Aspects of High Level Computer Vision Using Fuzzy Sets

James M. Keller
Computer Engr. & Computer Sci. Dept.
University of Missouri-Columbia
Columbia, MO 65211 USA
keller@cece.missouri.edu

Pascal Matsakis
Institut de Recherche en Informatique
Universite Paul Sabatier
31062 Toulouse Cedex FRANCE
Matsakis@irit.fr

Abstract

Fuzzy set theory is making many inroads into the handling of uncertainty in various aspects of image processing and computer vision. High level computer vision is a place that holds great potential for fuzzy sets because of its natural linguistic capabilities. Scene description, i.e., the language-based representation of regions and their relationships, for either humans or higher automated reasoning provides an excellent opportunity. In this paper we discuss aspects of scene interpretation involving linguistic descriptions of spatial relations between image objects.

1. Introduction

Natural scene understanding is an important aspect of computer vision. It has received considerable attention, but has not had the success of low level and mid level vision techniques. This is at least partially due to the fact that sophisticated world models are needed and that reasoning at high levels is rife with uncertainty. Early approaches, such as ACRONYM [1] and constraint networks[2] exposed the difficult nature of scene interpretation. These systems were primarily constructed to locate objects within a given scene, and did not explicitly model uncertainty.

Walker et al. [3] developed a system for reasoning about lines, planes, and polygons in 2 and 3 dimensions. This was extended to incorporate fuzzy set theoretic operations to control perceptual grouping of primitive elements [4].

Antony constructed a framework within which spatial, temporal and hierarchical scene reasoning can take place, although no actual imagery was analyzed [5]. The primary focus was on creating structures to efficiently carry out scene analysis tasks. He also addressed the possibility of incorporating fuzzy set concepts into constraints such as "near", and used quadtree representations to determine (crisp) areas of an image that would correspond to the spatial concepts like "northeast".

In high level computer vision, spatial relations among image objects play a vital role in the description of a scene. Humans can judge the spatial relationship between two objects, e.g., "*B* is to the right of *A*", although it has been made quite clear that human intuition varies considerably. Both the vague concepts of what spatial relationships should mean, and the uncertainty of how they can model differing human perceptions make automated calculation and use of this important information problematic.

Because of its importance and connection to human scene understanding, the concept of spatial relations has been considered in many forms, from linguistic and psychological points of view to automated definition and reasoning systems [6-9]. Spatial relations such as ABOVE, RIGHT, and others defy precise definitions, and seem to be best modeled by fuzzy sets [10-15]. However, there are many fuzzy definitions available.

With all of the potential definitions, there has been a considerable amount of argument about the "best" method. Much of the debate centered on human intuition. Each paper makes such arguments. Interestingly, what we found was that for pattern recognition, the differences between fuzzy spatial relation definitions was not crucial to the final recognition rate [16]; the fact that the approaches provided reasonable (and distinct) estimates was the important factor. Further recent evidence for this claim with respect to scene analysis can be found in [17], but this was done in a limited application environment. Much more work needs to be done to exploit the benefits of fuzzy definitions of spatial relations for scene analysis.

In this paper, we examine the utility of fuzzy spatial relations for scene description in the following sense. First, we consider a fuzzy rule-based approach to image description as found in [18]. While the results are quite good for the images tested, this approach has the problem that it is based on estimates made for only the four primitive directional relations (LEFT, ABOVE, RIGHT, BELOW) plus SURROUND. Most of the rules are used to generate additional finer quantized directions. Also, additional geometric object features are needed to develop the primitive relations. Next, we briefly describe the approach of Matsakis [13] who postulated an axiomatic framework for functions that generate spatial relationships from which he generated the histogram of forces. By picking particular choices of the functions, he can recover the histogram of angles or a histogram of "gravitational forces" that incorporates metric information about the sets in question. Finally, we consider how this richer set of geometric and angular information can be used to develop a "language" to better describe the relationships between objects. We believe that this language can be tailored to match individual human users, which has applications for content-based image retrieval.

2. Fuzzy Rule-based Scene Description

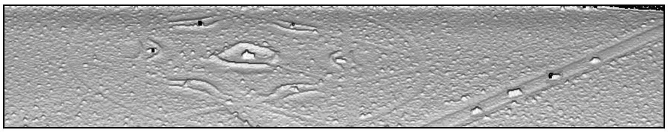
In [18], we presented a fuzzy rule-based approach for linguistic scene description involving spatial relations. Its objective was to construct linguistic descriptions of entire image scenes. Explicit modeling of the uncertainty of defining spatial relations was incorporated into the scene description methodology. We used spatial relationship values generated from neural networks trained on aggregate responses from a panel of people. These fuzzy spatial relationship values were combined with other world knowledge encoded in fuzzy logic rules to produce the final linguistic analysis. Details can be found in [18] with additional scene matching experiments in [17]

Based on outputs of spatial relationship neural networks, five spatial relation membership values (LEFT, ABOVE, RIGHT, BELOW, and SURROUND) were obtained for pairs of objects in a scene. Objects of like type were automatically grouped together to provide more concise descriptions. The fuzzy membership values were fed to a set of rules such as:

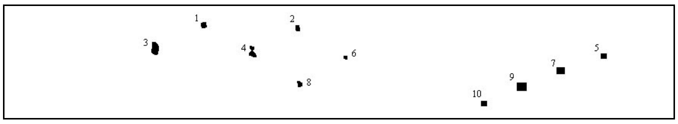
If	LEFT	is *LOW*	AND
	ABOVE	is *HIGH*	AND
	RIGHT	is *HIGH*	AND
	BELOW	is *LOW*	AND
	SURROUND	is *LOW*	
THEN	TOTALLEFT	is *LOW*	AND
	TOTALABOVE	is *LOW*	AND
	TOTALRIGHT	is *LOW*	AND
	TOTALBELOW	is *LOW*	AND
	ABOVELEFT	is *LOW*	AND
	ABOVERIGHT	is *HIGH*	AND
	BELOWLEFT	is *LOW*	AND
	BELOWRIGHT	is *LOW*	AND
	AMONG	is *LOW*	AND
	TOTALSURROUND is *LOW*		

The input values were the 5 spatial relation memberships from the spatial relationship neural networks. The 10 output variables are shown in the above rule. They indicated the confidence values of the 10 relations. There were over 200 basic rules in this system. Other rules were needed for specific problem domains.

In an application to Automatic Target Recognition, consider Figure 1(a) which shows a preprocessed LADAR range image (called "pseudo intensity") containing a SAM site and an embedded convoy of vehicles. Based on a series of detection/recognition algorithms [19], we produced the detection and labeling of the objects(Figure 1(b) and (c)).



(a)



(b)

Label	Object	Label	Object
1	Launcher	6	Launcher
2	Launcher	7	Vehicle
3	Launcher	8	Launcher
4	Center	9	Vehicle
5	Vehicle	10	Vehicle

(c)

Figure 1. An image of SAM site. (a) Pseudo-intensity image; (b) Detection and labeling of the targets; (c) Object labels from detection/recognition algorithms.

After application of the rule-based system, outputs were converted to linguistic labels. From the 10 output values, the variable with the highest confidence value was picked to describe the relation between objects in the scene. This is a simplistic linguistic approximation approach, but has produced good results in [18]. The automated description of Figure 1 is shown in Figure 2.

> There are 5 missile launchers (1,2,3,6,8).
> They surround a center vehicle (4).
> The image includes a SAM site.
> A convoy of vehicles (5,7,9,10) is belowright of the SAM site.

Figure 2. Linguistic output for Figure 1 from the fuzzy rule-based system.

The description in Figure 2 is typical of those obtained in [18]. How do we judge the performance of high level scene description? We only argued from an intuitive basis. That is, from comparing the original images to the linguistic representation, do we believe that the description captured the "essence" of the scene? As mentioned in the Introduction, appealing to intuition has been a standard means of justifying output of spatial relationship definitions. There needs to be a better method. One approach that involves using this system to produce a description to match against that generated by another means, e.g., a person or a different description package. Preliminary data is encouraging [17]. However, the "language" used in [18] is very coarse. Better descriptive terminology needs to be created and matched to the spatial relationship definitions, particularly if it can be tailored to particular individuals.

3. Relative Position and F-histograms

Now we describe the new approach employing force histograms and show how it can be used to develop finer linguistic descriptions between objects. Let A and B be two image objects. For any direction θ, the function value $F^{AB}(\theta)$ represents the total weight of the arguments that can be found in order to support the proposition "A is in direction θ of B". More precisely, it is the resultant of elementary forces that are exerted by the points of As on those of B, where each tends to move B in direction θ (Figure 3). If the elementary forces are in inverse ratio to d^r, where d represents the distance between the points considered and r is a positive real, then F is denoted by F_r. For instance, the F function associated with the universal law of gravitation is F_2. When considering all angles, F^{AB} is called the *histogram of forces associated with* (A,B), or the F*–histogram associated with* (A,B). It is demonstrated in [13] that F_0–histograms coincide with angle histograms, It is interesting that F_0–histogram and F_2–histogram have very different characteristics (Figure 4). The former gives a global view of the situation. It considers the closest parts and the farthest parts of the objects equally, whereas F_2–histograms focus on the closest parts.

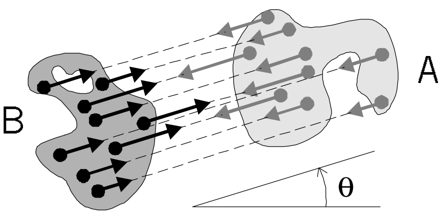


Figure 3. Computation of FAB(θ). It is the scalar resultant of forces (black arrows). Each one tends to move B in direction θ.

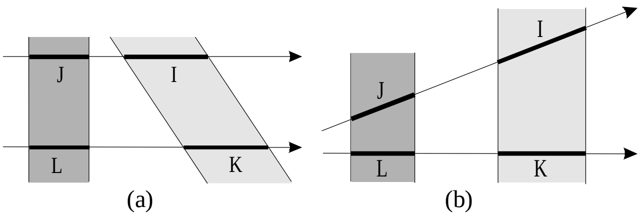


Figure 4. Main characteristics of the F_0 and F_2. a) Independence from distance (F_0–histograms): the force exerted by I on J is equal to that by K on L; b) Independence from scale (F_2–histograms): the force exerted by I on J is equal to that exerted by K on L.

3.1 F-Histograms and Directional Relations

Consider for instance the proposition "A is to the right of B"(other directions are similar). While the F-Histograms can supply a much richer Linguistic directional quantization, we restrict ourselves to the 4 principle directions here. In order to assess the degree of truth of this proposition, the set of directional forces is divided into four quadrants (Figure 5). The forces $F_r^{AB}(\theta)$ of the first and fourth quadrants are elements which, to various degrees, weaken the proposition "A is to the right of B"; the forces of the second and third quadrants are elements which support the proposition. Some forces of the third quadrant are used to compensate — as much as possible — the contradictory forces of the fourth one (Figure 5, Figure 6(a)). Forces of the second quadrant are used in a symmetrical way to compensate the contradictory forces of the first one. The remaining forces are called the effective forces. Each effective force is now divided into two components, as in Figure 6(b). One is "optimal" and the other "sub-optimal". The division is a user defined directional sensitivity for the F_r–histogram. The optimal components support the idea that A is "perfectly" to the right of B. The average direction α_r(RIGHT) of the effective forces is then computed. Finally, the degree of truth a_r(RIGHT) of the proposition "A is to the right of B" is computed as:

$$a_r(\text{RIGHT}) = \mu(\alpha_r(\text{RIGHT})) \times b_r(\text{RIGHT})$$

In this expression, b_r(RIGHT) denotes the percentage of the effective forces (Figure 5), and μ the membership function of a fuzzy set of $[-\pi,\pi]$ that can be used to define RIGHT as in Figure 7 (similar to those in [11,12]). Note that the most optimistic point of view consists in saying that any effective force is optimal. Then, α_r(RIGHT) is equal to 0 radians and $\mu(\alpha_r(\text{RIGHT}))$ to 1. According to F_r^{AB}, the value b_r(RIGHT) therefore corresponds to the maximum degree of truth that can reasonably be attached to the proposition "A is to the right of B".

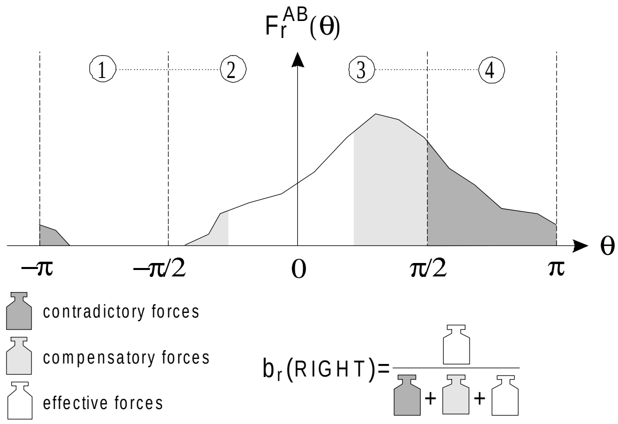


Figure 5. Breakdown of the forces for RIGHT (I).

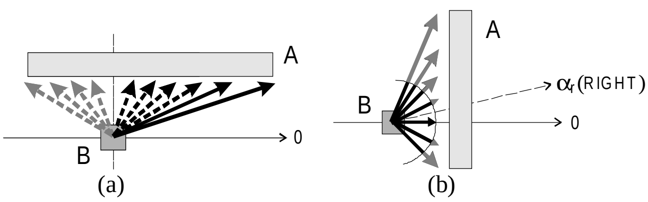


Figure 6. Breakdown of the forces for RIGHT (II). a) Gray dotted arrows: contradictory forces. Black dotted arrows: compensatory forces. Black continuous arrows: effective forces. b) Gray arrows: sub-optimal components. Black arrows: optimal components.

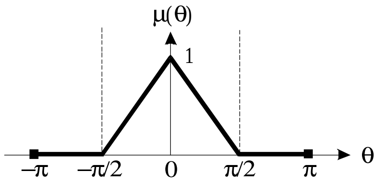


Figure 7. The degree of truth of the proposition "A is to the RIGHT of B".

4. Generation of Linguistic Descriptions

Let A and B be two objects. Our aim here is to give a linguistic description of the relative position between A and B that fundamentally relies on the sole primitive directional relationships: "to the right of", "above", "to the left of" and "below". The description is generated from F_0^{AB} and F_2^{AB}, which have different and interesting characteristics. We combine the opinions given by these two histograms.

4.1. The Numerical Features

Let Δ be the set of the four primitive directions: Δ={RIGHT,ABOVE,LEFT,BELOW}. Consider an element δ of Δ. A degree of truth $a(\delta)$ has to be attached to the proposition "A is in direction δ of B". Now, $a_0(\delta)$, the value proposed by F_0^{AB}, is never too optimistic, but is often too cautious. This drawback will be corrected considering F_2^{AB}. However, F_2^{AB}'s opinion may be excessive: sometimes excessively pessimistic, and sometimes excessively optimistic. There are actually three cases.

1. $a_2(\delta) > b_0(\delta)$

According to F_0^{AB}, the value $b_0(\delta)$ is the maximum degree of truth that can reasonably be attached to the proposition "A is in direction δ of B". Therefore, F_2^{AB} conflicts with F_0^{AB}. We temper F_2^{AB}'s enthusiasm and set $a(\delta) = b_0(\delta)$.

2. $a_0(\delta) > b_2(\delta)$

According to F_2^{AB}, the value $b_2(\delta)$ is the maximum degree of truth. Therefore, F_0^{AB} comes into conflict with F_2^{AB}. We ignore the excessive pessimism of F_2^{AB} and set: $a(\delta) = a_0(\delta)$.

3. $a_2(\delta) \leq b_0(\delta)$ and $a_0(\delta) \leq b_2(\delta)$

There is no conflict. We set: $a(\delta) = \max\{a_0(\delta),a_2(\delta)\}$

It is easy to see that in the three cases: $a(\delta) = \max\{a_0(\delta),\min\{a_2(\delta),b_0(\delta)\}\}$. Moreover, the first and second cases can be rewritten: $a(\delta)>\min\{b_0(\delta),b_2(\delta)\}$. The value $\min\{b_0(\delta),b_2(\delta)\}$ measures the agreement between the two sources of information and allows conflicts to be determined. Six parameters are extracted from the analysis of the histograms F_0^{AB} and F_2^{AB}, and used in order to constitute the linguistic description of the relative position between A and B. These values d_1, m_1, δ_1, d_2, m_2 and δ_2 are defined as follows:

$$d_1 = \max_{\delta\in\Delta} a(\delta) \qquad d_2 = \max_{\delta\in\Delta-\{\delta_1\}} a(\delta)$$
$$a(\delta_1) = d_1 \qquad a(\delta_2) = d_2$$
$$m_1 = \min\{b_0(\delta_1),b_2(\delta_1)\} \qquad m_2 = \min\{b_0(\delta_2),b_2(\delta_2)\}$$

Here, δ_1 is the primary direction, and δ_2 the secondary direction. The degree of truth $a(\delta)$ that we decided to attach to the proposition "A is in direction δ of B" is maximum when δ is δ_1. The linguistic description will generally be composed of three parts. The first part is the main part of the description (e.g., "A is to the *right* of B"). It involves the primary direction. The second part supplements the description (e.g., "but a little *above*"). It involves the secondary direction. The third part indicates to what extent the four primitive directional relationships are suited to describing the relative position of the objects (e.g., "The description is *satisfactory*"). In other words, it indicates to what extent it is necessary to turn or not to other spatial relations (*e.g.* "surrounds").

4.2. The System of Rules

A simple set of rules has been implemented to test out the descriptive nature of this approach. The preliminary rules are interval based, and hence, not fuzzy. The first part of the description uses the rules shown in Figure 8(a). In all parts of Figure 8, the thickness of the connections indicate the strength of the parameter. If the primary direction is RIGHT, and the two objects can be assimilated to points, then m_1 is then equal to 1 (no possibility of conflict between the sources of information F_0^{AB} and F_2^{AB}). The proposition "A is to the right of B" is more or less true, depending on its degree of truth, i.e. on d_1. The three linguistic terms *perfectly*, (--$_2$) (which is a void adverb - the standard situation) and *mostly* have been chosen here. If A and B cannot be assimilated to points, m_1 may be lower than 1. Depending on the amount of ambiguity, *perfectly* degenerates into (--$_1$) or *nearly*$_1$, (--$_2$) into *nearly*$_2$ or *loosely*$_1$, etc. Note that if m_1 or d_1 are very low values (very serious conflict, very ambiguous configuration), no rule is activated. In this case, no pertinent linguistic description relying on the sole primitive directional relationships can be given, and the description will be "???????". This dictionary of terms can be tailored to individual users.

The second part of the description depends on δ_2, d_2 and m_2. It uses the rules shown in Figure 8(b). Suppose for instance that the secondary direction is ABOVE. Depending on m_2, the proposition "A is above B" may turn into "A is shifted upward relative to B". The signification of the second expression is presented in Figure 9, and its connection with the primary is illustrated by Figure 10. Whatever the expression chosen, two adverbs compete, depending on the degree of truth of the proposition. If A and B can be assimilated to points, at least one of the values m_2 and d_2 is very low, and "A is above B" is the only possible choice. If both values are very low, no rule is activated, and the second part of the description is void.

In certain cases, one of the rules shown in Figure 8(c) is activated, and the two first parts of the description are combined, using one of the four compound directions: ABOVE−RIGHT, ABOVE−LEFT, BELOW−LEFT and BELOW− RIGHT.

If there is no ambiguity concerning A and B's relative position, the rule leading to the term ($--_3$) is the only one that can be activated. The connections of an expression such as “A is above-right of B” with the other kinds of expressions are illustrated in Figures 12 and 14.

If a pertinent linguistic description relying on the primitive directional relationships can be given, then the description assesses itself using the last rules shown in Figure 8(d).

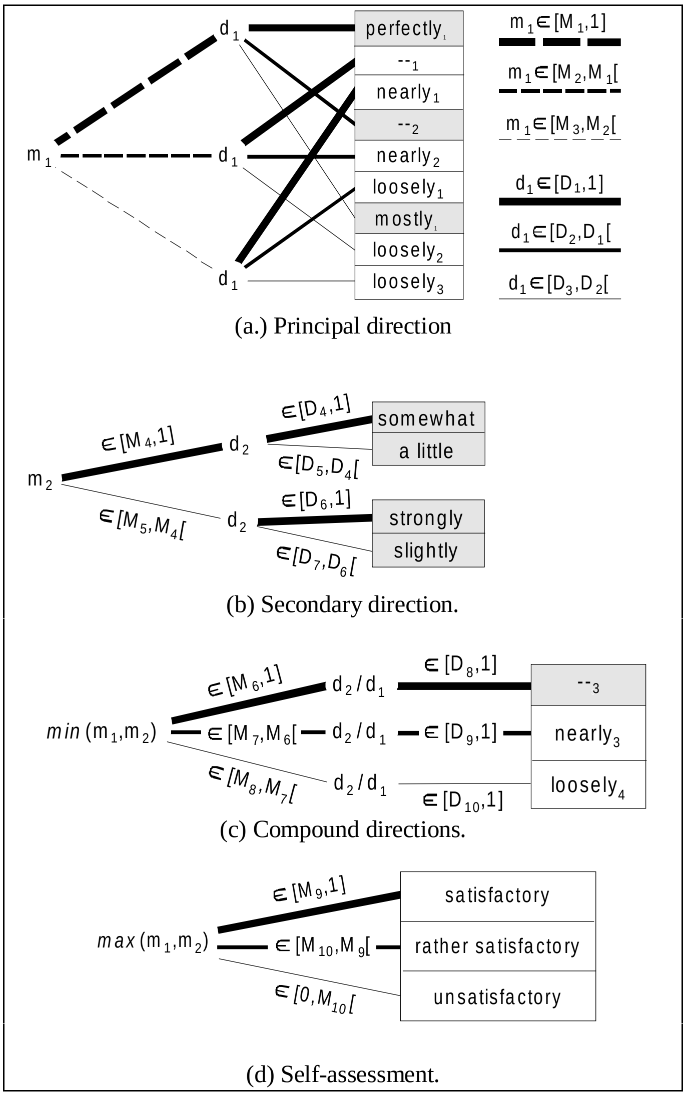


Figure 8. The system of rules.

There are many thresholds for these crisp rules. Most of the values D_1 to D_{10} have been deduced from geometric observations, the others have been determined empirically, considering typical configurations such as those presented in Figures 9 and 11.

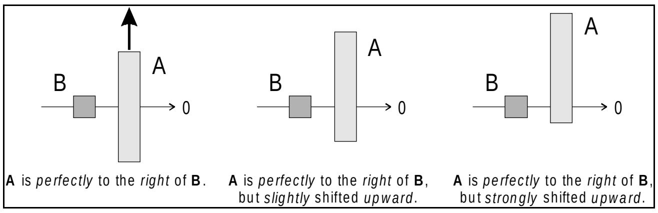


Figure 9. Secondary direction and Shifting. In each case, the self-assessment of the description is: “The description is *satisfactory*.”

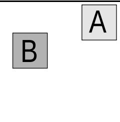
A is ($--_2$) to the *right* of **B**, but *strongly* shifted *upward* relative to **B**. The description is *satisfactory*.		**A** is ($--_2$) to the *right* of **B**, but *a little above*. The description is *satisfactory.*

Figure10. An example of equivalent descriptions.

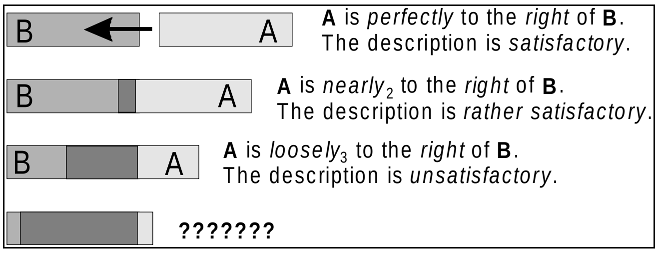
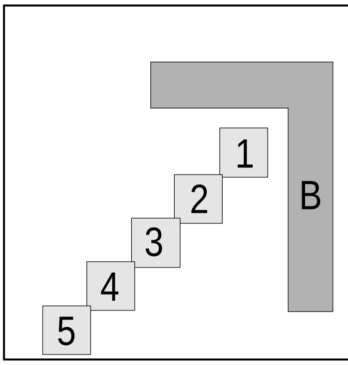


Figure 11. Training images and terminology: Self-assessment.



1 is loosely$_4$ below-left of **B**.
The description is unsatisfactory.
2 is loosely$_4$ below-left of **B**.
The description is rather satisfactory.
3 is loosely$_4$ below-left of **B**.
The description is satisfactory.
4 is nearly$_3$ below-left of **B**.
The description is satisfactory.
5 is ($--_3$) below-left of **B**.
The description is satisfactory.

Figure 12. Test image sequence to show connections of the two parts.

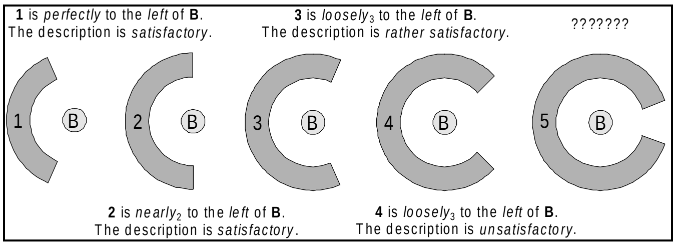


Figure 13. Test images to show the need for additional description.

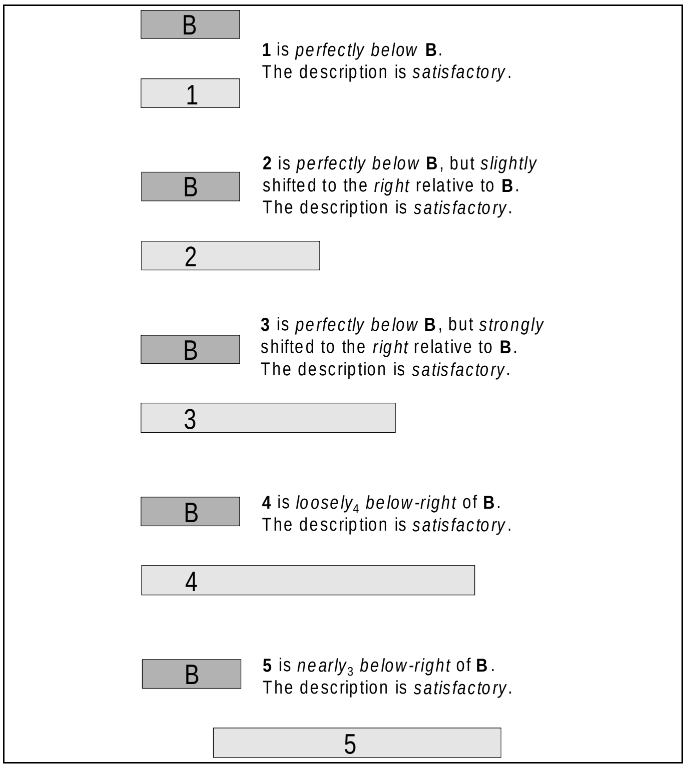


Figure 14. More test objects

5. Conclusions

In this paper, we indicate how spatial relations can be used with rule-based approaches for scene description. We introduced a new method to use F-Histograms to enhance the linguistic expressions by means of hedges and a self-assessment. Preliminary results are quite encouraging. Clearly more work is necessary to exploit the richness of the information contained in the F-Histograms.

Acknowledgements

J. Keller wishes to express his gratitude for support from the Office of Naval Research grant N00014-96-0439.

References

[1] R. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images", *Artificial Intelligence*, Vol. 17, 1981, pp. 285-348.

[2] D. M. Russell, "Where Do I Look Now?" *Proceedings, Pattern Recog. and Image Processing*, 1979, pp. 175-183.

[3] E. Walker, M. Herman and T. Kanade, "A Framework for Representing and Reasoning about Three-Dimensional Objects in Vision", *AI Mag.*, Vol. 9, No. 2, 1988, pp. 47-58.

[4] H. Kang and E. Walker, "Characterizing and Controlling Approximation in Hierarchical Perceptual Grouping", *Fuzzy Sets and Systems*, Vol. 65, 1994, pp. 187-224.

[5] R. Antony, "A Hybrid Spatial/Object-Oriented DBMS to Support Automated Spatial, Hierarchical, and Temporal Reasoning", in S. Chen(ed) Advances in Spatial Reasoning, Vol. 1, Ablex Publishing, Norwood, NJ, 1990, pp. 63-132.

[6] J. Freeman, “The Modeling of Spatial Relations,” *Computer Graphics and Image Processing* , Vol. 4, 1975, pp. 156-171.

[7] G. Retz-Schmidt, “Various Views on Spatial Relations,” AI Magazine, Summer 1988, pp. 95-105.

[8] S. Dutta, “Approximate Spatial Reasoning: Integrating Qualitative and Quantitative Constraints,” International J. of Approximate Reasoning , Vol. 5, pp. 307-331, 1991.

[9] P. Winston, "Learning Structural Descriptions from Examples", in P. Winston (ed.), *The Psychology of Computer Vision* , McGraw-Hill , New York, 1975.

[10] X. Wang, and J. M. Keller, "Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches", *Fuzzy Sets and Systems*, Vol. 101, No. 1, 1999, pp. 5-20.

[11] K. Miyajima and A. Ralescu, "Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations", *Fuzzy Sets and Systems*, 1994, pp. 225-236.

[12] A. Krishnapuram, J. M. Keller and Y. Ma, "Quantitative Analysis of Properties and Spatial Relations of Fuzzy Image Regions", *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 3, 1993, pp. 222-233.

[13] P. Matsakis and L. Wendling, "A New Way to Represent the Relative Position Between Areal Objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press, 1999.

[14] I. Bloch, "A Fuzzy Relative Position Between Objects in Images: a Morphological Approach", *Proceedings, ICIP'96*, Lausanne, Vol. II, 1996, pp. 987-990.

[15] P. Gader, "Fuzzy Spatial Relations Based on Fuzzy Morphology", *Proceedings, IEEE International Conference on Fuzzy Systems*, Barcelona, 1997, pp. 1179-1183.

[16] X. Wang, J. Keller and P. Gader, "Using Spatial Relationships as Features in Object Recognition", *Proceedings of NAFIPS'97*, Syracuse, NY, September, 1997, pp. 160-165.

[17] J. Keller, P. Gader and X. Wang, "LADAR Scene Description using Fuzzy Morphology and Rules", *Proceedings, IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, in conjunction with CVPR '99, Fort Collins, CO, June, 1999.

[18] J. Keller and X. Wang, "A Fuzzy Rule-based Approach for Scene Description Involving Spatial Relationships" *Computer Vision and Image Understanding*, revised, 1999.

[19] J. M. Keller, P. D. Gader, R. Krishnapuram, X. Wang, A. Hocaoglu, H. Frigui and J. Moore, "A Fuzzy Logic Automatic Target Detection System For LADAR Range Images", *Proceedings, FUZZ-IEEE’98*, Anchorage, Alaska, May 1998, pp. 71-76.