Relative positions in words: A system that builds descriptions around Allen relations

P. MATSAKIS, L. WAWRZYNIAK and J. NI

Department of Computing and Information Science
University of Guelph, Guelph, ON, N1G 2W1, Canada
matsakis@cis.uoguelph.ca, {lwawrzyn, jni}@uoguelph.ca

In this paper, we introduce a system able to generate an intuitive, human-like linguistic description of the topological relationships between two objects. The description includes approximate terms commonly found in daily communications. It attempts to capture the essential characteristics of the relationships, while leaving out superfluous and possibly overwhelming detail. The objects are 2-D image objects. They need not be convex, nor connected, and they may have holes in them. Each description is built around Allen relations, based on information extracted from F-histograms. It consists of a topological component that indicates the primary topological relationships, directional estimates of where these relationships are most prominent, and a self-assessment component which reflects the complexity of the situation. Experiments on synthetic and real data validate the approach.

Keywords: Topological and Directional Relationships; Allen Relations; F-Histograms; Natural Language Descriptions; Fuzzy Sets and Fuzzy Logic.

1 Introduction

In human-human interaction, people often communicate about space through natural language, drawings and gestures, which all carry complementary information. Interfaces based on the first two modalities are expected to expand the community of users and suit the demand for a wide-ranging access to computer systems, including

Geographic Information Systems (Stephanidis 2001). Machines that could comprehend the organization of objects in space and could reason and communicate linguistically about space, like humans do, would be immeasurably helpful in many areas. They would facilitate the routine work of the experienced user. They would be accessible to the visually impaired. They would benefit the everyday, non-expert user, who is not inclined to spend much time on training and does not want to struggle with state of the art WIMP (Windows, Icons, Menus, Pointers) interfaces.

In everyday situations, space is often viewed as a construct induced by spatial relationships, rather than as a container that exists independently of the objects located in it. The past ten to fifteen years have seen significant advancements in the development of mathematical and computational models of distance, directional, and topological relationships between image objects (Bloch 1999, Clementini and Di Felice 1996, Egenhofer and Herring 1994, Guesgen 2002, Krishnapuram *et al.* 1993, Matsakis and Wendling 1999, Nabil *et al.* 1995, Petry *et al.* 2002). Over the same period, several methods have been proposed for generating automated linguistic descriptions of these relationships.

Regier (1992), for instance, presents a neural network-based system that learns to associate spatial terms with pairs of 2-D objects. The terms are learned independently from each other, using positive and negative examples. The system can be trained to recognize English prepositions (e.g. *above*) as well as prepositions in other languages (e.g. the Russian preposition *iz-pod*). It can also be trained to recognize prepositions that convey the idea of motion (e.g. *over*). However, only two topological relationships are actually considered: inclusion and contact. They are modelled by crisp, all-or-nothing relations. Moreover, only simple convex objects, such as triangles, circles and rectangles, are handled.

Kopp (1994) also uses a connectionist approach. In the training phase, his system learns to recognize some objects and spatial relationships. It is presented with a series of images and associated textual expressions. In the testing phase, the system is presented an image only, and is asked to generate a textual expression on its own. The images, however, are 64×64 binary images, and the objects are 5×5 predefined masks. Only directional relationships are considered (*right*, *over*, *left*, *under*). Simplistic descriptions are produced, such as ‘frog under fly’, ‘frog left of cat’, or ‘cat right of frog under fly over dog’.

The linguistic expressions output by the system in Abella and Kender (1999) are far more elaborated. The system generates path descriptions for getting from a start location to a goal location, such as: ‘First, find the Chinese Pavilion which is the leftmost one near the German Pavilion. Then, find the Norwegian Pavilion which is the leftmost one next to the Chinese Pavilion. The Mexican Pavilion is the topmost one below the Norwegian Pavilion.’ (The example is set in Disney World.) The topological relationships, however, are neglected. Only *inside* can be handled by the system. Moreover, the objects are approximated by rectangles.

The approximation is rather crude, and some authors turn to angle histograms (Krishnapuram *et al.* 1993, Miyajima and Ralescu 1994). A histogram of angles is a quantitative representation of the relative position between two objects A and B. It encapsulates structural information about the objects as well as information about their spatial relationships. The objects are not necessarily convex, nor disjoint, and they are not approximated by rectangles or other elementary entities. In Keller and Wang (2000), angle histogram values feed neural networks trained on aggregate responses from a panel of people. The outputs of the networks are spatial relationship numeric values: they tell us to what extent object A *surrounds*, is *above*, *below*, to the *right* or *left* of object B. A system of fuzzy rules then produces a final linguistic analysis, such as: ‘The roof is right of the tree.’ ‘The stack buildings are surrounded by a pipe.’ ‘A convoy of vehicles is below-right of the SAM (Surface-to-Air Missile) site.’

The system in Matsakis *et al.* (2001) only handles the four directional relationships. It does not handle *surrounds*. It has, however, more descriptive power. Object pairs are represented by force histograms (Matsakis and Wendling 1999), which generalize and supersede angle histograms. Examples of outputs are: ‘The group of storehouses is loosely above-left of the stack buildings.’ ‘The storehouse 9 is perfectly above the stack buildings, but slightly shifted to the right.’ ‘The tower is to the left of the stack buildings, but a little above.’ (The example is set in a power plant.) The system is able to indicate how satisfactory each description is, i.e. to what extent it is necessary to turn or not to other relationships.

In Skubic *et al.* (2003), the system above is expanded and adapted to robot navigation. The linguistic terms (which are stored in a dictionary) are changed accordingly, and *near*, *far* and *surrounds* are added to the set of available relationships. The system is fed with readings from sonar sensors placed around the robot. Examples of outputs are: 'An object is mostly in front of the robot, but somewhat to the left. The object is far from the robot.' 'An object is on the left of the robot, but extends forward relative to the robot. The object is very close to the robot.' 'The robot is surrounded, but there is an opening on the rear-right.'

Angle and force histograms are ill-suited to topological information extraction. To address this issue, Wang and Makedon (2003) design the R-histogram, another quantitative representation of the relative position between objects. The R-histogram-based system presented in Wang *et al.* (2004), however, can only handle two topological relationships: *inside* and *outside*. Moreover, the system outputs spatial relationship values and stops short of generating natural language descriptions.

Intelligent computerized systems for spatial data processing should support linguistic queries and outputs that include approximate terms—which people commonly employ when describing spatial relationships. In Skubic *et al.* (2003), mentioned above, approximate terms are used in conjunction with distance and directional relationships. In Zhan (2002), they are used in conjunction with topological relationships: 'Region Q covers region R *a little bit*.' 'Region Q *nearly completely* covers region R.' 'Region Q covers region R *somewhat*.' A fuzzy set model of these terms is developed on the basis of cognitive evidence obtained through experiments with human subjects. Only convex regions and the relationship *covers* are, however, considered.

In this paper, we introduce a system able to generate a linguistic description of the topological relationships between two 2-D image objects. We designed it with the following in mind: the system should handle more than just one or two relationships; it should not be limited to convex regions; it should make use of approximate terms. It was argued in Wawrzyniak *et al.* (2004) that the descriptions could be built around Allen relations (Allen 1983), based on information extracted from F-histograms (Matsakis 1998, Matsakis and Wendling 1999). The F-histogram is a generic quantitative representation of the relative position between two objects. In Matsakis and Nikitenko (2005), it was shown that F-histograms could be coupled with any set of mutually exclusive and collectively exhaustive spatial relations between segments on an oriented line. The Allen relations constitute such a set. The descriptions generated by the system presented here could be built around other relations. The set of Allen relations, however, is a well-known set, of reasonable size, which has been extensively used. Moreover, as shown by Knauff (1999) through memory experiments on conceptual cognitive adequacy, the Allen relations seem to describe some important aspects of human conceptual knowledge about spatial relationships. The information they embed is similar to the information people use when representing and remembering spatial arrangements. The paper is organized as follows. In Section 2, we review the notion of Allen F-histograms. Section 3 deals with how the linguistic descriptions are generated from these histograms. Experimental results are in Section 4. Conclusions and directions of future work appear in Section 5.

2 Allen F-histograms

Consider two 2-D objects, A and B. An F-*histogram* F^{AB} is a numeric function used to represent the position of A (the *argument*) relative to B (the *referent*). The principle is illustrated by figure 1. For any direction θ, the value $F^{AB}(\theta)$ is computed by decomposing A and B into pairs (I_k,J_k) of aligned *longitudinal sections*. $F^{AB}(\theta)$ depends on how the values $F(I_k,J_k,\theta)$ shown in figure 1*(b)* are calculated and aggregated. For instance, I_k and J_k can be seen as metal rods of negligible diameter and $F(I_k,J_k,\theta)$ as the scalar resultant of gravitational forces: the forces exerted by the particles of I_k on those of J_k and that tend to move J_k in direction θ. The F-histogram F^{AB} is then called a *force histogram*. Force histograms are specialized in the modelling of directional relationships (e.g. see Matsakis 2002). *Allen* F-*histograms* constitute a different family of F-histograms. They are F-histograms coupled with Allen relations (Matsakis and Nikitenko 2005). Originally introduced to represent knowledge about time intervals, the thirteen Allen relations (Allen 1983) can be seen as mutually exclusive and collectively exhaustive spatial relations between segments on an oriented line. They are denoted by <, *m*, *o*, s, *fi*, *d*, =, *di*, *si*, *f*, *oi*, *mi* and > (figure 2). Let $\mathcal{A}$ be the set of Allen relations. Consider the Allen F-histogram F_r^{AB}, where $r\in\mathcal{A}$. Each pair (I_k,J_k) of longitudinal sections defines pairs (I_{ki},J_{kj}) of segments for which *r* can be readily assessed.

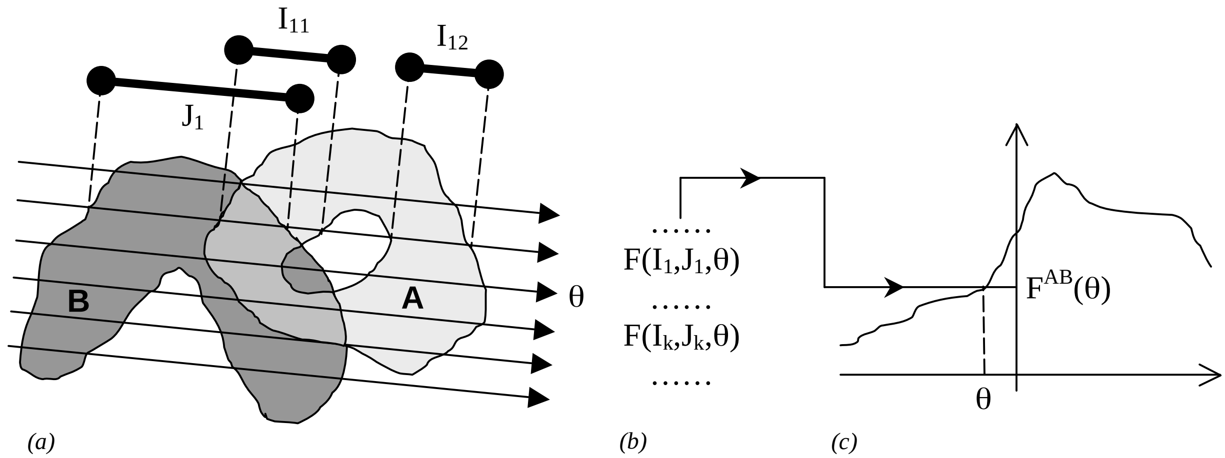


Figure 1. F-Histograms. Principle. *(a)* For each direction θ, the space is partitioned into parallel lines. As an example, the second line from the top intersects the objects A and B in $I_1=I_{11} \cup I_{12}$ and J_1, where I_{11}, I_{21} and J_1 are segments. I_1 is a *longitudinal section* of A, and J_1 is a longitudinal section of B. *(b)* A numeric value $F(I_1,J_1,\theta)$ is attached to these aligned sections I_1 and J_1. *(c)* The values $F(I_k,J_k,\theta)$ obtained for all lines in direction θ are combined into a single value $F^{AB}(\theta)$.

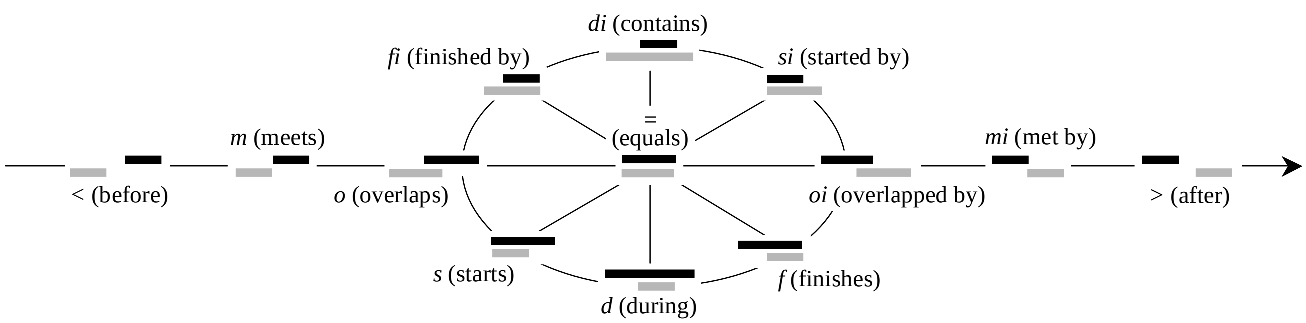


Figure 2. Allen relations. The black segment is the *referent*. The grey segment is the *argument*. As an example, the relations *oi* and *f* are linked because one can be obtained directly from the other by moving and resizing the referent or the argument in a continuous way: *oi* and *f* are *conceptual neighbours* (Freksa 1992).

$F_r(I_k,J_k,\theta)$ is obtained by aggregating the $r(I_{ki},J_{kj})$ values appropriately for all i and j. Finally, $F_r^{AB}(\theta)$ is a weighted sum of the $F_r(I_k,J_k,\theta)$ values for all k. It is such that $\sum_{r \in \mathcal{A}} F_r^{AB}(\theta)$ represents an area: this area, as illustrated by figure 3*(a)*, can be used to measure the extent of object interaction in direction θ.

The computation of $F_r(I_k,J_k,\theta)$ is not that simple, however, and we need to expand a bit. The Allen relations and longitudinal sections are actually fuzzified, to guarantee that the Allen F-histograms change continuously with respect to the object configuration. If an object is slightly shifted in space, for example, or rotated, or resized, or stretched, the histograms change equally slightly. As many authors early emphasized (Dutta 1991, Freeman

1975, Robinson 1988, Wang *et al.* 1990), fuzzy set theoretic approaches—e.g. see Klir and Yuan (1995)—are of great interest for spatial modelling and reasoning because of their capability to deal with imprecision and achieve robustness. Here, any r in the set $\mathcal{A}$ of Allen relations is seen as a continuous function onto [0,1] (i.e., r is fuzzified). See figure 3*(b)* for an illustrative example. Two properties are worth mentioning. Consider any two segments I_{ki} and J_{kj}. The (fuzzy) relations are collectively exhaustive: $\sum_{r\in\mathcal{A}} r(I_{ki},J_{kj}) = 1$. Non-direct neighbours in the graph of figure 2 are mutually exclusive: for any r_1 and r_2 in $\mathcal{A}$, if r_1 and r_2 are not direct neighbours then $r_1(I_{ki},J_{kj})=0$ or $r_2(I_{ki},J_{kj})=0$. Moreover, when any two segments I_{ki} and I_{kj} of any longitudinal section I_k get closer and closer to each other, it is considered that the points in between belong more and more to I_k (i.e., I_k is fuzzified). When close enough, the two segments will basically be seen as a single segment. For an illustrative example, see figure 3*(c)*. In the end, computing $F_r(I_k,J_k,\theta)$ actually consists in processing the α-cuts ${}^{\alpha}I_k$ and ${}^{\alpha}J_k$ of the (fuzzy) sections I_k and J_k and aggregating the $r({}^{\alpha}I_{ki}, {}^{\alpha}J_{kj})$ values appropriately for all α, i, and j.

Examples of object pairs and associated Allen F-histograms are shown in figure 4. The argument object A is in light grey, the referent B in dark grey, and areas of intersection in mid-grey. For each pair (A,B), the thirteen Allen F-histograms are plotted in the same diagram and arranged in layers. This is well illustrated in figure 4*(a)*. The thirteen $F_r^{AB}(\theta)$ values on the Y-axis (most of them are actually 0) describe the spatial relationships along direction θ. Their sum $\Sigma_{r\in\mathcal{A}} F_r^{AB}(\theta)$ is the total height of the layers, and measures the object interaction in direction θ, as mentioned earlier. Figures 4*(b)(c)(d)(e)* demonstrate the effects of fuzzifying Allen

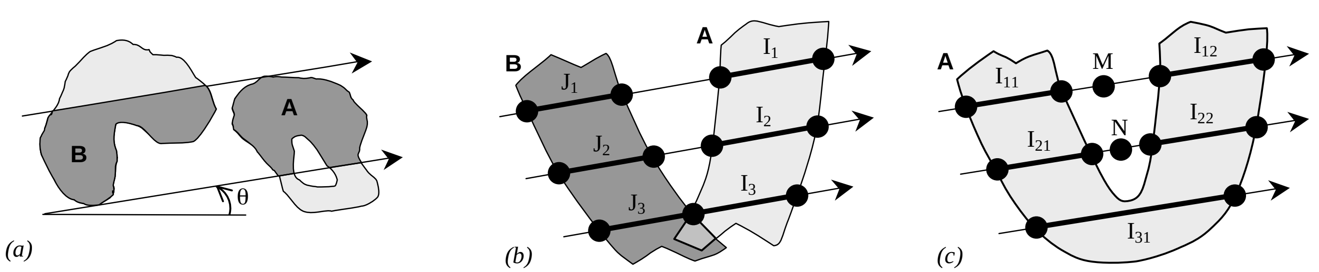


Figure 3. On F-histogram computation. *(a)* The sum $\Sigma_{r\in\mathcal{A}} F_r^{AB}(\theta)$ is the total area of the dark grey regions, which are facing each other in direction θ. *(b)* Before fuzzification of the Allen relations, $>(I_1,J_1)$ is 1 and $mi(I_1,J_1)$ is 0, $>(I_2,J_2)$ is 1 and $mi(I_2,J_2)$ is 0, $>(I_3,J_3)$ is 0 and $mi(I_3,J_3)$ is 1. After fuzzification, however, $>(I_2,J_2)$ and $mi(I_2,J_2)$ are both greater than 0 and less than 1. *(c)* Before fuzzification of the longitudinal sections, the points M and N do not belong at all to the sections $I_1=I_{11}\cup I_{12}$ and $I_2= I_{21}\cup I_{22}$. After fuzzification, the point N belongs to I_2 to some extent.

relations and longitudinal sections. In figure 4*(b)*, no fuzzification is in effect. For any direction in the plane, the only relations present are < and >. In figure 4*(c)*, *m* and *mi* appear as a result of fuzzifying Allen relations, and *di* appears as a result of fuzzifying the longitudinal sections of the argument. In figure 4*(d)*, the relations < and > disappear as B becomes adjacent to A, and the only fuzzification in effect is that of the longitudinal sections of the argument. Finally, in figure 4*(e)*, the hole in the referent disappears completely, and with it disappear all traces of fuzziness. The only relation remaining is *di*. The reader will find in Matsakis and Nikitenko (2005) all details concerning the computation of Allen F-histograms.

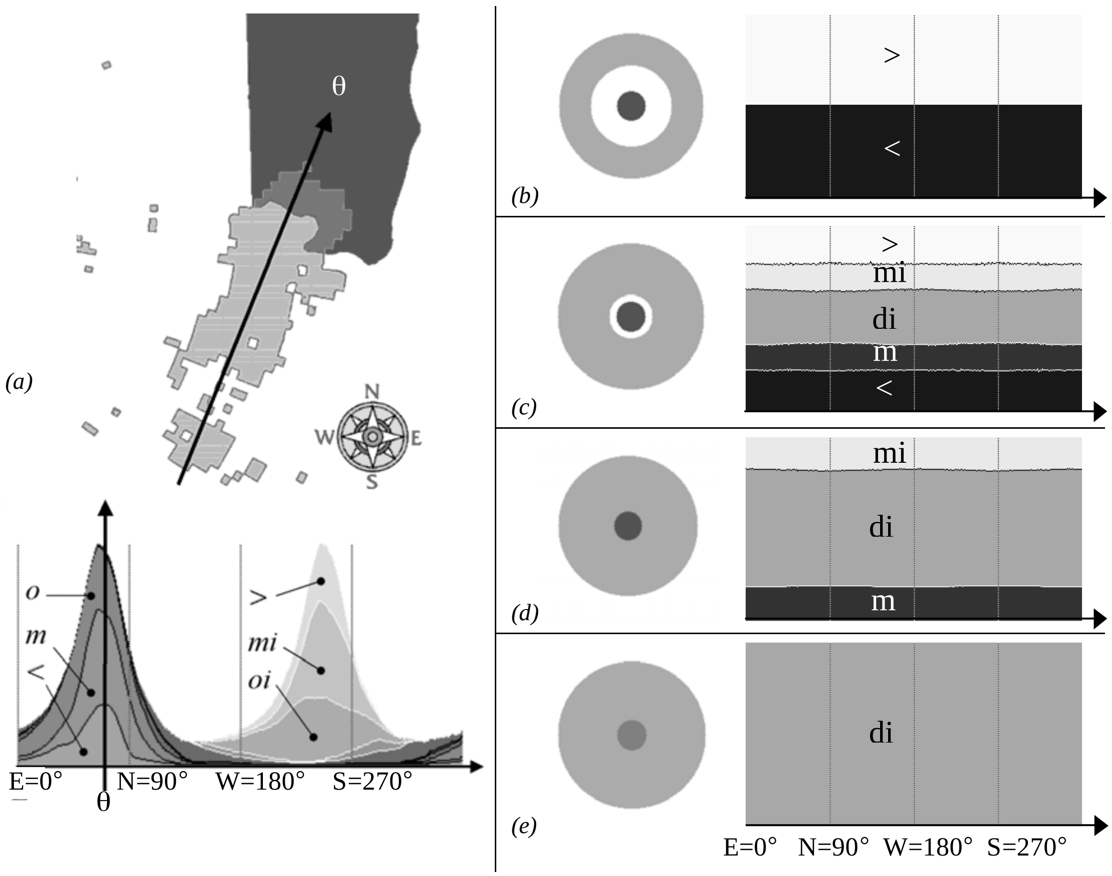


Figure 4. Examples of object pairs and associated Allen F-histograms. *(a)* The thirteen histograms are plotted in the same diagram and arranged in layers. *(b)(c)(d)(e)* Effects of fuzzifying Allen relations and longitudinal sections.

3 Generating linguistic descriptions

Our goal is to capture the essence of the topological relationships between two complex 2-D objects A and B with a natural language description $\mathcal{L}^{AB}$ that relies on the Allen relations. The description usually consists of three components. The *topological component* is built around the Allen relations that best represent the spatial relationships between the objects. The *self-assessment component* indicates to what extent the topological component succeeds in capturing the essence of the relationships. The *directional component* gives the directions along which the Allen relations mentioned in the topological component are most prominent. A typical description might read as follows:

$\mathcal{L}^{AB}$	A mostly *meets* but somewhat *overlaps* B.	*Topological Component*
	The description is *rather satisfactory*.	*Self-Assessment Component*
	A meets B to the *northeast*. A overlaps B loosely to the *east*.	*Directional Component*

The topological component cannot be built around any combination of Allen relations. This is explained in Section 3.1. Section 3.2 shows how some numeric values are extracted from the Allen F-histograms associated with the object pair (A,B). In particular, satisfactory indices are attached to the allowed combinations of Allen relations. Finally, Section 3.3 addresses the process of deriving the natural language description $\mathcal{L}^{AB}$ from the extracted values. Note that first definitions of the concepts presented in Sections 3.1.2 and 3.2.1 were given in Wawrzyniak *et al.* (2004). These concepts are here revisited and refined.

3.1 Allowed combinations of Allen relations

As shown in Section 3.1.1, the linguistic description $\mathcal{L}^{AB}$ actually relies on only eight Allen relations. Moreover, its topological component cannot be built around any combination of these eight relations. This is explained in Section 3.1.2.

3.1.1 Reorientation. Consider two segments I and J of an oriented line of direction θ. If I is *after* J in direction θ, then I is *before* J in the opposite direction θ+π. We say that < is the reorientation of > (Schlieder 1995). The two Allen relations embed different ordinal information, but they embed the same topological information (both represent *disjoint*). *m* and *mi*, *o* and *oi*, *s* and *f*, *fi* and *si* are other reorientation pairs. *d* is the reorientation of itself (if I *contains* J in direction θ, then I still *contains* J in the opposite direction θ+π). The same applies to = and *di*. For the purpose of linguistic description generation, the focus is given to topological information (in the topological component of $\mathcal{L}^{AB}$). Only eight Allen relations (say, <, *m*, *o*, *s*, *fi*, *d*, = and *di*) need, therefore, to be considered. Ordinal information is not ignored, but it is assigned less importance and handled separately (in the directional component of $\mathcal{L}^{AB}$). This is consistent with Knauff's results (1999). His experiments on the conceptual cognitive adequacy of Allen relations support the idea that topological information and ordinal information are represented separately in human memory, and that the former is more important than the latter (it is less consciously encoded and, therefore, more easily remembered). Reorientation reflects in the Allen F-histograms through the following equalities (some of which are well illustrated in figure 4*(a)*). For any θ: $F_{>}^{AB}(\theta) = F_{<}^{AB}(\theta+\pi)$, $F_{mi}^{AB}(\theta) = F_{m}^{AB}(\theta+\pi)$, $F_{oi}^{AB}(\theta) = F_{o}^{AB}(\theta+\pi)$, $F_{f}^{AB}(\theta) = F_{s}^{AB}(\theta+\pi)$, $F_{si}^{AB}(\theta) = F_{fi}^{AB}(\theta+\pi)$ and $F_{d}^{AB}(\theta) = F_{d}^{AB}(\theta+\pi)$, $F_{=}^{AB}(\theta) = F_{=}^{AB}(\theta+\pi)$, $F_{di}^{AB}(\theta) = F_{di}^{AB}(\theta+\pi)$. In other words, only eight Allen F-histograms are independent of each other. The other five need not be computed. Moreover, the F_d, $F_=$ and F_{di}-histograms are periodic with period π.

3.1.2 Coherent sets. Obviously, we cannot expect the linguistic expression $\mathcal{L}^{AB}$, which relies on the sole Allen relations, to perfectly describe the topological relationships between A and B. These objects are arbitrarily complex 2-D objects, not aligned segments, and even a human observer might have difficulty describing the configuration. At the very least, we should allow $\mathcal{L}^{AB}$ to be built around more than one Allen relation. As shown in figure 4*(a)*, different directions in space may yield different relations. Furthermore, different relations might coexist along the same direction. Here, however, we must be careful. Consider the following descriptions: 'A mostly *meets* but somewhat *overlaps* B', 'A *contains* and is *before* B', 'A mostly *overlaps*, is somewhat *before*, partially *meets*, and is marginally *equal* to B'. The first one sounds *coherent*, since it is easy to picture such objects A and B. The second one involves semantically contradictory relations and goes against intuition. As for the last expression, even if it is true to the actual relationships between the objects, it feels somewhat overwhelming. When one starts reading the sentence, one begins to build a mental picture of the objects being described. By the time one is finished reading the sentence, however, one no longer has a clear picture of what the relationships between the objects actually are. We would like to strike a balance between simplification and detail. Our description should relate the most meaningful information while leaving out superfluous and potentially overwhelming detail. Let us formalize this.

As explained in Section 3.1.1, the Allen relations that best represent the spatial relationships between the objects are selected from $\mathcal{A}' = \{<, m, o, s, fi, d, =, di\}$. A set c of Allen relations is *coherent* if and only if $c \in C$, where C is some subset of the power set of $\mathcal{A}'$. The relations within c are then considered not to semantically contradict each other and might be used together in a linguistic description without overloading it with information. For example, assume:

$$C = \cup_{r \in \mathcal{A}'} \{\{r\}\} \quad (1)$$

The only coherent sets are the eight singletons $\{<\}$, $\{m\}$, $\{o\}$, etc. In other words, it is considered that a coherent description cannot involve more than one Allen relation. The expression 'A *starts* B' is a possible output of the system, while 'A mostly *meets* but somewhat *overlaps* B' is not. As stated previously, such a choice might prove too restrictive for adequately describing the relationships between complex 2-D objects. Here is another option:

$$C = \cup_{r \in \mathcal{A}'} \{\{r\}\} \cup \{\{r, r'\} \subset \mathcal{A}' \mid \delta_{r,r'} = 1\}. \quad (2)$$

$\delta_{r,r'}$ denotes the length of the shortest path, in the graph of figure 2, between the Allen relations r and r'. We have $\delta_{m,m}=0$, $\delta_{m,o}=1$, $\delta_{m,s}=2$, etc. The set C now contains all singletons and all pairs of neighbour relations $\{<,m\}$, $\{m,o\}$, $\{o,s\}$, etc. 'A *starts* B' and 'A mostly *meets* but somewhat *overlaps* B' are possible outputs of the system, whereas 'A *contains* and is *before* B' is not. Note that through Knauff's experiments (1999), Freksa's conceptual neighbourhood theory did not appear to be cognitively relevant. At the same time, these experiments do not actually provide arguments for or against Eq. 2. Further cognitive investigation will be needed to determine the most appropriate set of coherent sets. In the remainder of the paper, C is defined as in Eq. 2. This choice prevents the generated linguistic descriptions from being overloaded with information. It is also compatible with the fact that, usually, people do not combine more than two spatial prepositions when translating visual information into natural language descriptions (Retz-Schmidt 1988).

3.2 Numeric values extracted from the Allen F-histograms

Which Allen relations best represent the spatial relationships between the objects A and B? A winning coherent set c_0 must be selected from C (Section 3.1.2). In Section 3.2.1, we strive to quantify how well a given coherent set c represents the relationships between the objects in a given direction θ. We define *local satisfactory indices* $\sigma_c(\theta)$ and, from then, in Section 3.2.2, *global satisfactory indices* s_c. The latter are used to find the 'best' coherent set c_0 and answer the question above. In Section 3.2.3, we turn our attention to directional information. In particular, for each relation r in c_0, we try to determine the direction θ_r where r is most prominent. The linguistic description $\mathcal{L}^{AB}$ will be derived from all these numeric values, $\sigma_c(\theta)$, s_c, θ_r, etc. which are extracted from the Allen F-histograms.

3.2.1 Local satisfactory indices. In this section, the focus is on a specific direction θ (hence the adjective 'local'). Assume the topological relationships between the objects along θ are represented by the Allen relations in some coherent set c. The purpose of a *local satisfactory index*, $\sigma_c(\theta)$, is to indicate how satisfactory this representation is. First, consider a relation r, its reorientation $\overline{r}$ (Section 3.1.1), and the quantity $v_r(\theta)$ defined as follows:

$$\begin{cases} r = \overline{r} \;\Rightarrow\; v_r(\theta) = F_r^{AB}(\theta) \,/\, \Sigma_{r'\in\mathcal{A}}\, F_{r'}^{AB}(\theta) \\ r \neq \overline{r} \;\Rightarrow\; v_r(\theta) = [\,F_r^{AB}(\theta)+F_{\overline{r}}^{AB}(\theta)\,] \,/\, \Sigma_{r'\in\mathcal{A}}\, F_{r'}^{AB}(\theta) \end{cases} \qquad (3)$$

Remember that r and $\overline{r}$ represent the same topological relation (Section 3.1.1). For instance, *before* (<) and *after* (>) represent *disjoint*; *meets* (*m*) and *met by* (*mi*) represent *adjacent*; *contains* (*di*) is its own reorientation and represents itself. The values $v_<(\theta)$, $v_m(\theta)$ and $v_{di}(\theta)$ can therefore be interpreted as the degrees of truth of the propositions 'A and B are *disjoint* in direction θ', 'A and B are *adjacent* in direction θ' and 'A *contains* B in direction θ', respectively. We now can define local satisfactory indices in a very simple way:

$$\sigma_c(\theta) = \Sigma_{r\in c}\, v_r(\theta). \tag{4}$$

$\sigma_c(\theta)$ is a number between 0 and 1. The value 1 is reached when the only Allen relations present along θ are the relations in c (and their reorientations). These are desirable properties. Assume, however, c is {<} and $v_<(\theta)=0.7$. It seems reasonable that $\sigma_{\{<\}}(\theta)$ should be higher if the relation *before* coexists with *meets* ($v_m(\theta)=0.3$) than if it coexists with *contains* ($v_{di}(\theta)=0.3$), which lies further from *meets* on the conceptual neighbourhood graph (figure 2). Equation 4 does not make this distinction. In both cases, $\sigma_{\{<\}}(\theta)$ is assigned the value 0.7. In the remainder of the paper, we adopt the following definition:

$$\sigma_c(\theta) = \max\,\{0, \Sigma_{r\in c}\, v_r(\theta) - \Sigma_{r'\in \mathcal{A}'-c}\, [\delta_{c,r'}(\theta)/4]\, v_{r'}(\theta)\} \tag{5}$$

$$\text{where} \quad \delta_{c,r'}(\theta) = [\Sigma_{r\in c}\, v_r(\theta)\, \delta_{r,r'}] \,/\, \Sigma_{r\in c}\, v_r(\theta).$$

$\delta_{r,r'}$ is the distance between r and r', as defined in Section 3.1.2, while $\delta_{c,r'}(\theta)$ is a weighted average distance between r' and the relations in c. In Eq. 5, the number 4 corresponds to the maximum possible value for $\delta_{r,r'}$ and $\delta_{c,r'}(\theta)$. The desirable properties mentioned above still hold. Moreover, any $r'\in\mathcal{A}'-c$ such that $v_{r'}(\theta)>0$ leads to a decrease of $\sigma_c(\theta)$, which becomes more pronounced as the distance between r' and c increases. $\sigma_c(\theta)\neq 0$ is however guaranteed if $\Sigma_{r\in c}\, v_r(\theta) > 0.5$. Finally, $\sigma_c(\theta)$ is continuous with respect to all the $v_r(\theta)$ values, and a continuous transition between coherent sets is achieved: $\sigma_{c'}(\theta) = \sigma_c(\theta)$ if c' and c are two coherent sets such that $c' = c \cup \{r'\}$ and $v_{r'}(\theta) = 0$.

3.2.2 Global satisfactory indices. The purpose of a *global satisfactory index*, s_c, is to measure the degree to which the Allen relations in some coherent set c satisfactorily represent the spatial relationships between the 2-D objects at hand. Let c_0 be the coherent set that 'best' represents these spatial relationships. The idea is that c_0 should give the highest index:

$$s_{c_0} = \max_{c \in C} s_c \,. \tag{6}$$

The simplest way to define s_c is to rely on local satisfactory indices (Section 3.2.1), as follows:

$$s_c = \max_\theta \sigma_c(\theta) \,. \tag{7}$$

Here, max can be interpreted as a fuzzy existential quantifier and s_c as the degree of truth of the proposition '*there exists* at least one direction along which the relations in c satisfactorily represent the topological relationships between the objects'. Assume, however, the relations in c are the only ones present along some direction θ_0 (i.e. $\sigma_c(\theta_0) = 1$). It seems reasonable that the global satisfactory index should be higher if the objects heavily interact along θ_0 than if they barely interact along θ_0. In the latter case, a linguistic description built around the relations in c would be of little value. Equation 7 does not make this distinction. In both cases: $s_c = 1$. Hence this second definition,

$$s_c = \max_\theta \min\{\sigma_c(\theta), i(\theta)\}, \tag{8}$$

where $i(\theta)$ denotes the normalized object interaction in direction θ:

$$i(\theta) = [\Sigma_{r \in \mathcal{A}} F_r^{AB}(\theta)] \,/\, \max_\alpha \Sigma_{r \in \mathcal{A}} F_r^{AB}(\alpha). \tag{9}$$

Contrary to $\Sigma_{r \in \mathcal{A}} F_r^{AB}(\theta)$, which represents an area (Section 2), $i(\theta)$ belongs to the interval [0,1]. A value of 1 indicates that the object interaction is maximal along θ. A value of 0 indicates that the objects are not involved in any Allen relations along θ. In Eq. 8, min can be interpreted as a fuzzy conjunction and s_c as the degree of truth of '*there exists* at least one direction along which: the objects heavily interact *and* the relations in c satisfactorily represent the topological relationships between the objects'. Other fuzzy conjunctions than min could be used. In particular, one might find the algebraic product a better choice:

$$s_c = \max_\theta \sigma_c(\theta)\, i(\theta). \quad (10)$$

$\Sigma_{r\in\mathcal{A}} F_r^{AB}(\theta)$ is a factor of $i(\theta)$ (Eq. 9) while $1/\Sigma_{r\in\mathcal{A}} F_r^{AB}(\theta)$ is a factor of $\sigma_c(\theta)$ (Eqs. 5 and 3). The sum $\Sigma_{r\in\mathcal{A}} F_r^{AB}(\theta)$ therefore cancels out of the product $\sigma_c(\theta)\, i(\theta)$. It could be objected that max and the algebraic product are not dual operators; the De Morgan laws are not satisfied (e.g. see Klir and Yuan 1995). However, there is no reason here to force duality. As pointed out by Trillas (2005), duality is not always desirable, especially when linguistic soundness comes into play. In Eq. 10 as in Eqs. 7 and 8, the focus, in the end, is given to a single direction. According to all three equations, $s_c=1$ if there exists some θ_0 such that $\sigma_c(\theta_0)=i(\theta_0)=1$ (i.e. such that the relations in c are the only ones present along θ_0 and the object interaction is maximal along θ_0). Is this property desirable? Assume, for instance, that θ_0 is 0, and that for any direction θ between 10 and 350 degrees the relations in c are not present at all while the relations in another coherent set c' largely dominate. A linguistic description built around the relations in c might not be very satisfactory. A way to get around this is to define s_c as follows:

$$s_c = \left[\int_{-\pi}^{+\pi} \sigma_c(\theta)\, i(\theta)\, d\theta\right] / \left[\int_{-\pi}^{+\pi} i(\theta)\, d\theta\right]. \quad (11)$$

Equation 7 makes it particularly easy for s_c to be 1. Equation 11 makes it particularly difficult: $s_c\neq 1$ if there exists some direction θ such that $\sigma_c(\theta)\neq 1$ and $i(\theta)\neq 0$ (i.e. such that the relations in c are not the only ones present along θ). In the remainder of the paper, s_c is defined as in Eq. 10, which constitutes a compromise. Note that in Eq. 11 the index s_c is defined as a weighted average of the $\sigma_c(\theta)$ values, while in Eq. 7 it is defined using max, the upper-bound of all averaging operations. In future work, other averaging operations will be investigated.

3.2.3 Directional information. The Allen relations that will appear in the linguistic description $\mathcal{L}^{AB}$ are the elements of the winning coherent set c_0 (Section 3.2.2). Assume c_0 is $\{m,o\}$ and the topological component of $\mathcal{L}^{AB}$ is 'A mostly meets but somewhat overlaps B'. Where does A meet B? Where does it overlap B? Let *r* be an element of c_0 and let $a_r(\theta)$ be the degree of truth of the proposition 'direction θ is the direction where A *r* B is most true'. The direction θ_r where *r* is most prominent satisfies:

$$a_r(\theta_r) = \max_\theta a_r(\theta). \tag{12}$$

The simplest way to define $a_r(\theta)$ is as follows:

$$a_r(\theta) = F_r^{AB}(\theta) \,/\, \max_\alpha F_r^{AB}(\alpha). \tag{13}$$

This equation, however, displays a complete lack of consideration for the directions where r is present to a lesser or equal degree. In Matsakis *et al.* (2001), the degree of truth of the proposition ‘A is in direction θ of B’ was extracted from the force histogram associated with (A,B). The authors proceeded by imposing physical considerations on the histogram and distinguishing between effective, compensatory and contradictory forces. The aim was to build, around directional relations, an automated linguistic description of the relative position between two objects. Here, $a_r(\theta)$ is extracted from F_r^{AB} using exactly the same technique. Assume A meets B precisely to the west (i.e. the eastern part of A meets the western part of B; the Allen relation m holds when looking in direction θ=0, from west to east). Then, when computed as in Matsakis *et al.* (2001), the value $a_m(0)$ is high (and Eq. 12 gives θ_m=0). Now, assume A meets B not only to the west, but to the northwest and the southwest as well. Then, $a_m(0)$ is low (but is still the highest $a_m(\theta)$ and, again, θ_m=0). If $a_r(\theta)$ was defined by Eq. 13, the degree of truth $a_m(0)$ would be high in both cases, and no distinction could be made.

3.3 From numeric values to words

So far, we have discussed the extraction of numeric values from Allen F-histograms. In this section, we turn our attention to converting these values into words. The Allen relations that will appear in the linguistic description $\mathcal{L}^{AB}$ are the elements of the winning coherent set c_0. For instance, if $c_0 = \{m,o\}$, the topological component of $\mathcal{L}^{AB}$ might be ‘A *mostly* meets but *somewhat* overlaps B’, or ‘A meets but *marginally* overlaps B’. The most appropriate adverbs and other approximate terms will be determined and modelled in the framework of fuzzy set theory. Fuzzy set theory is an obvious choice because of its links to natural language. Moreover, Allen F-histogram computation is based on the fuzzification of Allen relations and longitudinal sections. Finally, there is cognitive evidence regarding the fuzzy nature of approximate linguistic terms in the description of topological relationships (Zhan 2002). In Zhan (2002), these terms are modelled by fuzzy sets determined through human subject experiments. At this stage of our work, we have chosen to trust our own intuition and empirically model them using standard trapezoidal fuzzy sets. Each trapezoid is symmetrical (unless truncated), and its major basis is twice the length of its minor basis. Trapezoidal fuzzy sets have been widely used in the literature because of their computational simplicity and efficiency. Moreover, most practitioners have found that they are adequate for developing approximate solutions (Yen and Langari 1999). The subsections below are in the spirit of Zadeh's idea (1996, 1999) of ‘computing with words’—a methodology which involves a fusion of natural languages and computation with fuzzy variables. Note, however, that Zadeh’s focus is on reasoning with perceptions that are expressed in words; it is not on translating (visual) perceptions into words, as in the present paper.

3.3.1 Topological component. The winning coherent set c_0 contains at most two elements (Section 3.1.2, Eq. 2) and satisfies $s_{c_0} = \max_{c \in C} s_c$, where $s_c = \max_\theta \sigma_c(\theta)\, i(\theta)$ (Section 3.2.2). Actually, the Allen relations in c_0 best represent the topological relationships between A and B along some direction θ_0, the *direction of major object interaction*. θ_0 is determined at the same time as c_0. The pair (c_0,θ_0) is defined by:

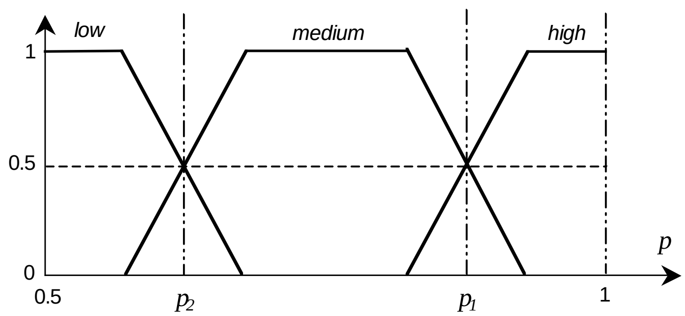
$$\sigma_{c_0}(\theta_0)\, i(\theta_0) = \max_{\theta,c} \sigma_c(\theta)\, i(\theta). \tag{14}$$

Assume (c'_0,θ'_0) and (c''_0,θ''_0) both maximize $\sigma_c(\theta)\, i(\theta)$. If c'_0 and c''_0 contain the same number of elements, the winning pair is chosen randomly. If c'_0 has fewer elements, (c'_0,θ'_0) wins against (c''_0,θ''_0) (which would lead to a less concise description). The typical case is when $c''_0 = c'_0 \cup \{r'\}$ and $\theta''_0 = \theta'_0$, with r' such that $v_{r'}(\theta'_0) = 0$. Since the transition between coherent sets is continuous (Section 3.2.1): $\sigma_{c''_0}(\theta''_0)\; i(\theta''_0) = \sigma_{c'_0}(\theta'_0)\; i(\theta'_0)$. In the end, the winning pair (c_0,θ_0) satisfies: $\forall r \in c_0,\ v_r(\theta_0) > 0$.

If $c_0 = \{r_0\}$ and $s_{c_0}=1$, then the topological component of $\mathcal{L}^{AB}$ takes the form ‘A perfectly r_0 B’. If $c_0 = \{r_0\}$ and $s_{c_0} \neq 1$, it takes the form ‘A r_0 B’. Finally, if $c_0 = \{r_0, r_1\}$, with $v_{r_0}(\theta_0) \geq v_{r_1}(\theta_0)$, then it takes the form ‘A [⟨*adverb*$_0$⟩] r_0 ⟨*connective*⟩ [⟨*adverb*$_1$⟩] r_1 B’. As indicated by the brackets [ ], adverbs may or may not be present in this expression. The idea is to reflect the predominance of r_0 over r_1 along θ_0, which can be quantified as:

$$p = \mathrm{v}_{r_0}(\theta_0) \,/\, [\mathrm{v}_{r_0}(\theta_0) + \mathrm{v}_{r_1}(\theta_0)] \in [0.5;1) \,. \tag{15}$$

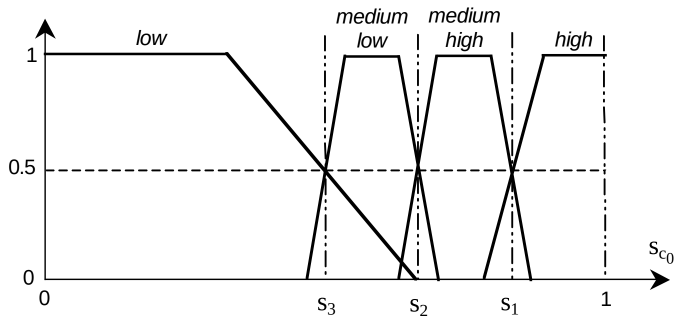
Figure 5 shows how the adverbs and connective are chosen depending on the value of p. Assume, for instance, that r_0 is < and r_1 is m. If p is *high*, 'A is before *but marginally* meets B'. If p is *medium*, 'A is *mostly* before *but somewhat* meets B'. If p is *low*, 'A is before *as much as* it meets B'.



p		
low	*medium*	*high*
—	mostly	—
as much as	but	but
—	somewhat	marginally

Figure 5. Rules for the generation of the topological component of the description.
In our experiments, p_2 =0.63 (midpoint between 0.5 and 0.75) and p_1 =0.88 (midpoint between 0.75 and 1).

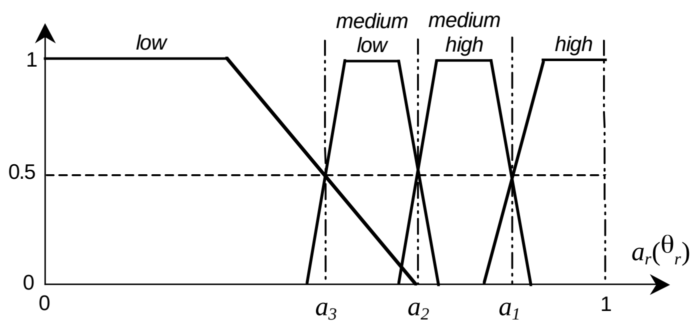
3.3.2 Self-assessment component. The topological relationships between two arbitrarily complex 2-D objects cannot always be satisfactorily described through a concise linguistic expression. The self-assessment component of $\mathcal{L}^{AB}$ gives the user a measure of confidence in the topological component of the description. This self-assessment component takes the form: 'The description is [⟨*adverb*⟩] ⟨*adjective*⟩.' Figure 6 shows how the adverb and adjective are chosen depending on the value of s_{c_0}, the global satisfactory index of the winning coherent set c_0. If s_{c_0} is *medium high*, for instance, then the self-assessment component of $\mathcal{L}^{AB}$ is: 'The description is *rather satisfactory*.' Note that a special action is taken when s_{c_0} is *low*. The system then considers that no pertinent linguistic description relying on the sole Allen relations can be given. All three components of $\mathcal{L}^{AB}$ are discarded and replaced by the message 'Only unsatisfactory descriptions could be found.'



s_{c_0}			
*low**	*medium low*	*medium high*	*high*
— unsatisfactory	rather unsatisfactory	rather satisfactory	— satisfactory

Figure 6. Rules for the generation of the self-assessment component of the description.
In our experiments, s_3 is 0.5, s_2 is 0.67 (midpoint between s_3 and s_1) and s_1 is 0.83 (midpoint between s_2 and 1).

3.3.3 Directional component. The third component of the linguistic description gives a directional estimate for each Allen relation *r* in the set $c_0 - \{d,=,di\}$ (which might be empty). The relation = is not considered. 'Where does A *meet* B?' sounds a reasonable question. 'Where is A *equal* to B?' does not. This stems from the fact that the reorientation of = is itself (Section 3.1.1). The same applies to *d* and *di*. The estimate associated with *r* usually takes the form: 'A *r* B [$\langle adverb_r \rangle$] to the $\theta_r^{\diamond}$', as in 'A overlaps B *loosely* to the *northwest*'. Note here that 'to the northwest' means 'to the northwest of the referent B'. In other words, A overlaps the northwestern part of B, and the Allen relation *o* holds when looking in direction $\theta = -45°$ (from northwest to southeast). Figure 7 shows how the adverb is chosen depending on the value of $a_r(\theta_r)$ (Section 3.2.3). The symbol $\theta_r^{\diamond}$ refers to the compass point (N, S, E, W, NE, NW, SE, SW, NNE, NNW, SSE, SSW, ENE, WNW, ESE or WSW) that coincides (roughly) with the angle $\theta_r + \pi$. Assume, for instance, that $c_0 = \{s, d\}$ and $\theta_s = 20°$. The system then produces a single directional estimate. If $a_s(\theta_s)$ is *high*, the third component of the linguistic description is 'A starts B to the *west-southwest*'. If $a_s(\theta_s)$ is *medium high*, 'A starts B *primarily* to the *west-southwest*'. Note that a special action is taken when $a_r(\theta_r)$ is *low*, i.e. when there exists much directional ambiguity as far as the relation *r* is concerned. The expression 'A *r* B *barely* to the $\theta_r^{\diamond}$' is discarded and replaced by 'A *r* B in multiple directions'.



$a_r(\theta_r)$			
*low**	*medium low*	*medium high*	*high*
barely	loosely	primarily	—

Figure 7. Rules for the generation of the directional component of the description.
In our experiments, a_3 is 0.5, a_2 is 0.67 (midpoint between a_3 and a_1) and a_1 is 0.83 (midpoint between a_2 and 1).

4 Experiments

For our experiments, we used one training dataset and three test datasets. Each dataset is composed of a few image sequences, and each sequence of a few raster images. The images in the third test dataset correspond to real, Doppler radar data. The other images represent synthetic data. For each object pair, 360 directions were considered when computing the Allen F-histograms. In the following figures, the argument A is shown in light grey, the referent B in dark grey, and areas of intersection in mid-grey.

4.1 Training dataset

Image sequences like those in figure 8 can be used to tune the parameters defining the linguistic values. Each sequence is such that only one component of the description changes from one image to another. The first sequence, figure 8*(a)*, shows variations of the topological component (tuning of p_1 and p_2, Section 3.3.1). Figure 8*(b)* shows variations of the self-assessment component (tuning of s_1, s_2 and s_3, Section 3.3.2) and figure 8*(c)* variations of the directional component (tuning of a_1, a_2 and a_3, Section 3.3.3). The first and last images of each sequence depict limit configurations. For instance, if the argument A was slightly bigger in figure 8*(a1)*, then the corresponding description would be as in figure 8*(a2)*. If A was slightly smaller in figure 8*(a4)*, the corresponding description would be as in figure 8*(a3)*. The second and third images of each sequence, however, depict middle configurations. The description for figure 8*(a2)* would not change if A was slightly smaller or bigger. In

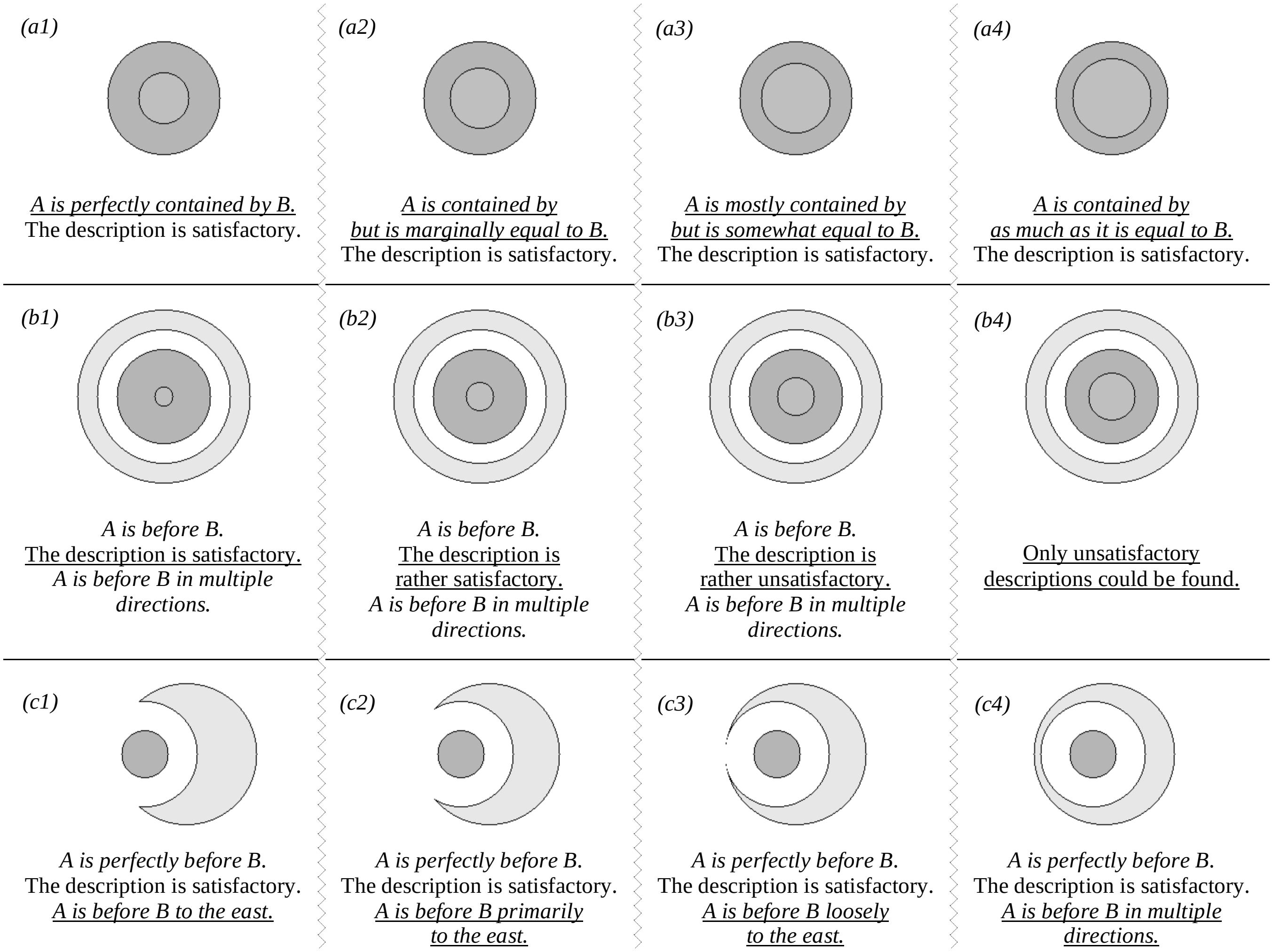


Figure 8. Training dataset. In each sequence, *(a)*, *(b)* and *(c)*, object B is a disk with constant diameter.
(a) Object A is the smaller disk. Its diameter increases. The *topological component* of the description varies accordingly.
(b) Object A is the union of the ring and the small disk, whose diameter increases. The self-assessment component varies accordingly.
(c) Object A is the crescent. Its shape changes. The *directional component* varies accordingly.

the second sequence, figure 8*(b)*, the argument A is the union $A_1 \cup A_2$ of a ring A_1 (around B) and a disk A_2 (inside B). When the diameter of A_2 increases, the system gets confused and is not quite satisfied with the topological component. The reason is that we have chosen not to consider $\{<, d\}$ a coherent set ($\{<, d\}$ does not belong to C, see Eq. 2). When A_2 gets too big, the system gives up and admits its inability to provide a useful description (figure 8*(b4)*). For our experiments, the parameters defining the linguistic values were set as in figures 5-7, by rule of thumb. The three image sequences were used for validation only. No fine-tuning was performed.

4.2 Test dataset 1

Figure 9 shows pairs of disjoint or adjacent objects. In the first image sequence, figure 9*(a)*, the argument A and referent B can be thought of as simple 2-D extensions of aligned segments. In *(a1)*, the winning coherent set c_0 is $\{m\}$, and '*A perfectly meets B*'. If a single column of pixels is introduced between A and B, as in *(a2)*, then $c_0 = \{<, m\}$, and '*A meets but is marginally before B*'. One might believe the intermediate description '*A meets B*' should be generated here, and the relation *before* should only appear when A moves farther away from B. This can be easily achieved. Remember that if $c_0=\{r_0, r_1\}$ then 'A [⟨*adverb*$_0$⟩] r_0 ⟨*connective*⟩ [⟨*adverb*$_1$⟩] r_1 B', where adverbs and connective are chosen depending on the value of p (Section 3.3.1). One might want to modify the set of linguistic values (figure 5) and add the following rule: if p is *very high* then 'A r_0 B'. In *(a3)*, '*A is before as much as it meets B*'. How far apart the two objects need to be before this description is generated depends, of course, on the way the Allen relations are fuzzified (Section 2). A similar comment applies to *(a4)*. In the second image sequence, figure 9*(b)*, the argument is of variable size and touches the referent in various ways. *(b1)* should be compared with *(a3)*. In both cases, but for totally different reasons, '*A is before as much as it meets B*'. Figure *(b3)* clearly shows the advantage of having independent directional estimates. Figure 9*(c)* also makes a good argument for independent estimates. Although the directional relationships are fairly complex, the topological relationships are very simple, and the topological component is always judged to be satisfactory. The linguistic descriptions generated by the system are, of course, invariant to translation and scaling. The last sequence, figure 9*(d)*, illustrates the fact that the topological and self-assessment components are also invariant to rotation (compare *(d1)* with *(a1)*, *(d2)* with *(a2)* and *(d3)* with *(d4)*). Moreover, when permuting the objects, all components are affected in a predictable and logical manner. If A *contains* B, for instance, then B is *contained by* A. If A is *before* B to the *south*, then B is *before* A to the *north* (compare *(d3)* with *(c3)*).

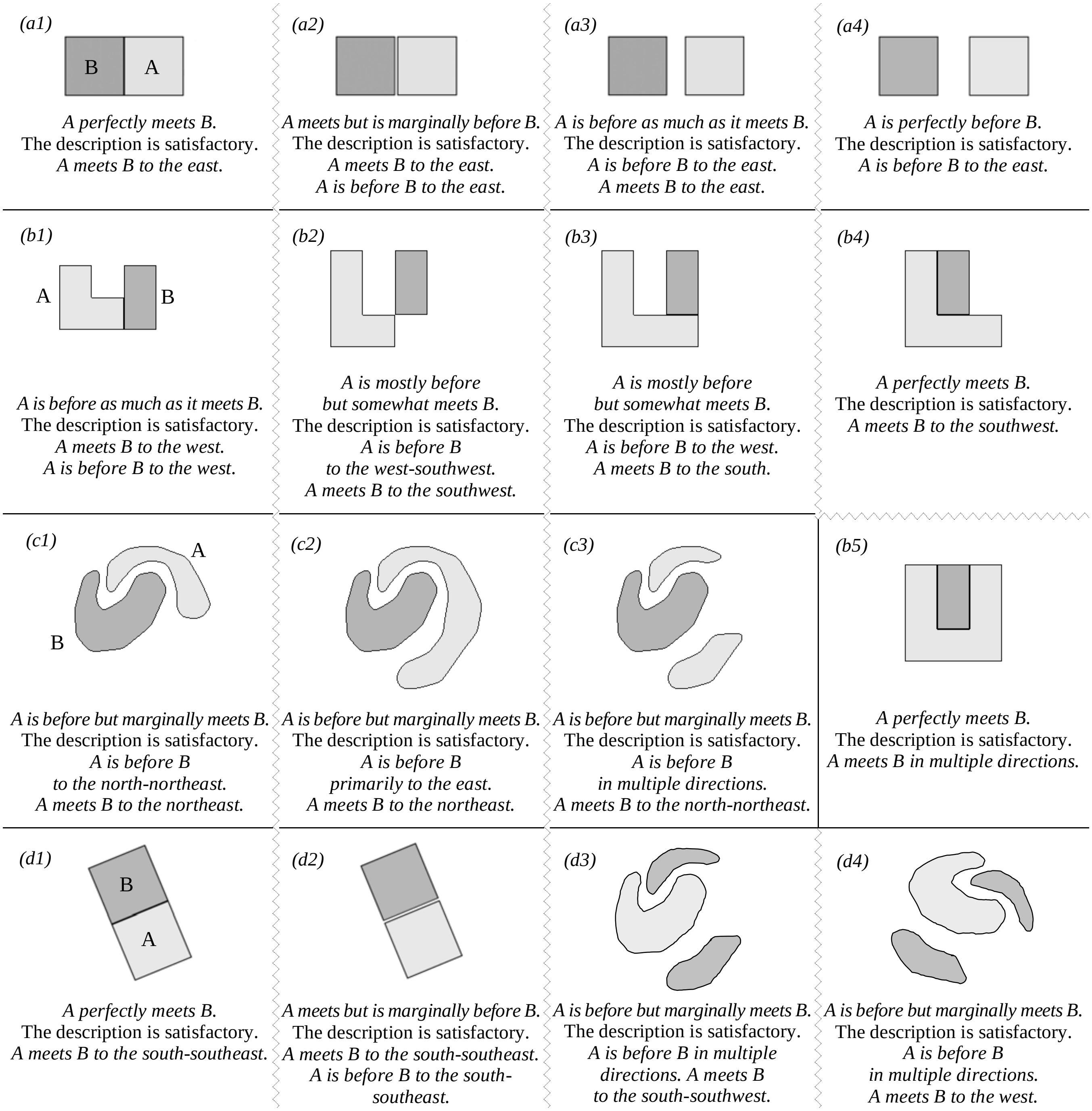


Figure 9. Test dataset 1.

4.3 Test dataset 2

The second test dataset, figure 10, focuses on intersecting objects. In the first image sequence, figure 10*(a)*, the argument moves 'into' the referent. In *(a1)*, the only Allen relation present is *before*. In *(a2)*, *overlaps* dominates. *meets* occurs due to the fuzzification of *m* and *o*. In *(a3)*, the argument mostly *starts* the referent, but its eastern part is shifted downward into the referent, hence the presence of *during*. Figure 10*(b)* is pretty much the same sequence. The presence of the hole in the referent has little impact on the descriptions. The most notable differences concern *(b4)*. In *(b4)*, the direction of major object interaction lies near $\pi/4$ (looking from the southwest). The interaction of the argument with the northern edge of the hole causes a small amount of the relations *overlaps* and *starts* to appear. *starts* is coherent with *during* and it is included in the description. *overlaps* is not. It contributes to the lowering of the global satisfactory index of the coherent set $\{s,d\}$. The self-assessment component therefore drops to 'rather satisfactory'. In the next sequence, figure 10*(c)*, the argument seems to 'envelop' the referent along its border. In *(c2)*, *(c3)* and *(c4)*, *equals* gains more and more importance. Since we have chosen not to consider $\{s,=,o\}$ a coherent set, $\{s,=\}$ competes (and wins) against $\{s,o\}$. The rivalry, however, affects the self-assessment component. A similar phenomenon can be observed in figure 10*(d)*. In *(d3)*, for instance, the description, although fairly reasonable, is judged to be 'rather unsatisfactory'. The conflict here lies in choosing between the coherent sets $\{<,m\}$ and $\{m,o\}$. The objects do not overlap much along the (horizontal) direction of major object interaction, and *meets* coexists with *before* and *overlaps* due to the fuzzification of the Allen relations. Had $\{<,m,o\}$ been allowed to enter the lists, a satisfactory description (at least from the system's point of view) could have been generated. In *(d4)*, the objects overlap more and the presence of *meets* diminishes. Since no coherent set contains both *before* and *overlaps*, the competition between these relations is fierce, and the system gives up. The last image sequence, figure 10*(e)*, also poses problems for the system. Consider *(e3)*. There is no conflict along the horizontal and vertical directions, but the interaction index (Eq. 9) is rather low. Object interaction is maximum along the diagonal directions, but conflict is high. In the end, $\theta_0 = 90°$, $c_0 = \{d\}$ and $s_{c_0} = i(\theta_0) = 0.67$. The topological component of the description com-

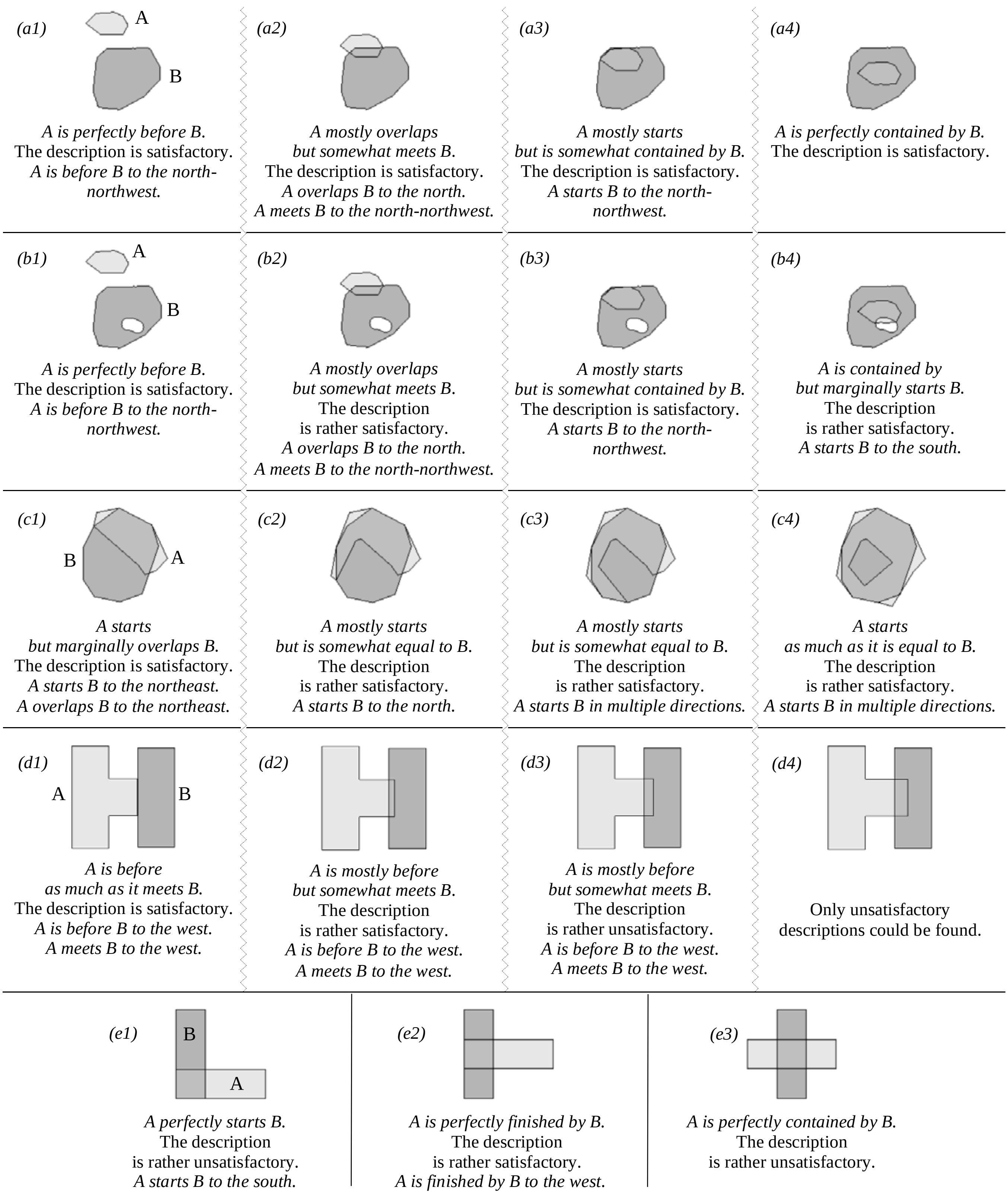


Figure 10. Test dataset 2.

pletely disregards the portions of the argument lying outside of the referent. The system realizes that, but its self-assessment is obviously not harsh enough. One might want to redefine the global satisfactory index, as hinted in Section 3.2.2, or to modify the linguistic values associated with s_{c_0} (figure 6). Various image sequences can be used to fine-tune the corresponding parameters (e.g. see figure 8*(b)*). Another alternative is to make use of ancillary information. For instance, if the degree of subsethood of A in B (often defined by $|A \cap B|/|A|$) is low, then A cannot be '*perfectly contained by*' B.

4.4 Test dataset 3

Figure 11 represents a sequence of Doppler radar images from Detroit/Pontiac National Weather Service (http://www.crh.noaa.gov/dtx/mayfly.php). The sequence, captured on June 26, 2001, shows a mayfly aerial courtship over St. Clair County, Michigan. In *(a1)*, only the relation *before* is present. The mayfly swarm (argument object) is born outside of the county (referent). In *(a2)*, the swarm becomes a disconnected object. Some of the smaller fragments *meet* the referent at the western border. In *(a3)*, the swarm has grown considerably and moved over the county. The relation *during* clearly dominates along $\theta = \pi/2$, which is the direction of major object interaction. *starts* is present to various degrees looking from the southwest, southeast, northeast… This is reflected in the directional estimate. A smaller portion of the swarm lies just outside the southwestern part of the county and is left out of the description. In *(a4)*, the swarm has contracted towards the eastern border of the county. In *(a5)*, it breaks apart into numerous fragments. *before*, *meets*, and *overlaps* are present to a roughly equal degree. Since we have chosen not to consider $\{<, m, o\}$ a coherent set, $\{<, m\}$ competes (and wins) against $\{m, o\}$. The rivalry, however, severely affects the self-assessment component.

5 Conclusion

A system for generating linguistic descriptions of the topological relationships between two-dimensional objects has been introduced. The objects need not be convex, nor connected, and they may have holes in them. The descriptions are human-like in structure and richness of language. The system makes use of approximate terms and

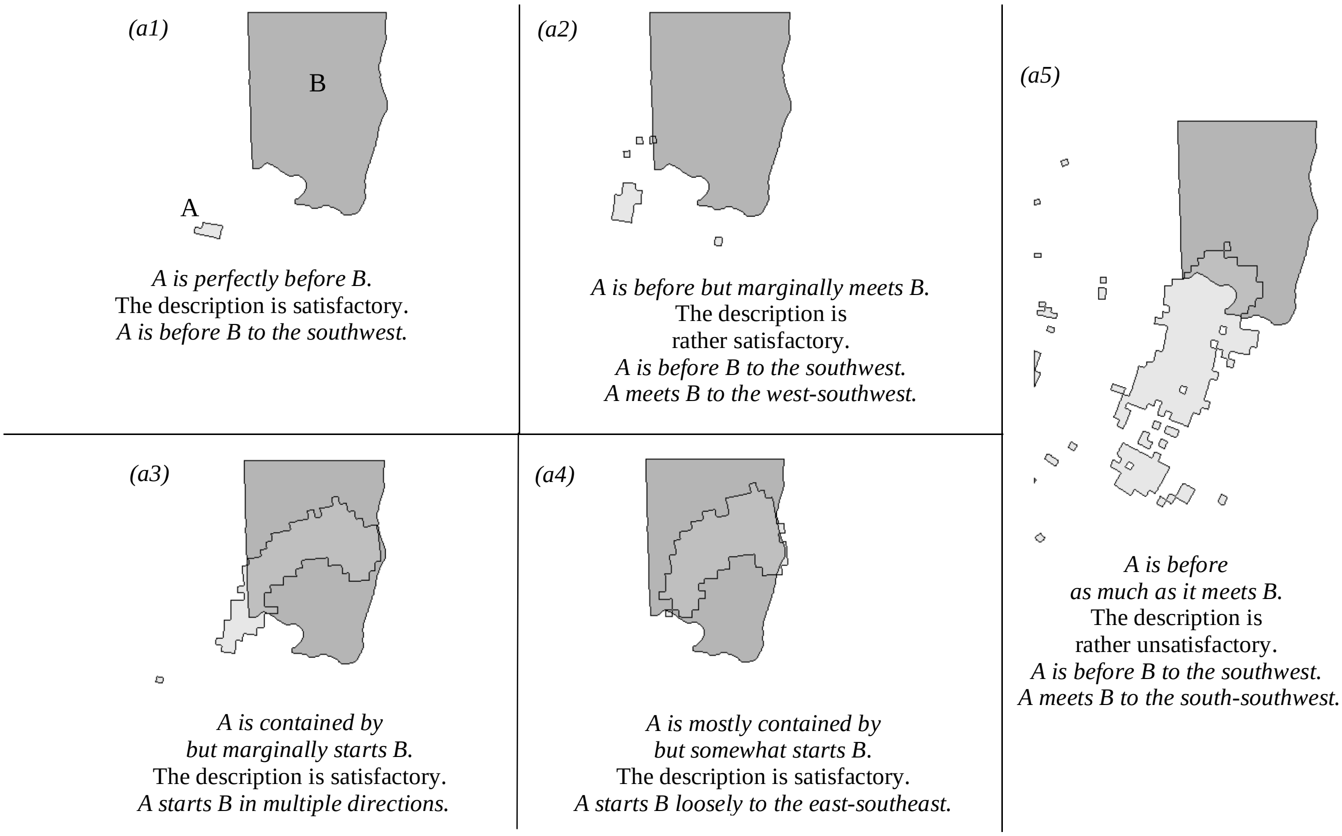


Figure 11. Test dataset 3.

operates at a fine level of granularity, which involves way more than just one or two topological relationships. To our knowledge, it is the first system to combine these characteristics. The descriptions are derived, through simple fuzzy rules, from numeric values extracted from Allen F-histograms. Each linguistic expression is built around the Allen relations that best represent the spatial relationships between the objects. A self-assessment component indicates to what extent the essence of these relationships has been successfully captured. The system also gives the directions along which the Allen relations mentioned in the description are most prominent.

As demonstrated with numerous examples, the system performs well in a large number of cases and produces intuitive descriptions. This, of course, is a rather subjective statement. Two human subjects may perceive the same spatial concept differently, and by implication, describe it differently (Mark *et al.* 1995, Robinson 2000, Worboys 2001). The assumption here is that the user is familiar with the building blocks of the system-generated descriptions, i.e. with the Allen relations. Obviously, the descriptions may not seem intuitive if the Allen relations themselves, and the way they are referred to, are not found intuitive. Issues such as the cognitive validity of computational and mathematical models of spatial relations, or the appropriateness of the names given to these relations, (Knauff 1999, Mark and Egenhofer 1994, Renz *et al.* 2000, Riedemann 2005) are not issues this paper intended to address. However, Knauff's experiments on conceptual cognitive adequacy tend to support the idea of building linguistic descriptions around Allen relations.

Considering the nature of language and spatial cognition, it is clear that subjectivity cannot be avoided. Methods for performance evaluation of systems with linguistic, subjective outputs have yet to be investigated. For example, one might want to adapt the system presented here to a prototypical perception, which would have to be captured through cognitive experiments involving multiple subjects. For training and validation purposes, the semantic similarity between prototypical and system-generated linguistic descriptions would then have to be measured. Although there are still few related publications, the need for semantic similarity measures between sentences is being increasingly recognized (Li *et al.* 2006). The procedure just described is far beyond the scope of this paper. The most important at this point is that the system can adapt to a given perception, application, context, through various mechanisms. The terms (nouns, prepositions, adverbs, adjectives…) can easily be changed, and the vocabulary can be expanded or shrunk according to need. The linguistic values in the fuzzy rules (such as *high*, *low*) can be tuned using calibration image sequences. The transition between two neighbouring relations (e.g. *before* and *meets*, *meets* and *overlaps*) depends on the way these relations are fuzzified, and can be redefined. The combinations of Allen relations allowed in a description can be chosen freely. Furthermore, a different set of relations could be considered.

These mechanisms, nonetheless, cannot prevent the system from generating some descriptions that are clearly counter-intuitive, less satisfactory than the self-assessment component would have the user believe. Several measures have been suggested to remedy the problem. For a given description, the system selects the best combination of Allen relations based on the computation of satisfactory indices—which can be defined in different ways. Ancillary information can also be exploited. These measures will be explored in future work, and should allow the system to make fewer counter-intuitive statements and provide more insightful descriptions. A first attempt in this direction was presented in Wawrzyniak *et al.* (2005). Misleading descriptions such as those in figure 10*(e)* are detected by incorporating subsethood information into the local satisfactory index. An independent module, which does not rely on Allen F-histograms, then produce alternative descriptions of the form: 'The central | southern | western… part of A coincides with the central | southern | western… part of B.' In the case of figure 10*(e3)*, for instance, the module is invoked and outputs 'The central part of A coincides with the central part of B.'

In Matsakis *et al.* (2001), the linguistic descriptions were generated exclusively from force F-histograms, and they were built around directional relations (*is to the right of*, *is above*, *is to the left of*, *is below*). Here, they are generated exclusively from Allen F-histograms and are built around topological relations (the Allen relations). These works show the specificity and limits of each type of histogram, and they show how each one can contribute to the generation of natural language expressions that capture the essence of spatial relationships. Such results are fundamental for the design of systems combining F-histograms with other sources of information. We have mentioned Wawrzyniak *et al.* (2005), which exploits subsethood ancillary information. Wawrzyniak *et al.* (2006) is an attempt in what can be seen as an opposite direction: the descriptions are built around 2-D set relations (*is disjoint from*, *overlaps*, *includes*, *is included in*, *is equal to*), and Allen F-histograms only provide ancillary information. Having machines that, like humans, can comprehend the organization of objects in space, and can reason and communicate linguistically about space, is an ambitious goal. There is no doubt in our mind that, ultimately, various sources of information will have to be combined, and various systems will have to cooperate.

Acknowledgements

The authors want to express gratitude for support from the Natural Science and Engineering Research Council of Canada (NSERC), grant 262117. They would also like to thank the anonymous reviewers for their constructive comments and suggestions.

References

ABELLA, A. and KENDER, J.R., 1999, From images to sentences via spatial relations. In *Proceedings, ICCV '99 Workshop on the Integration of Image and Speech Understanding*, 1999, Corfu, Greece.

ALLEN, J.F., 1983, Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), pp. 832-43.

BLOCH, I., 1999, Fuzzy relative position between objects in image processing: A morphological approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(7), pp. 657-64.

CLEMENTINI, E. and DI FELICE, P., 1996, A model for representing topological relationships among complex geometric features in spatial databases. *Information Sciences*, 90, pp. 121-36.

DUTTA, S., 1991, Approximate spatial reasoning: Integrating qualitative and quantitative constraints. *Int. J. of Approximate Reasoning*, 5, pp. 307-31.

EGENHOFER, M.J. and HERRING, J.R., 1994, Categorizing topological spatial relations between point, line, and area objects. In *The 9-Intersection: Formalism and its Use For Natural-Language Spatial Predicates*, M.J. Egenhofer, D.M. Mark and J.R. Herring (Eds) (Santa Barbara, CA: National Center for Geographic Information and Analysis, 1994).

FREEMAN, J., 1975, The modelling of spatial relations. *Computer Graphics and Image Processing*, 4, pp. 156-71.

FREKSA, C., 1992, Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54, pp. 199-227.

GUESGEN, H.W., 2002, Fuzzifying spatial relations. In *Applying Soft Computing in Defining Spatial Relations*, P. Matsakis and L. Sztandera (Eds), 106, pp. 1-16 (*Studies in Fuzziness and Soft Computing*, Physica-Verlag, 2002).

KELLER, J.M. and WANG, X.A., 2000, Fuzzy rule-based approach to scene description involving spatial relationships. *Computer Vision and Image Understanding*, 80(1), pp. 21-41.

KLIR, G.J. and YUAN, B., 1995, *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Prentice Hall).

KNAUFF, M., 1999, The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. *Spatial Cognition and Computation*, 1(3), pp. 261-90.

KOPP, L., 1994, A neural network for spatial relations: connecting visual scenes to linguistic descriptions. *Lund University Cognitive Studies*, 32.

KRISHNAPURAM, R., KELLER, J.M. and MA, Y., 1993, Quantitative analysis of properties and spatial relations of fuzzy image regions. *IEEE Trans. on Fuzzy Systems*, 1(3), pp. 222-33.

LI, Y., MCLEAN, D., BANDAR, Z., O'SHEA, J. and CROCKETT, K., 2006, Sentence similarity based on semantic nets and corpus statistics. *IEEE Trans. on Knowledge and Data Engineering*, 18(8), pp. 1138-50.

MARK, D.M., COMAS, D., EGENHOFER, M.J., FREUNDSCHUH, S.M., GOULD, M.D. and NUNES, J., 1995, Evaluating and refining computational models of spatial relations through cross-linguistic human-subjects testing. In *Spatial Information Theory: A Theoretical Basis for GIS*, A.U. Frank and W. Kuhn (Eds), pp. 553-68 (*Lecture Notes in Computer Sciences*, Springer-Verlag, 1995).

MARK, D.M. and EGENHOFER, M.J., 1994, Modeling spatial relations between lines and regions: Combining formal mathematical models and human subjects testing. *Cartography and Geographical Information Systems*, 3, pp. 195-212.

MATSAKIS, P., 1998, *Relations Spatiales Structurelles et Interprétation d'Images*, PhD Thesis, Institut de Recherche en Informatique de Toulouse, France.

MATSAKIS, P., KELLER, J., WENDLING, L., MARJAMAA, J. and SJAHPUTERA, O., 2001, Linguistic description of relative positions in images. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 31(4), pp. 573-88.

MATSAKIS, P. and NIKITENKO, D., 2005, Combined extraction of directional and topological relationship information from 2D concave objects. In *Fuzzy Modeling with Spatial Information for Geographic Problems*, M. Cobb, F. Petry and V. Robinson (Eds), pp. 15-40 (Springer-Verlag, 2005).

MATSAKIS, P. and WENDLING, L., 1999, A new way to represent the relative position of areal objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(7), pp. 634-43.

MIYAJIMA, K. and RALESCU, A., 1994, Spatial organization in 2-D segmented images: Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems*, 65(2/3), pp. 225-36.

NABIL, M., SHEPHERD, J. and NGU, A.H.H., 1995, 2D projection interval relationships: A symbolic representation of spatial relationships. In *Proceedings, SSD'95 Advances in Spatial Databases*, 1995, pp. 292-309.

PETRY, F., COBB, M., ALI, D., ANGRYK, R., PAPRZYCKI, M., RAHIMI, S., WEN, L. and YANG, H., 2002, Fuzzy spatial relations and mobile agent technology in geospatial information systems. In *Applying Soft Computing in Defining Spatial Relations*, P. Matsakis and L. Sztandera (Eds), 106, pp. 123-55 (*Studies in Fuzziness and Soft Computing*, Physica-Verlag, 2002).

REGIER, T.P., 1992, *The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization*, PhD Thesis, University of California at Berkeley, California.

RENZ, J., RAUH, R. and KNAUFF, M., 2000, Towards cognitive adequacy of topological spatial relations. In *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Models, and Practical Applications*, C. Freksa *et al.* (Eds), pp. 184-97 (*Lecture Notes in Computer Science* 1849, Springer, 2000).

RETZ-SCHMIDT, G., 1988, Various views on spatial prepositions. *Artificial Intelligence*, 9(2), pp. 95-105.

RIEDEMANN, C., 2005, Matching names and definitions of topological operators. In *Spatial Information Theory: Cognitive and Computational Foundations*, A.G. Cohn and D.M. Mark (Eds) (*Lecture Notes in Computer Science* 3693, Springer, 2005).

ROBINSON, V., 1988, Implications of fuzzy set theory for geographic databases. *Computers, Environment, and Urban Svstems*, 12, pp. 89-98.

ROBINSON, V., 2000, Individual and multipersonal fuzzy spatial relations acquired using human-machine interaction. *Fuzzy Sets and Systems*, 113(1), pp. 133-45.

SCHLIEDER, C., 1995, *The Construction of Preferred Mental Models in Reasoning with the Interval Relations*, Technical Report 5/95, Institut für Informatik und Gesellschaft der Universität Freiburg, Germany. Appears also in *Mental Models in Discourse Comprehension and Reasoning*, C. Habel *et al.* (Eds).

SKUBIC, M., MATSAKIS, P., CHRONIS, G. and KELLER, J., 2003, Generating multi-level linguistic spatial descriptions from range sensor readings using the histogram of forces. *Autonomous Robots*, 14(1), pp. 51-69.

STEPHANIDIS, C., 2001, User interfaces for all: New perspectives into human-computer interaction. In *User Interfaces for All: Concepts, Methods, and Tools*, C. Stephanidis (Ed), pp. 3-17 (Mahwah, NJ: Lawrence Erlbaum Associates, 2001).

TRILLAS, E., 2005, What about fuzzy logic's linguistic soundness? *Fuzzy Sets and Systems*, 156(3), pp. 334-40.

WANG, F., HALL, G.B. and SUBARYONO, 1990, Fuzzy information representation and processing in conventional GIS software: Database design and application. *Int. J. of Geographical Information Systems*, 4, pp. 261-83.

WANG, Y. and MAKEDON, F., 2003, R-histogram: Quantitative representation of spatial relations for similarity-based image retrieval. In *Proceedings, 11th ACM Int. Conf. on Multimedia*, 2003, Berkeley, California, pp. 323-26.

WANG, Y., MAKEDON, F., FORD, J., SHEN, L. and GOLDION, D., 2004, Generating fuzzy semantic metadata describing spatial relations from images using the R-histogram. In *Proceedings, Joint ACM/IEEE Conf. on Digital Libraries*, 2004, Tucson, AZ, pp. 202-11.

WAWRZYNIAK, L., MATSAKIS, P. and NIKITENKO, D., 2004, Representing topological relationships between complex regions by F-histograms. *SDH'2004 Int. Spatial Data Handling Symposium*, 2004, Leicester, UK. In *Developments in Spatial Data Handling*, P.F. Fisher (Ed), pp. 245-58 (Springer Publications, 2004).

WAWRZYNIAK, L., NIKITENKO, D. and MATSAKIS, P., 2005, Describing topological relationships in words: Refinements. In *Proceedings*, *FUZZ-IEEE 2005 Int. Conf. on Fuzzy Systems,* 2005, Reno, Nevada, pp. 743-48.

WAWRZYNIAK, L., NIKITENKO, D. and MATSAKIS, P., 2006, Speaking with spatial relations. In *Int. J. of Intelligent Systems Technologies and Applications*, Special Issue on *Intelligent Image and Video Processing and Applications: The Role of Uncertainty*, 1(3/4), pp. 280-300.

WORBOYS, M.F., 2001, Nearness relations in environmental space. *Int. J. of Geographical Information Science*, 15(7), pp. 633-51.

YEN, J. and LANGARI, R., 1999, *Fuzzy Logic Intelligence, Control, and Information* (Prentice Hall).

ZADEH, L.A., 1996, Fuzzy logic = Computing with words. *IEEE Trans. on Fuzzy Systems*, 4(2), pp. 102-11.

ZADEH, L.A., 1999, From computing with numbers to computing with words — From manipulation of measurements to manipulation of perceptions. *IEEE Trans. on Circuits and Systems, Part I*, 45(1), pp. 105-19.

ZHAN, F. B., 2002, A fuzzy set model of approximate linguistic terms in descriptions of binary topological relations between simple regions. In *Applying Soft Computing in Defining Spatial Relations*, P. Matsakis and L. Sztandera (Eds), 106, pp. 179-202 (*Studies in Fuzziness and Soft Computing*, Physica-Verlag, 2002).