0]1858 436941
E info@optimaxonline.com
Registered in England
No. 5562754
VAT No. 780 4974 94

Optimax
Imaging Inspection Measurement

Certificate of calibration

Certificate number: CAL.17.8970 **Date:** 30th November 2017
Model type: Alicona IFM G4 Vg **Serial number:** 017104701811

Vertical Calibration:

Lens	Adjusted	Calibrated	Nominal size	Deviation	Uncertainty
x2.5	N/A	N/A	1000.01µm	N/A	480nm
x5	N	Y	1000.01µm	784nm	200nm
x10*	N	Y	1000.01µm	440nm	200nm
x20	N	Y	1000.01µm	523nm	200nm
x50	N	Y	1000.01µm	610nm	200nm
x100	N	Y	1000.01µm	714nm	200nm

The above results are within specified and acceptable limits (Yes or No): YES

Ambient temperature at time of calibration: 20.4°C

The above errors are the maximum recorded deviation from the nominal truth and are corrected to 20 degrees centigrade. This calibration certificate is void if any adjustments are made to software error correction data or mechanical alignments.

Equipment used:

This certificate is to certify that the above instrument has been calibrated using a grade AA granite square and Mitutoyo digital indicator for parallelism, straightness and squareness. Glass scale grid and ruler for X, Y linear accuracy and grade 1 tungsten carbide gauge blocks for Z axis. All reference standards used are UKAS certified and are traceable to national standards. Maximum errors are documented in the certificate of calibration with a full breakdown of information contained within the calibration and service schedule.

Reference standards used:

AA Granite square – Serial number 2751, UKAS cert number N3030568D
Mitutoyo digital indicator – Serial number 010505, UKAS cert number N3030569D
Two dimensional glass grid – Serial number PGR/200/115, UKAS cert number N3030570D
Alicona Calibration tool
comprising carbide step blocks
and glass grids and dots: Serial number 002217001712, cert mark 06ALICONA12

Uncertainty Statement

The reported expanded uncertainty of 0.005mm is based on a standard uncertainty multiplied by a coverage factor k = 2, providing a coverage probability of approximately 95 %.

*x10 objective lens is used as the reference lens for the service software correction

Uncertainties quoted against objectives refer to traceable artefact

Approved checking signatories: Ms A Aldhous

Name engineer: P Reed **Counter Signature engineer:** P.P. A Aldhous

Bibliography

Abdul-Rahman, H., Xiang, J., & Scott, P. (2013). Freeform Surface Filtering using the Lifting Wavelet Transform. *Precision Engineering*, 187-202.

AFTE Criteria for Identification Commitee. (1992). Theory of Identification, Range of Striae Comparison Reports and Modified Glossary Definitions- an AFTE Criteria for Identification Comitee Report. *AFTE Journal*, 336-340.

Alicona. (2017, 5 3). *Fastest optical form- and roughness*. Retrieved from InfiniteFocus: http://www.alicona.com/products/infinitefocus/

Allen, G., & Dempsey, N. (2016). *Firearm Crime Statistics: England and Wales.* London: Home Office.

ATF. (2016, 9 1). *Fact Sheet- Firearms Trafficking & Interdiction Branch*. Retrieved from ATF: https://www.atf.gov/resource-center/fact-sheet/fact-sheet-firearms-trafficking-interdiction-branch

Bachrach, B. (2002). Development of a 3D-based Automated Firearms Evidence Comparison System. *Journal of Forensic Sciences*, 1253-1264.

Bachrach, B., Jain, A., Jung, S., & Koons, R. (2010). A Statistical Validation of the Individuality and Repeatability of Striated Tool Marks: Screwdrivers and Tongue and Groove Pliers. *Journal of Forensic Sciences*, 348-357.

Baiker, M., Keereweer, I., Pieterman, R., Vermejj, E., van der Weerd, J., & Zoon, P. (2014). Quantitative Comparison of Striated Toolmarks. *Forensic Science International*, 186-199.

Baiker, M., Pieterman, R., & Zoon, P. (2015). Toolmark Variability and Quality Depending on the Fundamental Parameters: Angle of Attack, Toolmark Depth and Substrate Material. *Forensic Science International*, 40-49.

Balscan. (2017, 02 13). *Balscan Products*. Retrieved from www.forensic.cz/en/products/balscan

Barrett, M., Tajbakhsh, A., & Warren, G. (2011). Portable Forensic Ballistics Examination Instrument: Advanced Ballistics Analysis System (ALIAS). *AFTE*, 74-79.

BBC. (2010, 10 1). Pair Jailed for Real IRA Gun Plot. *BBC News*.

Beesly, L. (1961). Formation of Rifling in Small Arms Barrels by a Precision Forging Process. *Production Engineer*, 382-388.

Bell, S. (2012). *A Dictionary of Forensic Science.* Oxford: Oxford Univeristy Press.

Bello, S., & Soligo, C. (2008). A New Method for the Quantitative Analysis of Cutmark Micromorphology. *Journal of Archaeological Science*, 1542-1552.

Benecke, M. (2001). A Brief History of Forensic Entomology. *Forensic Science International*, 2-14.

Bennett, T., & Holloway, K. (2004). Possession and Use of Illegal Guns Among Offenders in England and Wales. *The Howard Journal of Criminal Justice*, 237-252.

Bertin, J., & Henifin, M. (1994). Science, Law, and the Search for Truth in the Courtroom: Lessons from Daubert v. Merrell Dow. *The Journal of Law, Medicine & Ethics*, 6-20.

Biasotti, A. (1959). A Statistical Study of the Individual Characteristics of Fired Bullets. *Journal of Forensic Sciences*, 34-50.

Blunt, L. (2008). *EU Patent No. 0811985.*

Blunt, L., & Xiang, J. (2003). *Advanced Techniques for Assessment Surface Topography: Development of a basis for 3D Surface Texture Standards "surfstand".* London: Kogan Page Science.

Bolton-King, R. (2012). Classisication of Barrel Rilfling Transitions for the Forensic Idenfication of Firearms. Nottingham Trent University.

Bolton-King, R. (2016). Preventing Miscarriages of Justice: A Review of Forensic Firearm Identification. *Science & Justice*, 129-142.

Bolton-King, R., Bencsik, M., Evans, J., Smith, P., Painter, D., Allsop, C. L., Painter, J.D., Caranton, W.M., Wayne, M. (2012). Numerical Classification of Curvilinear Structures for the Identification of Pistol Barrels. *Forensic Science International*, 197-209.

Bolton-King, R., Evans, J., Smith, C., J.D, P., Allsop, D., & Cranton, W. (2010). What are the Prospects of 3D Profiling Systems Applied to Firearms and Toolmark Identification. *AFTE Journal*, 23-33.

Bonfanti, M., & De Kinder, J. (1999). The Influence of Manufacturing Processes on the Identification of Bullets and Cartridge Cases- A Review of the Literature. *Science & Justice*, 3-10.

Brinck, T. (2008). Comparing the Performance of IBIS and BulletTRAX-3D Technology Using Bullets Fired Through 10 Consecutively Rifled Barrels. *Journal of Forensic Sciences*, 677-682.

Bunch, S. (2000). Consecutive Matching Striation Criteria: A General Critique. *Journal of Forensic Sciences*, 955-962.

Bunch, S., & Wevers, G. (2013). Application of Likelihood Ratios for Firearm and Toolmark Analysis. *Science & Justice*, 223-229.

Bunch, S., & Wevers, G. (2013). Application of Likelihood Ratios for Firearm and Toolmark Analysis. *Science & Justice*, 223-229.

Caddick, A., & Porter, L. (2012). Exploring a Model of Professionalism in Multiple Perpetrator Violent Gun Crime in the UK. *Criminology & Criminal Justice*, 61-82.

Cadevall, C., & Schwarz, N. (2013, 10 22). Forensic Inspection for Firearms: Comparison Techniques in Optical Profilometry. *Forensic Magazine*.

Cadre Forensics. (2017, 7 27). *TopMatch-GS 3D*. Retrieved from http://www.cadreforensics.com/index.html

Capri, A. (2011). *Progress in Molecular and Environmental Bioengineering - From Analysis and Modeling to Technology Applications.* Intech.

Chu, W., Song, J., Vorburger, T., & Ballou, S. (2010). Striation Density for Predicting the Identifiability of Fired Bullets with Automated Inspection Systems. *Journal of Forensic Sciences*, 1222-1226.

Chu, W., Song, J., Vorburger, T., Yen, J., Ballou, S., & Bachrach, B. (2010). Pilot Study of Automated Bullet Signature Identification Based on Topography Measurements and Correlations. *Journal of Forensic Sciences*, 341-347.

Chu, W., Thompson, R., Song, J., & Vorburger, T. (2013). Automatic Identification of Bullet Signatures Based on Consecutive Matching Striae (CMS) Criteria. *Forensic Science International*, 137-141.

Committee on Identifying the Needs of the Forensic Sciences Community . (2009). *Strengthening Forensic Science in the United States : A Path Forward.* Washington: National Academies Press.

Cork, D., Nair, V., & Rolph, J. (2010). Some Forensic Aspects of Ballistic Imaging. *Fordham Urban Law Journal*, 473-502.

Cork, D., Rolph, J., & Meieran, E. (2008). *Ballistic Imaging.* Washington: National Academies Press.

Crowson, I. (2017, 12 21). Rogue Firearms Dealer who Supplied Bullets Found in Derbyshire Jailed for 30 Years. *Derby Telegrah*.

Danzl, R., Helmli, F., & Scherer, S. (2011). Focus Variation--a Robust Technology for High Resolution Optical 3D Surface Metrology. *Strojniski Vestnik/Journal of Mechanical Engineering*, 245-256.

Danzl, R., Hemli, F., Rubert, P., & Prantl, M. (2008). Optical Roughness Measurements on Specially Designed Roughness Standards. *Proceedings SPIE 7102, Optical Fabrication, Testing, and Metrology III* (p. 1). Glasgow: SPIE.

Davies, M., Wells, C., Squires, P., T.J, H., & Lecky, F. (2012). Civilian Firearm Injury and Death in England and Wales. *Emergency Medicine Journal*, 10-14.

De Ceuster, J., & Dujardin, S. (2015). The Reference Ballistic Imaging Database Revisited. *Forensic Science International*, 82-87.

De Ceuster, J., Hermsen, R., Mastaglio, M., & Nennstiel, R. (2012). A Discussion on the Usefulness of a Shared European Ballistic Image Database. *Science & Justice*, 237-242.

de Groot, P. (2011). *Coherence Scanning Interferometry.* Berlin: Springer.

De Kinder, J., Tulleners, F., & Thiebaut, H. (2004). Reference Ballistic Imaging Database Performance. *Forensic Science International*, 207-215.

Digital Surf. (2017, 1 22). *Filtration Techniques for Surface Texture*. Retrieved from Digital Surf: http://www.digitalsurf.com/en/guide-filtration-techniques.html

Edmond, G. (2015). Forensic Science Evidence and the Conditions for Rational (jury) Evaluation. *Melbourne University Law Review*, 77-127.

Ender, A., & Mehl, A. (2013). Accuracy of Complete-arch Dental Impressions: A New Method of Measuring Trueness and Precision. *The Journal of Prosthetic Dentistry*, 121-128.

European Network of Forensic Science Institutes. (2010). *ENFSI Guideline for Evaluative Reporting in Forensic Science.* Budapest: ENFSI .

Evett, I. (1998). Towards a Uniform Framework for Reporting Opinions in Forensic Science Casework. *Science & Justice*, 198-202.

Evofinder. (2017, 02 13). *Company History*. Retrieved from Evofinder: www.evofinder.com/company/history

Faden, D., Kidd, J., L.S, C., Morris, M., Genalo, L., & Kreiser, J. (2001). Statistical Confirmation of Empirical Observations Concerning Tool Mark Striae. *AFTE Journal*, 205-214.

Fahy, F. (2001). Appendix 4 - Coherence and Cross-Correlation A2. In *Foundations of Engineering Acoustics* (pp. 397-400). London: Academic Press.

FTI. (2017, 02 13). *IBIS*. Retrieved from Ultra Forensic Technology: www.ultra-forensictechnology.com/ibis

Gambino, C., McLaughlin, P., Kuo, L., Kammerman, F., Shenkin, P., Diaczuk, P., Petraco, N, Hamby, J., Petraco, N.D.K. (2011). Forensic Surface Metrology: Tool Mark Evidence. *Scanning*, 272-278.

George, W. (2004). A Validation of the Brasscatcher Portion of the NIBIM/IBIS System. *AFTE Journal*, 286-288.

Hallswoth, S., & Silverstone, D. (2009). 'That's life innit': A British Perspective on Guns, Crime and Social Order. *Criminology & Criminal Justice*, 359-377.

Hamilton, D., & Wilson, T. (1982). Three-Dimensional Surface Measurement Using the Confocal Scanning Microscope. *Applied Physics B*, 211-213.

Hamzah, N. (2016). Development of an Objective Method for the Comparison of Fired Projectiles using an Air Pistol as a template. *Forensic Science International*, 106-112.

Hannam, A. (2010). Trends in Converted Firearms in England & Wales as Identified by the National Firearms Forensic Intelligence Database (NFFID) between September 2003 and September 2008. *Journal of Forensic Sciences*, 757-766.

Heard, B. (2013). *Forensic Ballistics in Court: Interpretation and Presentation of Firearms Evidence.* Chichester: Wiley-Blackwell.

Helmli, F. (2011). *Focus Variation Instruments.* Berlin: Springer.

Hiersemenzel, F., Petzing, J., Leach, R., Helmli, F., & Singh, J. (2012). Areal Texture and Angle Measurements of Tilited Surfaces Using Focus Variation Methods. *3rd International Conference on Surface Metrology* (p. 1). Annecy: Universite de Savoie.

Home Office. (1968). *Firearms Act .* London: Home Office.

Huang, Z., & Leng, J. (2010). A Novel Binarization Algorithm for Ballistics Imaging Systems. *2010 3rd International Congress on Image and Signal Processing* (pp. 1287-1291). Yantai: IEEE.

INTERPOL. (2017, 1). *INTERPOL Ballistic Information Network.* Retrieved from https://www.interpol.int/Crime-areas/Firearms-trafficking/INTERPOL-Ballistic-Information-Network-IBIN

Johnson, M., & Adelson, E. (2009). Retrographic Sensing for the Measurement of Surface Texture and Shape. *Conference on Computer Vision and Pattern Recognition* (pp. 1070-1077). Miami: IEEE.

Jones, B., & Guerci, J. (1997). Intelligent Image Capture of Cartridge Cases for Firearms Examiners. *Proceedings Volume 2942, Investigative Image Processing* (pp. 94-104). Boston: SPIE.

Kassin, S., Dror, I., & Kukucka, J. (2013). The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions. *Journal of Applied Research in Memory and Cognition*, 42-52.

Kaufman, H. (2001). The Expert Witness. Neither Frye nor Daubert Solved the Problem: what can be done? *Science & Justice*, 7-20.

Lambelet, P. (2011). Parallel Optical Coherence Tomography (pOCT) for Industrial 3D Inspection. *Optical Measurement Systems for Industrial Inspection VII* (pp. 1-12). Munich: SPIE.

Leach, R. (2010). *Fundamental Principles of Engineering Nanometrology.* Amsterdam: William Andrew.

Leach, R. (2011). *Optical Measurement of Surface Topography.* Teddington: Springer.

Leach, R., Brown, L., Jiang, X., & Blunt, R. (2008). *Measurement Good Practice Guide no. 108. Guide to the Measurement of Smooth Surface Topography using Coherence Scanning Interferometry.* Teddington: National Physical Laboratory.

Leeds Micro. (2013, 2 04). *Introducing the Discovery!* Retrieved from Leeds Micro: https://www.leedsmicro.com/blog/2013/02/04/introducing-leeds-discovery/

Li, D. (2003). Image Processing for the Positive Identification of Forensic Ballistics Specimens. *Proceedings of the Sixth International Conference of Information Fusion, 2003.* (pp. 1494-1498). Cairns: IEEE.

Lucy, D. (2005). *Introductory Statistics for Forensic Scientists.* John Wiley & Sons Ltd.

Lyon, D. (2010). The Discrete Fourier Transform, part 6: Cross Correlation. *Journal of Object Technology*, 1.

Ma, L., Song, J., Whitenton, E., Zheng, A., Vorburger, T., & Zhou, J. (2004). NIST Bullet Signature Measurement System for RM (reference material) 8240 Standard Bullets. *Journal of Forensic Sciences*, 1-11.

Mackowski, C., & White, K. (2013). *Last Days of Stonewall Jackson : The Mortal Wounding of the Confederacy's Greatest Icon.* El Dorado Hills, UNITED STATES: Savas Beatie.

Mahat, M., Aris, A., Jais, U., Yahya, M., & Ramli, R. (2011). Infinite Focus Microscope (IFM): Microbiologically influenced corrosion (MIC) behavior on mild steel by Pseudomonas aeruginosa. *2011 International Symposium on Humanities, Science & Engineering Research (SHUSER)* (pp. 106-110). Kuala Lumpur: IEEE.

Martire, K., Kemp, R. I., Sayle, M., & Newell, B. (2014). On the Interpretation of Likelihood Ratios in Forensic Science Evidence: Presentation Formats and the Weak Evidence Effect. *Forensic Science International*, 61-68.

McGuire, S. (1996). The Dunblane Effect. *Newsweek*, 46.

NABIS. (2017, 4 22). *Welcome to the National Ballistics Intelligence Service*. Retrieved from NABIS: http://www.nabis.police.uk/

National Crime Agency. (2017, 3 17). *Illegal Firearms*. Retrieved from National Crime Agency: http://www.nationalcrimeagency.gov.uk/crime-threats/firearms

Nayar, S. (1989). Surface Reflection: Physical and Geometrical Perspectives. *Carnegie Mellon Univeristy Research Showcase*, 611-634.

Nennstiel, R., & Rahm, J. (2006). A Parameter Study Regarding the IBIS™ Correlator. *Journal of Forensic Sciences*, 18-23.

Nichols, R. (2007). Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges. *Journal of Forensic Sciences*, 586-594.

NIST. (2014, 06 27). *Firearms and Toolmarks Subcommittee*. Retrieved from NIST: https://www.nist.gov/topics/forensic-science/firearms-and-toolmarks-subcommittee

Nordgaard, A., & Rasmusson, B. (2012). The Likelihood Ratio as Value of Evidence-More than a Question of Numbers. *Law, Probability and Risk*, 303-315.

O'Neill, S., & Hamilton, F. (2011, 22 10). Lithuanians Jail Real IRA Gunrunner who fell for an MI5 sting; Real IRA. *The Times*.

Page, M., Taylor, J., & Blenkin, M. (2011). Forensic Identification Science Evidence since Daubert: Part II--Judicial Reasoning in Decisions to Exclude Forensic Identification Evidence on Grounds of Reliability. *Journal of Forensic Sciences*, 913-917.

Papillon Systems. (2017, 02 13). *Papillon Systems*. Retrieved from www.papillon.ru/eg/7/

Parker, C. (2012). *Forensic Investigation*. Retrieved from https://www.studyblue.com/notes/note/n/forensic-investigation/deck/2744067

Pernkopf, F., & O'Leary, P. (2003). Image Acquisition Techniques for Automatic Visual Inspection of Metallic Surfaces. *NDT&E International*, 608-617.

Rendle, D. (2005). Advances in Chemistry Applied to Forensic Science. *Chemical Society reviews*, 1021-1030.

Rifling Manufacturing . (2017, 07 25). Retrieved from Firearms History: http://firearmshistory.blogspot.co.uk/2010/05/rifling-manufacturing-cut-rifling.html

Riva, F., & Champod, C. (2014). Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases. *Journal of Forensic Sciences*, 637-647.

Roady, J. (2017). The PCAST Report. *Criminal Justice*, 9-39.

Roberts, C., & Innes, M. (2009). The 'Death' of Dixon?: Policing Gun Crime and the end of the Generalist Police Constable in England and Wales. *Criminology & Criminal Justice*, 337-357.

Sakarya, U., Leloğlu, U., & Tunalı, E. (2008). Three-Dimensional Surface Reconstruction for Cartridge Cases using Photometric Stereo. *Forensic Science International*, 209-217.

Sarıbey, A., & Hannam, A. (2013). Comparison of the Class and Individual Characteristics of Turkish 7.65 mm Browning/.32 Automatic Caliber Self-Loading Pistols with Consecutive Serial Numbers. *Journal of Forensic Sciences*, 146-150.

Sarıbey, A., Hannam, A., & Tarımcı, Ç. (2009). An Investigation into Whether or Not the Class and Individual Characteristics of Five Turkish Manufactured Pistols Change During Extensive Firing. *Journal of Forensic Sciences*, 1068-1072.

Schroettner, H., Schmied, M., & Scherer, S. (2006). Comparison of 3D Surface Reconstruction Data from Certified Depth Standards Obtained by SEM and an Infinite Focus Measurement Machine (IFM). *Microchimica Acta*, 279-284.

Scott, P., Zeng, W., & Xiang, J. (2011). *Wavelet Analysis for the Extraction of Morphological Features for Orthopaedic Bearing Surfaces.* Intech.

Senin, N., Groppetti, R., Garofano, L., Fratini, P., & Pierni, M. (2006). Three-Dimensional Surface Topography Acquisition and Analysis for Firearm Identification. *Journal of Forensic Sciences*, 282-295.

Song, J. (2015). Proposed "Congruent Matching Cells (CMC)" Method for Ballistic Idenfitication and Error Rate Estimation. *AFTE Journal*, 177-185.

Song, J., Ma, L., Whitenton, E., & Vorburger, T. (2005). 2D and 3D Surface Texture Comparisons using Autocorrelation Function. *Key Engineering Materials*, 437-440.

Song, J., Vorburger, T. V., Ballou, S., Thompson, R. M., Yen, J., Renegar, T. B., Zheng, Silver, R.M., Ols, M. (2012). The National Ballistics Imaging Comparison (NBIC) Project. *Forensic Science International*, 168-182.

Song, J., Vorburger, T., Ballou, S., Ma, L., Renegar, T., Zheng, A., & Ols, M. (2009). Traceability for Ballistics Signature Measurements in Forensic Science. *Measurement*, 1433-1438.

Spapens, T. (2007). Trafficking in Illicit Firearms for Criminal Purposes within the European Union. *European Journal of Crime, Criminal Law and Criminal Justice*, 359-381.

Spiegelman, C., & Tobin, W. A. (2013). Analysis of Experiments in Forensic Firearms/toolmarks Practice Offered as Support for Low Rates of Practice Error and Claims of Inferential Certainty. *Law, Probability and Risk*, 115-133.

Sutherland, L. (1996). Managing Forensic Technology goes Ballistic Computers / The Montreal Firm has Developed a Scanning System that is Revolutionizing Police Work. *The Globe and Mail*.

The Economist. (2013, 10 3). Guns for Hire. *The Economist*.

the International Organisation for Standardisation. (2012). Surface texture: Areal- Part 2: Terms, definitions and surface texture parameters. *Geometrical product specifications (GPS)*. the International Organisation for Standardisation.

Thomas, J. (2011). The Importance of Scientific and Technical Innovation in the Police Investigation of Gun Crime. The University of Huddersfield.

Thomas, T. (1999). *Rough Surfaces.* London: Imperial College Press.

Thomson, G. (2017, 2 18). *Integrated Ballistics Identification System (IBIS)*. Retrieved from http://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/integrated-ballistics-identification-system-ibis

Tobin, W., & Blau, P. (2013). Hypothesis Testing of the Critical Underlying Premise of Disernable Uniqueness in Firearms- Toolmarks Forensic Practice. *Jurimetrics*, 121-142.

Tulleners, F. (2001). *Technical Evaluation: Feasibility of a Ballistics Imaging Database for all new Handgun Sales.* Evalutation for California Department of Justice.

Turner, I. (2017). Arming the Police in Britain. *The Police Journal*, 107-127.

Udupa, G., Singperumal, M., Sirohi, R., & Kothiyall, M. (2000). Characterization of Surface Topography by Confocal Microscopy: I. Principles and the Measurement System. *Measurement Science and Technology*, 305-314.

Vickery, W. (2015). *Advanced Gunsmithing : A Manual of Instruction in the Manufacture, Alteration, and Repair of Firearms.* New York: Skyhorse Publishing.

Vorburger, T. V., Song, J., Chu, W., Ma, L., Bui, S. H., Zheng, A., & Renegar, T. B. (2011). Applications of Cross-Correlation Functions. *Wear*, 529-533.

Vorburger, T., Song, J., & Petraco, N. (2015). Topography Measurements and Applications in Ballistics and Tool Mark Identifications. *Surface Topography; Metrology and Properties*, 1-35.

Waite, M. (2012). *Paperback Oxford English Dictionary.* Oxford: Oxford University Press.

Walton, K., Addinall, K., Zeng, W., & Blunt, L. (2017). Focus Variation with Integrated Form Removal for Forensic Comparison of Bullet Striations. *21st Triennial meeting of the IAFS International Association of Forensic Sciences 2017* (p. 201). Toronto: Elsevier.

Walton, R. (2006). *Cold Case Homicides: Practical Investigative Techniques.* CRC.

Wang, X., Shi, T., Liao, G., Zhang, Y., Hong, Y., & Chen, K. (2017). Using Wavelet Packet Transform for Surface Roughness Evaluation and Texture Extraction. *Sensors*, 933.

Warlow, T. (2012). *Firearms, the Law, and Forensic Ballistics.* CRC.

Weiss, K., Watson, C., & Xuan, Y. (2014). Frye's Backstory: a Tale of Murder, a Retracted Confession, and Scientific Hubris. *The Journal of the American Academy of Psychiatry and the Law*, 226-233.

Whitehouse, D. (2011). *Handbook of Surface and Nanometrology.* Boca Raton: CRC.

Wilson, R., Jopek, L., & Bates, C. (2010). Sharing Ballistics data across the European Union. In *Computation tools 2010, The first internation conference on computational logics, algebras, programming, tools and benchmarking.* Lisbon: IARIA.

Xie, F., Xiao, S., Blunt, L., Zeng, W., & Jiang, X. (2009). Automated Bullet-Identification System based on Surface Topography Techniques. *Wear*, 518-522.

Yates, S., Akghar, B., Bates, C., Jopek, L., & Wilson, R. (2011). A Platform for Discovering and Sharing Confidential Ballistic Crime Data. *International Journal of Knowledge Web Intelligence*, 202-218.

Zahouani, H., Mezghani, S., Vargiolu, R., & Dursapt, M. (2008). Identification of Manufacturing Signature by 2D Wavelet Decomposition. *Wear*, 480-485.

Zeng, W., Jiang, X., & Scott, P. (2010). Fast Algorithm of the Robust Gaussian Regression Filter for Areal Surface Analysis. *Measurement Science and Technology*, 1-9.

Zhang, H., Song, J., Tong, M., & Chu, W. (2016). Correlation of Firing Pin Impressions based on Congruent Matching Cross-sections (CMX) method. *Forensic Science International*, 186-193.

Zou, Y., Li, Y., Kaestner, M., & Reithmeier, E. (2016). Low-coherence Interferometry Based Roughness Measurement on Turbine Blade Surfaces using Wavelet Analysis. *Optics and Lasers in Engineering*, 113-121.