Abstract. The analysis and the description of complex visual scenes characterized by the presence of many objects of interests involve reasoning on spatial relations such as “above”, “below”, “before”, “after” and “between”. In this context, we have defined these semantic concepts in terms of ternary spatial relations and we have formalized them using the clock model which is based on the clock-face division and the semantic notions of hours to describe relative spatial positions. The presented approach has been efficiently applied for the automated understanding of spatial relations between multiple objects in real-world computer vision image datasets.

Keywords: Ternary Spatial Relations; Clock Model; Qualitative Spatial Reasoning; Computer Vision; Visual Scene Understanding.

1 Introduction

Modelling spatial relations among objects of visual scenes is greatly of benefit to visual applications [1]. Indeed, their integration into vision systems brings an additional level in the task of automatic image understanding, leading to the processing of semantic information besides those provided by visual features. Furthermore, the definition of spatial relations allows the action of reasoning on semantically meaningful concepts which is a major advantage [3] compared with traditional vision approaches using only quantitative techniques or annotating images with just sparse words.

In the literature, most of the spatial relations [2], [4] have been defined as binary ones, such as the topological spatial relations like the RCC-8 model [5] or the cardinal spatial relations and their fuzzy extension [6].

In context of visual scene description and analysis, [7] introduced a new formalism for modeling the image space as a clock face and they proposed a series of related spatial relations, including the directional spatial relations of the scene objects and the far/close relations. However, to perform reasoning on spatial relations among a greater number of objects (at least three), there is a need for relations such as the ternary ones. In fact, little attention has been paid to study them. Their formal geometric modeling has been mostly studied for geographic information systems (GIS) [8]. Indeed, [9] developed the 5-intersection model,
but its formalism leads to a restricted range of applications. On the other hand, [10] proposed for the biomedical imaging purpose a fuzzy definition of the ternary spatial relation *between*. Despite its improvement by [11], it does not fit well for computer vision applications such as crowd’s behaviour study.

In this work, we present the extension of the clock approach [7] to formalize the fundamental ternary spatial relation, namely, *between* (*bt*), and to model the semantic concepts *above* (*ab*), *below* (*bl*), *before* (*bf*), and *after* (*af*) as ternary spatial relations.

We implement these relations using Description Logics (DL) [12] which have been widely adopted for knowledge representation in visual systems [13], [14], [15], [16].

Thus, our clock-modeled ternary spatial relations define a set of useful notions to characterize visual scenes involving numerous objects of interest as well as to acquire knowledge about them, and could be incorporated in a complete system for automatic reasoning on spatial relations among objects detected in images.

The contributions of this paper are as follows:

– the modeling of ternary spatial relations using the clock-face approach;
– the architecture of the full system combining visual face recognition and spatial reasoning.

The paper is structured as follows. In Section 2, we present our approach using the clock formalism to model ternary spatial relations such as *above*, *below*, *before*, *after*, and *between*. All these relations have been integrated in a framework for automatic face recognition and reasoning as described in Section 3. The resulting system has been successfully tested on still image datasets as reported and discussed in Section 4. Conclusions are drawn up in Section 5.

2 Clock-Modeled Ternary Spatial Relations

In Section 2.1, we first introduce the clock model which is semantically meaningful and used to defined spatial relative relations, while the definitions of the ternary spatial relations formalized with this clock model are presented in Section 2.2.

2.1 Clock Model

The clock concept introduced by [7] consists in dividing the image plane in twelve parts around any object of interest of the scene as illustrated in Fig. 1(a). Hence, each portion of the space is then corresponding to an hour. This leads to a semantically meaningful division of the space as a clock face. This concept helps in reducing the uncertainty on the directional relative positions between objects in crowded scenes, in which case traditional binary relations such as *left* or *right* are not enough discriminant as demonstrated in [7]. In this work, the clock notion is used for the formal specification of our ternary relative directional relations *above*, *below*, *before*, *after*, and *between*. 

For this purpose, we introduce the Quadrant \((Q)\) concept as shown in Fig. 1(a). To provide an example of how we have formalized it, we define \(\text{isInQuadrant}I\) in DL as follows. Let \(O_{\text{REF}}\) be the object of reference and the \(O_{\text{REL}}\) the target object. Giving that \(\text{Angle}\) is the relative angle between the line \(O_{\text{REF}} - O_{\text{REL}}\) and the axis \(X\) of the analyzed image plane, then

\[
\text{isInQuadrant}I \subseteq \text{Spatial\_Relation} \land \exists \text{hasReferentObject.}O_{\text{REF}} \land \exists \text{hasTargetObject.}O_{\text{REL}} \land (\exists \text{hasAngle.Angle}12 \land \exists \text{hasAngle.Angle}1 \land \exists \text{hasAngle.Angle}2)
\]

\[(1)\]

with

\[
\text{Angle}2\text{clock} \equiv \text{Angle} \land \exists \text{Angle.value} \leq \frac{\pi}{6} \land \exists \text{Angle.value} > 0
\]

\[(2)\]

Equation (2) denotes the set of angles which have values lower than or equal to \(\pi/6\) and higher than 0. In this example, \(\text{value} \leq \frac{\pi}{6}\) is a predicate over the real number domain \(\mathbb{R}\). \(\text{Angle}12\text{clock}\) and \(\text{Angle}1\text{clock}\) concepts are defined similarly. We can note also that any \(O_{\text{REL}}\) lies at least in one of the four quadrants \(QI, QII, QIII\) or \(QIV\).

### 2.2 Ternary Spatial Relations

We adopt the notation \(rl(A, B, C)\) for a ternary relation \(rl\) among three objects \(A, B,\) and \(C\). The first object \(A\) involved in this relation is considered to be the target object, whereas the two other objects \(B\) and \(C\) are the reference objects. Thus, \(rl(A, B, C)\) denotes that “\(A\) is in the relation \(rl\) with \(B\) and \(C\)”.

In fact, the order of the reference objects \(B\) and \(C\) is important as it affects the orientation of the relation, in this case from the reference object \(B\) to the reference object \(C\). While in some relations the role of the three objects can be exchanged without

![Fig. 1. Illustration of the ternary spatial relations between visual objects using the clock model (a) and representing the semantic concepts which are above \((ab)\), below \((bl)\), before \((bf)\), after \((af)\), and between \((bt)\) (b).](image-url)
affecting the relation, in some relations the swapping of the arguments leads to a change of the relation.

In the remaining of this section, we present the detailed definitions of the relations above, below, before, and after which are formalized in this work as ternary spatial relations in opposite to [10] or [11], as well as the relation between, and we mention the cases when the exchange of the arguments modifies the described relations. It is worth to note that the definitions are valid for both convex and concave objects and that the center of each related clock is set to the centroid of the corresponding object.

Spatial Relation Above We consider the relation above as a ternary spatial relation where \( \text{ab}(O_{\text{REL}}, O_{\text{REF}1}, O_{\text{REF}2}) \) means that the target object \( O_{\text{REL}} \) is above both the reference object \( O_{\text{REF}1} \) and the reference object \( O_{\text{REF}2} \). For this relation, the order of the reference objects cannot be inverted otherwise the type of the relation is modified. Indeed, if the target object is above only one of the reference objects, other ternary spatial relations can be then applied. Hence, this ternary modeling leads to a more discriminating relation than the traditional ones in particular in the case of crowded scene analysis.

In DL, the concept \( \text{isAbove} \) is defined as follows. Let \( O_{\text{REF}1} \) and \( O_{\text{REF}2} \) be the two objects of reference, while \( O_{\text{REL}} \) is the object of interest. Considering definitions such as expressed by Eqs. (1) and (2), then

\[
isAbove \sqsubseteq \text{Spatial Relation} \cap \text{Ternary Spatial Relation} \\
\sqcap \exists \text{hasReferentObject.} O_{\text{REL}} \sqcap \exists \text{hasReferentObject.} O_{\text{REF}1} \\
\sqcap \exists \text{hasReferentObject.} O_{\text{REF}2} \\
\sqcap \exists \text{hasTargetObject.} O_{\text{REL}} \sqcap (\exists \text{isInQuadrantIVOf.} O_{\text{REF}1} \sqcup \exists \text{isInQuadrantIVOf.} O_{\text{REF}2}) \\
\sqcap (\exists \text{isInQuadrantIVOf.} O_{\text{REF}1} \sqcup \exists \text{isInQuadrantIVOf.} O_{\text{REF}2}) .\]

(3)

This concept is illustrated in Fig. 1(b).

Spatial Relation Below We consider the relation below as a ternary spatial relation where \( \text{bl}(O_{\text{REL}}, O_{\text{REF}1}, O_{\text{REF}2}) \) means that the target object \( O_{\text{REL}} \) is below both the reference object \( O_{\text{REF}1} \) and the reference object \( O_{\text{REF}2} \). For this relation, the order of the reference objects cannot be inverted otherwise the type of the relation is modified. Indeed, if the target object is below only one of the reference objects, other ternary spatial relations can be then applied. Hence, this ternary modeling leads to a more discriminating relation than the traditional ones in particular in the case of crowded scene analysis.

In DL, the concept \( \text{isBelow} \) is defined as follows. Let \( O_{\text{REF}1} \) and \( O_{\text{REF}2} \) be the two objects of reference, while \( O_{\text{REL}} \) is the object of interest. Considering
definitions such as expressed by Eqs. (1) and (2), then

\[
\text{isBelow} \sqsubseteq \text{Spatial Relation} \cap \text{Ternary Spatial Relation}
\]

\[
\sqcap \exists \text{hasReferentObject.}O_{REF1} \sqcap \exists \text{hasReferentObject.}O_{REF2}
\]

\[
\sqcap \exists \text{hasTargetObject.}O_{REL} \sqcap (\exists \text{isInQuadrantIIOf.}O_{REF1}
\]

\[
\sqcup \exists \text{isInQuadrantIIIOf.}O_{REF1} \sqcup (\exists \text{isInQuadrantIIIOf.}O_{REF2}
\]

\[
\sqcup \exists \text{isInQuadrantIVOf.}O_{REF2}).
\]

(4)

A representation of this concept is depicted in Fig. 1(b).

**Spatial Relation Before** We consider the relation *before* as a ternary spatial relation where \(bf(O_{REL},O_{REF1},O_{REF2})\) means that the target object \(O_{REL}\) is before both the reference object \(O_{REF1}\) and the reference object \(O_{REF2}\). For this relation, the order of the reference objects cannot be inverted otherwise the type of the relation is modified. Indeed, if the target object is before only one of the reference objects, other ternary spatial relations can be then applied. Hence, this ternary modeling leads to a more discriminating relation than the traditional ones in particular in the case of crowded scene analysis.

In DL, the concept *isBefore* is defined as follows. Let \(O_{REF1}\) and \(O_{REF2}\) be the two objects of reference, while \(O_{REL}\) is the object of interest. Considering definitions such as expressed by Eqs. (1) and (2), then

\[
\text{isBefore} \sqsubseteq \text{Spatial Relation} \cap \text{Ternary Spatial Relation}
\]

\[
\sqcap \exists \text{hasReferentObject.}O_{REF1} \sqcap \exists \text{hasReferentObject.}O_{REF2}
\]

\[
\sqcap \exists \text{hasTargetObject.}O_{REL} \sqcap (\exists \text{isInQuadrantIIOf.}O_{REF1}
\]

\[
\sqcup \exists \text{isInQuadrantIIIOf.}O_{REF1} \sqcup (\exists \text{isInQuadrantIIIOf.}O_{REF2}
\]

\[
\sqcup \exists \text{isInQuadrantIVOf.}O_{REF2}).
\]

(5)

An illustration of this concept is depicted in Fig. 1(b).

**Spatial Relation After** We consider the relation *after* as a ternary spatial relation where \(af(O_{REL},O_{REF1},O_{REF2})\) means that the target object \(O_{REL}\) is after both the reference object \(O_{REF1}\) and the reference object \(O_{REF2}\). For this relation, the order of the reference objects cannot be inverted otherwise the type of the relation is modified. Indeed, if the target object is after only one of the reference objects, other ternary spatial relations can be then applied. Hence, this ternary modeling leads to a more discriminating relation than the traditional ones in particular in the case of crowded scene analysis.

In DL, the concept *isAfter* is defined as follows. Let \(O_{REF1}\) and \(O_{REF2}\) be the two objects of reference, while \(O_{REL}\) is the object of interest. Considering
definitions such as expressed by Eqs. (1) and (2), then

\[
\text{isAfter} \subseteq \text{Spatial	extunderscore Relation} \cap \text{Ternary	extunderscore Spatial	extunderscore Relation} \\
\quad \cap \exists \text{hasReferentObject}.O_{\text{REF}_1} \cap \exists \text{hasReferentObject}.O_{\text{REF}_2} \\
\quad \cap \exists \text{hasTargetObject}.O_{\text{REL}} \cap (\exists \text{isInQuadrantIOf}.O_{\text{REF}_1} \\
\quad \cup \exists \text{isInQuadrantIIOf}.O_{\text{REF}_1}) \cap (\exists \text{isInQuadrantIOf}.O_{\text{REF}_2} \\
\quad \cup \exists \text{isInQuadrantIIOf}.O_{\text{REF}_2}).
\]

(6)

This concept could be visualized in Fig. 1(b).

Spatial Relation Between The relation between is intrinsically a ternary spatial relation. Indeed, \(bt(O_{\text{REL}}, O_{\text{REF}_1}, O_{\text{REF}_2})\) means that the target object \(O_{\text{REL}}\) is between the reference object \(O_{\text{REF}_1}\) and the reference object \(O_{\text{REF}_2}\). In this case, the order of the reference objects can be inverted without changing the semantic meaning of this relation.

In DL, the concept \(isBetween\) is defined as follows. Considering definitions such as expressed by Eqs. (3)-(6), then

\[
\text{isBetween} \subseteq \text{Spatial	extunderscore Relation} \cap \text{Ternary	extunderscore Spatial	extunderscore Relation} \\
\quad \cap \exists \text{inverse.isAbove} \cap \exists \text{inverse.isBelow} \\
\quad \cap \exists \text{inverse.isBefore} \cap \exists \text{inverse.isAfter}.
\]

(7)

This concept is illustrated in Fig. 1(b).

3 Implementation

The ternary spatial relations described in Section 2 could be embedded into a system for the automatic analysis of people localization in imaged scenes as presented in Fig. 2. Indeed, understanding images with groups of people is a complex process which requires more information that just those contained in the extracted visual features. In [17], they propose to add social relations in order to improve the automatic analysis of this kind of images, but their estimations are less satisfactory compared to those we obtain (see Section 4) by adding the presented spatial relations to the vision system.

The developed system is composed of three main phases. The first two steps constitute a vision system for face detection, which has been implemented using the well-established method of [18]. Firstly, faces are learned by training the system on sets of positive and negative examples, respectively. Secondly, the resulting face detector is applied on an image and automatically computes faces’ locations which are then included in corresponding rectangles and labeled. Then, the quantitative data which are extracted by this process are transferred in a similar way to [19] or [14] in order to populate an ontology such as [13]. This ontology is enhanced with the proposed ternary spatial relations. Next, qualitative reasoning is performed on these spatial relations and FaCT++ is used as the reasoner. This last phase of the system thus consists in reasoning on the ternary spatial relations and has been assessed in Section 4.
Experiments and Discussion

The goals of the presented experiments are twofold. On one hand, we assess in a quantitative way the performance of the all five proposed ternary relations compared to the 5-intersection model which is the only one also defining above, below, before, after as ternary relations, but using a different formalism from ours. On the other hand, qualitative assessment of our relations is performed against the ground truth.

In order to evaluate the performance of our formalism of all the ternary proposed relations, our relations have been embedded in the overall system developed for the computer vision application consisting in the analysis of photos with groups of people such as presented in Fig. 2.

To carry out these tests, we have firstly retrieved from Internet images of choirs using Google Image. The aim of the search of choir images was to ensure the finding of pictures of groups of people to analyze spatial relations among them. Indeed, the direct keywords “groups of people” did not produce relevant results. Then, we have constituted a dataset with these 500 retrieved choirs images where faces have been detected and labeled as explained in Section 3. The picture of Fig. 3(a) is an example of the images composing the dataset. Although the study of the face detection problem is out of purpose of the present paper, we can mention that the obtained general precision rate was 95% and that the undetected faces were manually added for the completeness of the dataset.

The adopted criterion for the quantitative assessment of a ternary spatial relation \( r(A, B, C) \) is the satisfaction degree computed as follows

\[
s(A) = \frac{|\text{area}(A) \cap \Gamma_{BC}|}{|\text{area}(A)|},
\]

where \( A \) is the target object and \( \Gamma_{BC} \) is the area between the two reference objects \( B \) and \( C \).
Fig. 3. Results of our system tested for an image of a choir. First column: Face recognition results. Second column: Schematic spatial representation of the above, below, before, after, and between relations of the visual objects detected in the image under study.

The qualitative evaluation of the system is carried out by asking different questions whose answers are boolean. The two main types of possible queries are:

– what are the relation(s) among three given objects $O_{REL}$, $O_{REF1}$, and $O_{REF2}$?
– which is/are the object(s) $O_{REL}$ that has/have the relation $rl$ with the given objects $O_{REF1}$ and $O_{REF2}$?

In the case of the image of the Fig. 3(a), the quantitative and qualitative results are reported in Tables 1 and 2, respectively.

When compared with the ternary relations of [9], we assume the semantic correspondence between their leftside concept and our above concept as well as between their rightside concept and our below concept.

Table 1. Quantitative evaluation of the ternary relations for the objects in the choir image in Fig. 3(a).

<table>
<thead>
<tr>
<th>Approaches</th>
<th>from [9]</th>
<th>ours (Sec. 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_{REL}$</td>
<td>$O_{REF1}$</td>
<td>$O_{REF2}$</td>
</tr>
<tr>
<td>$ab$</td>
<td>$bl$</td>
<td>$bf$</td>
</tr>
<tr>
<td>$B$</td>
<td>$A$</td>
<td>$D$</td>
</tr>
<tr>
<td>$C$</td>
<td>$A$</td>
<td>$B$</td>
</tr>
<tr>
<td>$C$</td>
<td>$A$</td>
<td>$D$</td>
</tr>
<tr>
<td>$C$</td>
<td>$D$</td>
<td>$E$</td>
</tr>
<tr>
<td>$D$</td>
<td>$A$</td>
<td>$C$</td>
</tr>
</tbody>
</table>
Table 2. Qualitative evaluation of the ternary relations for the objects in the choir image in Fig. 3(a).

<table>
<thead>
<tr>
<th>Approaches</th>
<th>ground truth</th>
<th>ours (Sec. 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_{REL}$</td>
<td>$O_{REF1}$</td>
<td>$O_{REF2}$</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

In Table 1, we can observe that in the case of the relation $rl(C, A, D)$, we find that the object $C$ is between $A$ and $D$ when applying our formalism, whereas [9] considers that the object $C$ is also below the objects $A$ and $D$ that is not complying with the human intuition. For the relations such as $rl(B, A, D)$ or $rl(C, D, E)$, our approach provides values which indicate the dominant relation between these objects and which is each time conformed with the human perception of the scene. In opposite, the figures computed using [9] give a large uncertainty about the type of the relations among the objects, e.g. by finding (i) 50% for $C$ below $D$ and $E$ and (ii) 50% for $C$ between $D$ and $E$ which does not indicate which semantic relation is correct and makes confusion between true (ii) and false (i) statements.

In the results of the qualitative reasoning on the proposed ternary spatial relations as reported in Table 2, we note the excellent concordance between the ground truth values set by human users and those computed with our developed system. The overall precision of our system tested for the entire dataset is of 99.5 ± 0.5 %.

Hence, the evaluation of the results shows that our clock-based formalism provides a more accurate and consistent definition of these concepts than the state-of-the art ones.

5 Conclusions

In this paper, we have applied new ternary spatial relations, namely, above, below, before, after, and between, in order to automatically understand and interpret images with complex content such as groups of people. Formalizing the presented relations using the clock model and defining them as ternary relations has provided new powerful semantic concepts to describe the relative position of an object of interest towards two other distinct visual objects. As demonstrated, this conceptualization brings a new insight in the automated analysis of crowded visual scenes.
References