A MULTI-VIEW INTEREST POINT BASED APPROACH TO PHOTOMETRIC REALISM WITHIN AUGMENTED REALITY SYSTEMS

M. Bingham, D. Taylor, D, Gledhill And Z. Xu
University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ABSTRACT

Integrating virtual objects into real environments presents a number of technical and aesthetic challenges which limit the impressiveness of current augmented reality applications. Although some of these challenges have recently been addressed, others are still active areas of research. The quality of the models and textures used, accurate geometric world alignment and the quality of illumination correspondence all need to be considered. Literature shows the latter to be the least mature field and is the focus of this paper. This paper introduces a new photometric registration technique that makes use of image interest points. The technique attains illumination correspondence without the need for pre-calibration or unnatural calibration objects, instead natural image features such as shadow and object interest points are use. The operational complexity of the technique is low and real-time processing of live data is achieved. Current progress and future work are discussed in the paper.

Keywords augmented reality, realism, interest points, shadow

INTRODUCTION

Augmented reality (AR) is the term used to describe the concept of superimposing virtuality over images of the real world, effectively combining real and artificial environments. AR has many areas of application and in recent years the field has begun to receive interest from a number of sectors such as manufacturing, military, medical and the computer games industry. One example is the Battlefield Augmented Reality System (BARS) developed by the US Naval Research Facility. BARS is a wearable device that attempts to gather intelligence from, and provide real-time information on, a soldier's surroundings using augmented reality [5]. AR gaming applications such as ARQuake have been developed and allow users to interact with virtual enemies in their own every day environment. The authors of [17] have implemented an AR system known as Tinmith which allows the user to construct AR outdoor structures via visually tracked hand movements. The Tinmith system has since been adapted for a number of applications including medical, security, entertainment, navigation, shopping, maintenance and has military potential. The alignment between the real and virtual worlds must be accurate in order to achieve realistic augmentation. The process of obtaining such alignment is known as geometric registration. A number of approaches have been proposed that use either sensor data, visual cues or a hybrid combination of both. Photometric registration is the matching of light conditions between worlds. This involves detecting the pose and light qualities of any illuminant affecting the real component of the scene. A number of photometric registration methods have been explored. Researchers attempt to estimate real-world illumination conditions by gathering various metrics from the real scene. This data is then used to illuminate the artificial component. An overview of such techniques is given in section 2.

Despite recent advances in virtual reality systems, the believable integration of real and virtual components is still a challenge. The realism of an augmented reality scene is massively dependant on sound geometric and photometric registration. The geometric problem has mostly been addressed by use of either fiducial markers or interest point based tracking however photometric registration is still problematic. This is primarily due to the unpredictable nature and complexities of the real world. Existing photometric registration techniques have limitations. These limitations include high computational complexity, the need to pre-calibrate the scene, continuous artificial object-based calibration during runtime and constraints in the operational environment. Computationally complex techniques would take too long to perform the necessary calculations. Any lag time this induced would cause desynchronisation between the two worlds, ultimately reducing realism and would therefore not be suitable for augmented reality application. If it took too long to recalculate an illuminant position the virtual lighting conditions may not match the real conditions for some time. Or worse, the computation may reduce the output framerate causing the scene to jitter or freeze. Techniques that require pre-calibration are often less computationally intense, however they make the assumption that lighting conditions and camera position do not change. A number of techniques only operate under certain conditions, for example, in a room of known geometry where the light sources have been manually positioned in the virtual scene. Techniques that require constant calibration at runtime require some form of calibration object. These objects are usually unnatural in appearance and therefore destroy the believability of the scene in the same way that the deployment of fiducial marker would when geometrically registering a scene. This paper focuses on photometric registration and conceptually outlines an interest point based technique that would allow for fast detection of a single illuminant using natural calibration geometry and shadows present within a scene.

RELATED WORK

Literature shows that researchers have attempted to photometrically register augmented reality worlds in a number of ways. Most existing techniques work well in constrained environments but fail if certain conditions are not met. Other techniques operate well in less constrained environments, however computational complexity is high and therefore unfeasible for real-time augmented reality processing. Feng [6] suggests a technique that makes use of spheres with Lambert surfaces as calibration objects in order to gather illumination parameters. The author claims to achieve an identical match between real and virtual components, the result being a seamless augmented reality scene. This technique operates in real-time with relatively low operational complexity but fails if multiple real light sources are present. The technique is not suited for combination with any geometric registration approach as a stationary camera is required once pre-calibration has taken place. Feng does not observe or attempt to reproduce cast shadows.

Jacobs [7] presents a real-time rendering solution that simulates colour consistent virtual shadows in real-scenes. Shadow regions are estimated and then confirmed using texture information and are segmented using canny edge detection. A binary mask is then used to track which pixels contain shadow information. This information is then analyzed and the data gathered is passed to either a shadow volume or shadow map algorithm which allows for the casting of shadows from artificial objects. Shadows are cast onto both virtual and real objects. Jacobs technique out performs other shadow matching techniques as the technique correctly combines real and virtual shadows without producing an unrealistic overlap. Feng identifies a number of illumination methods for augmented reality and classifies them into two categories. These are common illumination and relighting. Common illumination matching techniques attempt to simulate consistent lighting when artificial objects are inserted into a real context. Relighting techniques modify the real component in response to the insertion of a virtual object. He performs both by making use of the inverse illumination technique discussed by Patow [18]. This technique collects illumination parameters such as the Bidirectional Reflectance Distribution Function (BRDF) from the real scene for use within the virtual. The technique requires that approximate knowledge of real scene geometry be known prior to augmentation. State et al [9] propose an AR system that favours the use of shadow maps and Haller [10] suggests the use of shadow volume techniques. Both techniques allow the AR system to simulate shadows at low operational cost, after the real-world illumination data has been acquired. Yao [11] and Siala [12] present methods of locating shadows within an image, but do not perform any analysis of the data obtained.

Wang[13] presents a method of detecting multiple illuminants within a scene and accurately estimating their pose. This method does not require the use of a pre-calibration object. Additionally the data collected from the technique allows for the virtual recreation of three dimensional object shapes. The illuminant detection results the technique yields are directly applicable to development of realistic augmented reality systems, however the calculations required are slow and therefore would not be capable of processing a live video stream in real-time. Wang's technique provides good results compared to a number of other techniques as it analyzes both shadows and the shading of arbitrary scene objects. The technique finds it easy to obtain multiple illuminant information from shading when specular reflections are present but finds the task difficult when observing diffuse reflections alone. This technique adds robustness as it is less prone to error caused by cast shadows moving outside the camera's field of view, or being occluded *by* techniques that observe either object shading or cast-shadows exclusively. Zhang [14] presents a robust method of estimating the azimuth of a single illuminant.

Successful interest point (IP), shadow and object detection and geometric registration are a pre-requisite of the technique presented within this paper. An interest point is defined as a two-dimensional signal change; for example, where there is a corner, an edge or where the texture changes significantly [1]. Much work has been undertaken in the field of interest point detection. Many IP detection techniques have been presented by Harris and Stephens [2] and Lowe [15]. Much has been accomplished in the field of detecting and segmenting shadows and object within images and video footage. The task of geometrically registering AR by aligning the real and virtual worlds has been achieved using fiducial markers as per [4] and by tracking world interest points as detailed by State [9].

THE MULTI-VIEW INTEREST POINT-BASED TECHNIQUE

The proposed technique aims to detect the illumination conditions of a real scene by studying shadow and object interest points. It is assumed that only one illuminant is present and scene geometry is casting shadows onto planar surfaces. The technique functions by observing corresponding shadow and object interest points. This allows for the estimation of the illuminant position, in 2D, for two input images. These results are then reverse projected into 3D and combined to obtain 3D coordinates. The resultant data can be used to create augmented reality scenes with virtual components that are lit the same way as the real world.

The eight primary stages are:

1. Acquire input images
2. Identify IPs
3. Identify IP correspondences
4. Locate illuminant vector from two different angles
5. Convert illuminant vector into 3D
6. Locate illuminant in 3D space
7. Perform geometric registration
8. Augment reality

Input images are typically captured from two real-world input devices simultaneously. However three dimensional renders can also be used. The images can be taken from almost any angle. The prototype system can automatically detect the angle between cameras and providing that sufficient information is present within both images such as a marker or object of known geometry. Figure 1 shows sample input data in the form 3D virtual renders. An extensive data set has been generated to test the response of the prototype to various conditions in a controlled manner that would be difficult to reproduce using photographs.

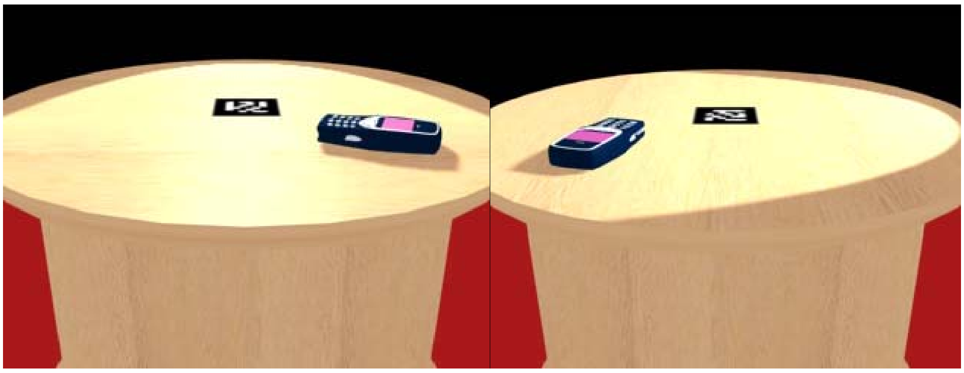

Figure 1 – Observing a scene from the front & side

variant Feature Transform (SIFT) method presented by Lowe [18] is preferred as it makes available additional information that is of use when detecting the correspondence between shadow interest points and object interest points. Figure 2 shows SIFT features extracted from an image. The image shows many more SIFT descriptors than would be required for the intended purpose. SIFT allows for a threshold to be applied that would reduce the number of feature extractions, limiting output to corners on objects or shadows within the scene. Other feature detection techniques such as presented by Harris [2] would provide interest point information however the direction data that assists with finding shadow and IP correspondences.

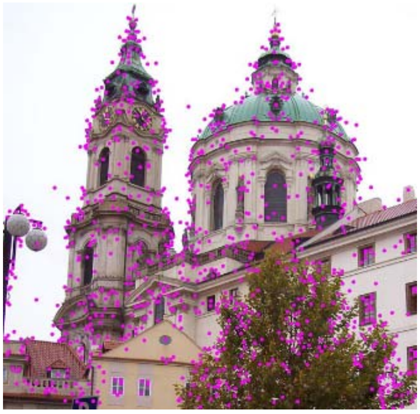

Figure 2 –SIFT IPs

Once interest points have been obtained they need to be classified as being associated with either a cast shadow or object geometry. Correspondences between object IPs and shadow IPs need to be defined. We need at least two correspondences per image, but more will usually offer improved accuracy. Figure 3 shows correspondences between shadow and object IPs.

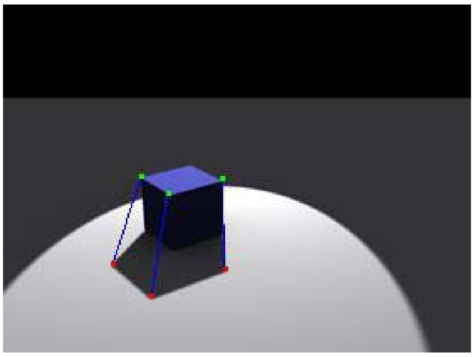

Figure 3 – Shadow & object IPs

By casting a line through the shadow and object interest point for multiple correspondences the approximate location of the illuminant in 2D coordinate space can be obtained. Intersections between two or more lines indicate a potential illuminant position. Any anomalous results are discarded and the average intersection point is taken as the 2D illuminant position. It should be noted that the actual illuminant position may be outside the image boundary; this does not present a problem. Figure 4 shows two correspondence lines and where these intersect for both images.

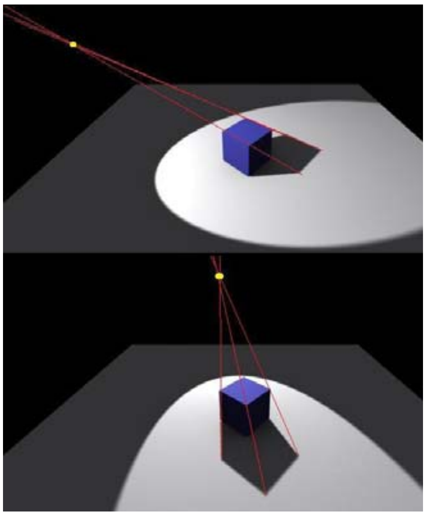

Figure 4 – The 2D intersection

When displaying 3D geometry on a flat 2D surface such as a computer monitor we need to perform a projection transform. This is the conversion from 3D coordinate space to 2D screen space. The 2D coordinates within the image can be treated as screen coordinates post projection. Therefore a reverse projection transform can be performed to obtain the 3D coordinates. A reverse projection allows for the derivation of a ray projecting into 3D space for any given screen pixel. Exact depth information is not available; therefore two calculations are performed at different depths. This gives us two points on a line within 3D space on which the illuminant lies. To obtain depth information we repeat the process for the second input image. Once two lines in 3D space have been obtained and the angle between cameras is known the position of the illuminant can be determined. This is achieved by rotating the second line by the angle between the two input cameras. We then calculate the 3D intersection of these two points. The point attained is an approximate illuminant position. Figure 5 shows the 2D intersections for both images. A ray is then cast through each plane at both 2D intersection points. Although the figure does not account for projective distortion the prototype application does. The 3D intersection point of these rays can be seen and the derived illuminant position is marked.

Once the light source has been located we can perform the augmentation. Geometric registration techniques should be applied to ensure accurate alignment between the real and virtual worlds. Once an artificial illuminant is positioned and augmented objects are registered to the same coordinate system, augmented and real objects will appear to be lit in the same manner. The scene can now be passed through AR shadow casting techniques such as those presented by Jacobs [7] and Haller [10]. The generated shadows will appear to be cast from the same light source as actual shadows. Once conditions are correctly matched, the world is photometrically and geometrically registered.

4 CONCLUSIONS

Both geometric and photometric registration are important for any realistic augmented reality application. As geometric techniques become more reliable, research focus has turned to photometric registration. A number of photometric registration techniques have been proposed; however they are often computationally complex and although they may work well for static images, real-time techniques are required for live video based systems. Existing real-time techniques attempt require either pre-calibration or continuous calibration using known objects. Such techniques either introduce additional artificial components to the scene or place constraints on the system. Realistic real-time photometric registration has not yet been achieved. The technique outlined in section 3 addresses this problem and, by utilising an interest point based approach, low computational complexity is achieved. Although some assumptions are made, the technique does not massively constrain the operational environment beyond the need for sufficient cast shadows and scene geometry to be visible. The technique introduced in this paper is currently being refined and its robustness to phenomena such as occlusions, complex light conditions and input image quality are being tested.

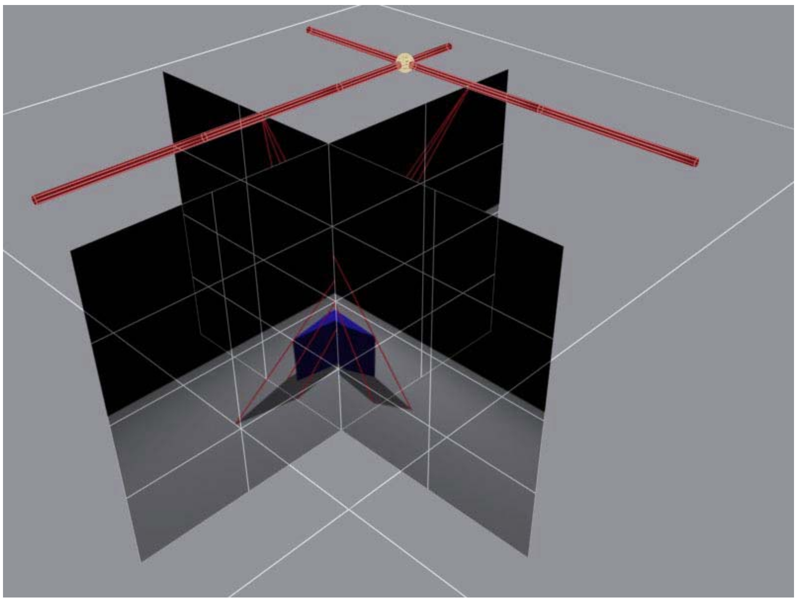

Figure 5 – The 3D intersection

Current work is creating a mathematical model of the working system and using it to parameterise and optimise the variables in the system; camera angles, camera focal length, image resolution, shadow clarity, distance from interest points to light source, shadow projection etc

REFERENCES

[1] D. Gledhill and G. Tian and D. Taylor and D. Clarke.Panoramic Imaging - A Review. Computers and Graphics, 2003.
[2] C. Harris and M. Stephens.A Combined Corner and Edge Detector. 4th Alvey Vision Conference, 1988.
[3] B. Thomas and C. Sandor.What Wearable Augmented Reality Can Do For You IEEE Pervasive Computing, 2009.
[4] X. Zhang and S. Fronz and N. Navab.Visual Marker Detection and Decoding in AR. In Proceeding International Symposium on Mixed and Augmented Reality, 2002.
[5] T. Cowper and M. Buerger.Improving Our View of The World: Police and Augmented Reality Technology.
[6] Y. Feng.Estimation of Light Source Environment for Illumination consistency or Augmented Reality. In Proceedings 2008 IEEE Congress on Image and Signal Processing, 2008.
[7] K. Jacobs and C. Angus and C.Loscos.Automatic Generation of Consistent Shadows for Augmented Reality. In Proceedings Graphics Interface, 2005.
[8] K. Jacobs and C. Loscos.Classification of Illumination Methods for Mixed Reality. In Proceedings Computer Graphics Forum, 2004.
[9] A. State and G. Hirota and C. Chen and W. Garrett and M. Livinston.Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. In Proceedings of the Computer Graphics Forum, 2006.
[10] M. Haller and S. Drab and W. Hartman.A Realtime Shadow Approach for An Augmented Reality Application Using Shadow Volumes. In Proceedings ACM Symposium on Virtual Reality Software and Technology, 2003.
[11] J. Yoa and Z. Zhang.Systematic Static Shadow Detection. In Proceedings 17th IEEE International Conference on Pattern Recognition, 2004.
[12] K. Siala and M. Chakchouk and O. Besbes and F. Chaieb.Moving Shadow Detection with Support Vector
Domain Description in the Color Ratios Space. In Proceedings of the 17th IEEE International Conference on Pattern Recognition, 2004.
[13] Y. Wang and D. Samaras.Estimation of Multiple Directional Light Sources for Synthesis of Augmented Reality Images. Special Issue on Pacific Graphics, 2003.
[14] X. Zhang and S. Fronz and N. Navab.Visual Marker Detection and Decoding in AR Systems: A Comparative Study. In Proceedings International Symposium on Mixed and Augmented Reality, 2002.
[15] D. Lowe.Distinctive Image Features from Scale- Invariant Keypoints. International Journal of Computer Vision, 2004.
[16] M. Tuceryan and D. Greer and R. Whitaker and D. Breen and C. Crampton and E. Rose and K. Ahlers.Calibration Requirements and Procedures for a Monitor Based Augmented Reality System. IEEE Transactions on Visualization and Computer Graphics, 1995.
[17] W. Piekarski and B Thomas.The Tinmith System: Demonstrating New Techniques for Mobile Augmented Reality Modeling. Computuer Science Community, 2002.
[18] G. Patow and X. Pueyo.A Survey of Inverse Rendering Problems. In Proceedings Computer Graphics Forum, 2006