In: Symbolic and quantitative approaches to reasoning and uncertainty.
A. Hunter and S Parsons, eds, pp. 104–115 (Springer-Verlag, Germany 1999)

Anchoring Symbols to Vision Data by Fuzzy Logic

Silvia Coradeschi[1] and Alessandro Saffiotti[2]

[1] Dept. of Computer and Information Science
Linköping University, S-581 83 Linköping, Sweden
silco@ida.liu.se

[2] Applied Autonomous Sensor Systems
Dept. of Technology and Science
Örebro University, S-701 82 Örebro, Sweden
alessandro.saffiotti@ton.oru.se

Abstract. Intelligent agents embedded in physical environments need the ability to connect, or *anchor*, the symbols used to perform abstract reasoning to the physical entities which these symbols refer to. Anchoring must rely on perceptual data which is inherently affected by uncertainty. We propose an anchoring technique based on the use of fuzzy sets to represent uncertainty, and of degree of subset-hood to compute the partial match between signatures of objects. We show examples where we use this technique to allow a deliberative system to reason about the objects (cars) observed by a vision system embarked in an unmanned helicopter, in the framework of the WITAS project.

1 Introduction

The focus of this paper is on autonomous systems embedded in a real-world, physical environment. A typical example is an autonomous mobile robot who has to providing services inside a factory, or to explore a far planet. Being embedded in the physical world, these systems need to incorporate processes at the sensori-motor level that provide the needed perceptual and execution capabilities. However, these systems also need the ability to perform high-level, abstract reasoning if they are to operate reliably in a dynamic and uncertain world without the need for human assistance. For example, a mail delivery robot faced with a closed door should decide whether to plan an alternative way to achieve its goal, or to reschedule its activities and try again this delivery later on.

The need to integrate low-level and high-level representations and processes is one of the major challenges of autonomous embedded systems, and most of the current architectures for autonomous robots address this challenge in some way or another [6]. In our work, we focus on one particular aspect of this integration problem: the connection between the abstract representations used by the high-level reasoning processes to denote a specific physical object, and the data in the low-level processes that correspond to that object. Following [8, 4], we call *anchoring* the process of establishing this connection.

In general, we assume that the high-level process associates each object in its universe of discourse to a unique name, and to a set of properties that (non-univocally) describe that object. For example, an object named 'car-3' with the description 'small red Mercedes on Road-61.' Anchoring this object then includes two steps: (i) use the perceptual apparatus to find an object whose observed features match the properties in the description; and (ii) update those properties using the observed values. In our example, anchoring 'car-3' means to: (i) perceptually find a small red Mercedes on Road-61; and (ii) update the description of 'car-3' using the observed size and color of the found car, and possibly other properties like position and speed.

One of the difficulties in the anchoring problem is that the data provided by the sensory system is inherently affected by a large amount of uncertainty. This may result in errors and ambiguities when trying to match these data to the high-level description of an object. In order to improve the reliability of the anchoring process, this uncertainty has to be taken into account in the proper way. Research in fuzzy logic has produced a number of techniques for dealing with different facets of uncertainty [5, 1] . In this work, we propose to use these techniques to define a *degree of matching* between a perceptual signature and an object description. The possibility to distinguish between objects that match a given description at different degrees is pivotal to the ability to discriminate perceptually similar objects under poor observation conditions. Moreover, degrees of matching allow us to consider several possible anchors, ranked by their degree of matching. Finally, these degrees can be used to reason about the quality of an anchor, and to perform higher-level decision making; for example, we can decide to engage in some active perception in order to get a better view on a candidate anchor.

In the rest of this paper, we deal with the anchoring problem in the context of an architecture for unmanned airborne vehicles (UAVs) used for traffic surveillance. This architecture, outlined in the next section, integrates several subsystems, including a vision system and an autonomous decision making system. The anchoring problem in this context means to make the decision making system and the vision system agree about the identity of the objects which they are talking about, like in our 'car-3' example. This is discussed in section 3. In Section 4, we show how we use fuzzy sets to represent the uncertain data in our domain, and to compute degrees of matching. Section 5 illustrates the use of these degrees by going through a couple of examples, which are run in simulation. Finally, section 6 discusses the results and traces future directions.

2 The WITAS Project

The WITAS project, initiated in January 1997, is devoted to research on information technology for autonomous systems, and more precisely to unmanned airborne vehicles (UAVs) used for traffic surveillance.

The general architecture of the system is a standard three-layered agent architecture consisting of a deliberative, a reactive, and a process layer. The deliberative layer generates at run-time probabilistic high-level predictions of the behaviors of agents in their environment, and uses these predictions to generate conditional plans. The reactive layer performs situation-driven task execution, including tasks relating to the plans generated by the deliberative layer. The reactive layer has access to a library of task and behavior descriptions, which can be executed by the reactive executor. The process layer contains image processing and flight control, and can be reconfigured from the reactive layer by means of switching on and off groups of processes. Besides vision, the sensors and knowledge sources of the system include: a global positioning system (GPS) that gives the position of the vehicle, a geographical information system (GIS) covering the relevant area of operation, and standard sensors for speed, heading and altitude.

The system is fully implemented in its current version. Because of the nature of the work most of the testing is being made using simulated UAVs in simulated environments, even though real image data has been used to test the vision module. In a second phase of the project, however, the testing will be made using real UAVs. More information about the project can be found at [10].

Of particular interest for this presentation is the interaction between the reactive layer and the image processing in the process layer. This is done by means of a specialized component for task specific sensor control and interpretation called the Scene Information Manager (SIM).

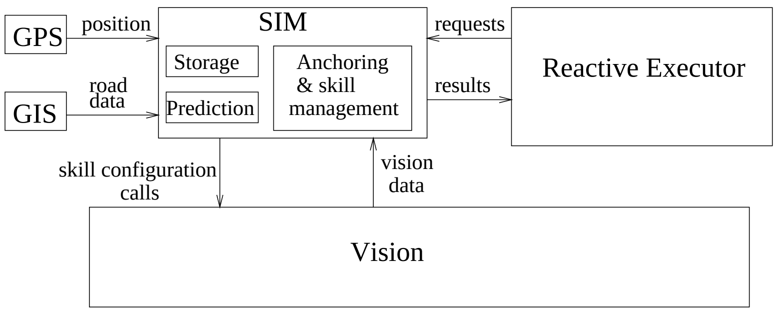

Fig. 1. Overview of the Scene Information Manager and its interaction with the Vision module and the Reactive Executor.

The SIM, figure 1, is part of the reactive layer and it manages sensor resources: it reconfigures the vision module, via skill configuration calls, on the basis of the requests of information coming from the reactive executor, it anchors symbolic identifiers to image elements (points, regions), and it handles simple vision failures, in particular temporary occlusion and errors in car re-identification.

In this paper we focus on the anchoring functionality of the SIM, and in particular on the matching of symbolic identifiers to image elements in the presence of uncertainty in the data provided by image processing. A description of the WITAS architecture and of the role of the SIM in it can be found in [3].

3 Anchoring in the SIM

Two of the main aspects of anchoring implemented in the SIM are identification of objects on the basis of a visual signature expressed in terms of concepts, and re-identification of objects that have been previously seen, but that have then been out of the image or occluded for a short period.

For identification and re-identification the SIM uses the visual signature of the object, typically color and geometrical description, and the expected positions of the object. For instance if the SIM has the task to look for a red, small Mercedes near a specified crossing, it provides the vision module with: the coordinates of the crossing; the HSV (hue, saturation and value) representation of "red"; and the length, width and area of a small Mercedes. The measurements done in the vision module have a degree of inaccuracy, and the SIM provides the vision module also with the intervals inside which the measurement of each of the features is acceptable. The size of the interval depend on how discriminating one wants to be in the selection of the objects and also, in the case of re-identification of an object, on how accurate previous measurements on the object were.

The vision module receives the position where to look for an object and the visual signature of the object and it is then responsible for performing the processing required to find the objects in the image whose measures are in the acceptability range and report the information about the objects to the SIM. The vision module moves the camera toward the requested position and it calculates for each object in the image and for each requested feature of the object an interval containing the real value. If the generated interval intersects with the interval of acceptability provided in the visual signature for the feature, the feature is considered to be in the acceptability range. The vision module reports to the SIM information about color, shape, position, and velocity of each object whose features are all in the acceptability range.

Intersection of intervals is a simple, but not very discriminating method to identify an object. As a consequence, several objects that are somehow similar to the intended one can be sent back by the vision module to the SIM. The SIM then needs to apply some criteria in order to perform a further selection of the best matching object between those reported by the vision module. The selection of the best matching object should depend on how well the objects match the different aspects of the signature, but also on the accuracy of the measurements performed by the vision and their reliability. In what follows, we show how we perform this selection using fuzzy signature matching.

4 Fuzzy Signature Matching

Let us look more closely at the process of anchoring a high-level description coming from the symbolic system (reactive executor) to the data coming from the vision module. As an example, consider the case in which a task needs to refer to 'a small red Mercedes.' The SIM system has to link two types of data: on the one side, the description containing the symbols 'red,' 'small' and 'Mercedes' received from the symbolic system; and on the other side, the values of the measurable features (HSV, length, etc.) of observed cars which are sent by the vision system. Anchoring implies to convert these representations to a common frame, and to find the car that best matches the description. In our case, we have chosen to convert symbols to the values used by the vision system.

4.1 Uncertainty Representation

In general, both the symbolic descriptions and the data coming from the vision system are affected by several types of inexactness. Symbolic descriptions use linguistic terms like 'red' and 'small' that do not denote a unique numerical value. Fuzzy sets are commonly considered to be an adequate representation of linguistic terms [11, 5], so in our system we have chosen to map each symbol of this kind to a fuzzy set over the relevant space. For example, we associate the term 'red' to three fuzzy sets: one for the hue characterizing the tint of color, one for the saturation characterizing the purity of the color, and one for value characterizing its intensity. Fig 2 (left) shows the fuzzy set for the hue. This fuzzy set is interpreted as follow: for each possible value of hue h, the value of $red(h)$ measures, on a $[0, 1]$ scale, how much h can be regarded as 'red'.

As a second example shows how we represent the linguistic term 'small-Mercedes' by a set of fuzzy sets over the space of the possible values of length, width and area of the car. Fig 2 (right) shows the fuzzy set for the area. The reason why we consider the term 'small-Mercedes' and not just 'small' is because what should be regarded as 'small' depends on the type of car we are talking about. In practice, we use a database that associates each car type to its typical length, size, and area, represented by fuzzy sets. Cars of unknown types are associated with generic fuzzy sets, like the 'small' (car) shown by the dotted lines in the picture. In our implementation, we only consider trapezoidal fuzzy sets for computational reasons.

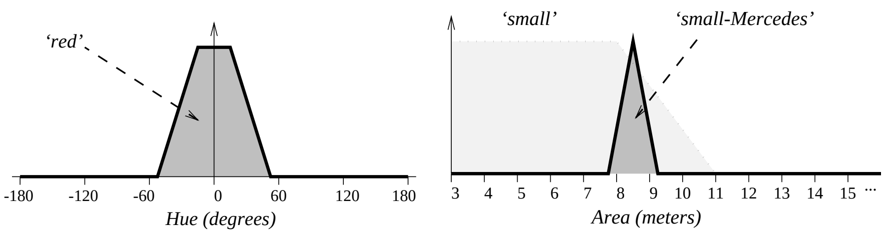

Fig. 2. Fuzzy sets for representing the symbols 'red' and 'small-Mercedes.'

Data from the vision system are affected by uncertainty and imprecision in several ways. Consider the measurement of the area of an observed car. Roughly, this measurement is done by identifying the edges of the car in the image, counting the pixels occupied by the car, and converting this count to a metric measure by some geometric computations. There are a number of factors that influence the correctness of the measured value. First, the discretization of the image limits the precision of the measure. Second, the measurement model may be inaccurate: for instance, cars are assumed to be rectangular, but this is not completely true in reality. The impact of both effects depends on the size of the car in the image, which in turn depends on its distance from the camera and on the focal length of the camera. Third, the measurement is affected by the perspective distortion due to the angle between the car plane and the optical axis: if the car is not perpendicular to the optical axis, its projection on the image will be shorter. Fourth, the geometric parameters needed to compute the length may not be know with precision: for example, the angle between the car and the optical axis depends on the inclination of the road and of that of the car, both of which are hard to evaluate. Finally, the measured value can be totally invalid if there has been an error in the identification of the edges of the car in the image; for instance, if the car has been merged with its shadow, or with another car in front of it.

The above discussion reveals that there is a great amount of uncertainty that affects the measured value for the length of an object; and that this uncertainty is very difficult to precisely quantify — in other words, we do not have a *model* of the uncertainty that affects our measures. Similar observations can be made for other features measured by the vision system: for example, the measurement of the color of an object is influenced by the spectral characteristics of the light that illuminates that object. Given these difficulties nature of the uncertainty in the data coming from the vision system, then, we have chosen to use (trapezoidal) fuzzy sets to represent the data coming from the vision system. A fuzzy set representation allows us to incorporate heuristic knowledge about the inexactness that affects our measures by choosing a specific shape for the corresponding fuzzy set. For example, Fig 3 show the fuzzy sets that represent the observed hue and area of a car object, respectively. The measure of the hue is rather precise in this case. The construction of the fuzzy set for the area value goes as follows.[1] The vision system has computed the interval $[9.2, 10.8]$ as the possible values for the area (this is an interval because of the image discretization). Knowing that the viewing angle between the camera and the car is about 30°, we slightly increase the upper bound of the interval, meaning that the car might be bigger than it appears. We then take this interval to be the core of our fuzzy set, and we "blur" the edges to account for the other possible sources of errors. Although the definition of the fuzzy sets used to represent measured features is mostly heuristic, it has resulted in good performance in our experiments.

[1] These fuzzy sets are currently constructed inside the SIM from the values provided by the vision system. We are now in the process of moving this construction into the vision system itself, where more informed heuristics can be applied.

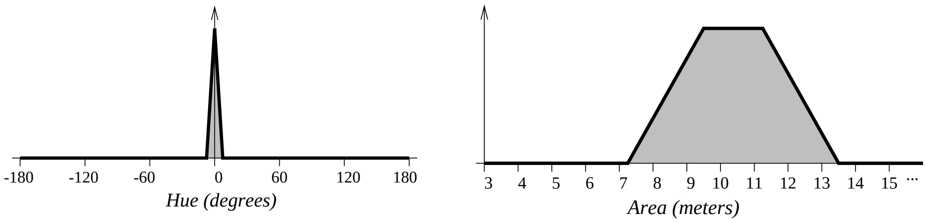

Fig. 3. Fuzzy sets for hue (left) and area (right) obtained from the vision system.

4.2 Fuzzy Matching of one Feature

Once we have represented both the desired description and the observed data by fuzzy sets, we can compute their *degree of matching* using fuzzy set operations. This choice is justified in our case since fuzzy sets can be given a semantic characterization in terms of degrees of similarity [7]. There is however a subtle difference between the notion of similarity and our intended notion of matching. Consider two fuzzy sets A and B over a common domain X which respectively represent the observed data and the target description. The degree of matching of A to B, denoted by $\mathrm{match}(A, B)$, is the degree by which the observed value A can be one of those that satisfy our criterium B. Thus, matching implies some sort of overlap between A and B, but it does not require that A and B have a similar shape. Moreover, matching is not required to be commutative.

In our work, we have tried two different definitions for a degree of matching (see, e.g., [1] for these and alternative definitions). In the first one, we measure how much A and B intersect by measuring the height of $A \cap B$. This gives us the following degree:

$$\mathrm{match}_1(A, B) \; = \; \sup_{x \in X} \min\{A(x), B(x)\} \tag{1}$$

In the second definition, we measure of how much A is a (fuzzy) subset of B by comparing the area of $A \cap B$ and the area of B:

$$\mathrm{match}_2(A, B) \; = \; \frac{\int_{x \in X} \min\{A(x), B(x)\}\, dx}{\int_{x \in X} B(x)\, dx} \tag{2}$$

Different definitions can be obtained using T-norm operators other than min.

The degrees of matching defined by equations (1) and (2) behave in two essentially different ways. Match_1 only depends on the existence of some common elements between A and B, while match_2 compares how much of A is inside B with how much of A is outside B. The difference is graphically illustrated in Fig. 4. When the cores of A and B have no common points (left), both definitions provide a degree of matching smaller than 1. As soon as the cores intersect (mid and right), match_1 always indicates total matching, while match_1 gives us only a partial degree whenever A is not entirely contained into B. In a sense, definition

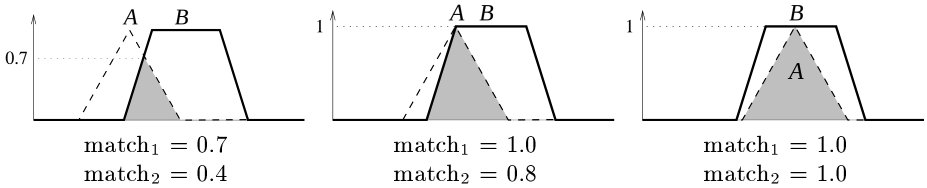

Fig. 4. Three examples of partial matching between a set A and a reference set B.

(1) tells us how much the observed value *may* satisfy our criterium B; while definition (2) tells us how much the observed value *must* satisfy it.

Measure (2) is more discriminating, and it has provided superior empirical results in our domain. We have thus adopted this measure in our system. For computational reasons, however, we approximate (2) by the ratio between the area of the inner trapezoidal envelope of $A \cap B$ and the area of B. When A and B are trapezoidal fuzzy sets, both areas can be easily computed from the four parameters of the trapezoids.

4.3 Fuzzy Matching of Several Features

Once we have computed a degree of matching for each individual feature, we need to combine all these degrees together in order to obtain an overall degree of matching between a description and a given object perceived by the vision system. The simplest way to combine our degrees is by using a *conjunctive* type of combination, where we require that each one of the features matches the corresponding part in the description. Conjunctive combination is typically done in fuzzy set theory by T-norm operators [9, 5], whose most used instances are min, product, and the Łukasiewicz T-norm $max(x+y-1, 0)$. In our experiments, we have noticed that the latter operator provides the best results. (See [2] for an overview of alternative operators.)

The overall degree of matching is used by the SIM to select the best anchor among the candidate objects provided by the vision module. For each candidate, the SIM first computes its degree of matching to the intended description; then it ranks these candidates by their degree, and return the full ordered list to the reactive executor. Having a list of candidates is convenient if the currently best one later turns out not to be the one we wanted. Also, it is useful to know how much the best matching candidate is better than the other ones: if the two top candidates have similar degrees of matching, we may decide to engage in further exploratory actions in order to disambiguate the situation before committing to one of them — for instance, we may request the vision system to zoom on each candidate in turn in the hope to get more precise data.

While conjunctive T-norm combination has produced a satisfactory behavior in our preliminary experiments, there are a few reasons why more complex types of combinations seem more adequate to our case. First, some of the features are more critical than others, and we would like their degree of matching to have a stronger impact on the overall degree. Second, in some situations some values are known not to be reliable and should have little impact on the overall degree of matching: for instance, the observed size of the car is not reliable when the viewing angle is large. Finally, some features have errors which are strongly correlated (e.g., length and width) and it might be wise to combine their individual degrees of matching by an idempotent operator. The search for a more adequate combination technique is part of our current development.

5 Fuzzy Signature Matching at Work

We illustrate the use of the fuzzy signature matching by two examples on the scenario taken from the WITAS project shown in Fig. 5. In this scenario, the deliberative system is interested in a red car of a specified model in the vicinity of a given crossing. Four cars are situated around that crossing, moving in different directions. The cars are all red, but of different models: a small van, a big Mercedes, a small Mercedes, and a Lotus. Discriminating between these cars is made more difficult by the fact that the helicopter views the crossing at an inclination of about 30 degrees: this results in some perspective distortions, thus introducing more uncertainty in the extraction of geometrical features.

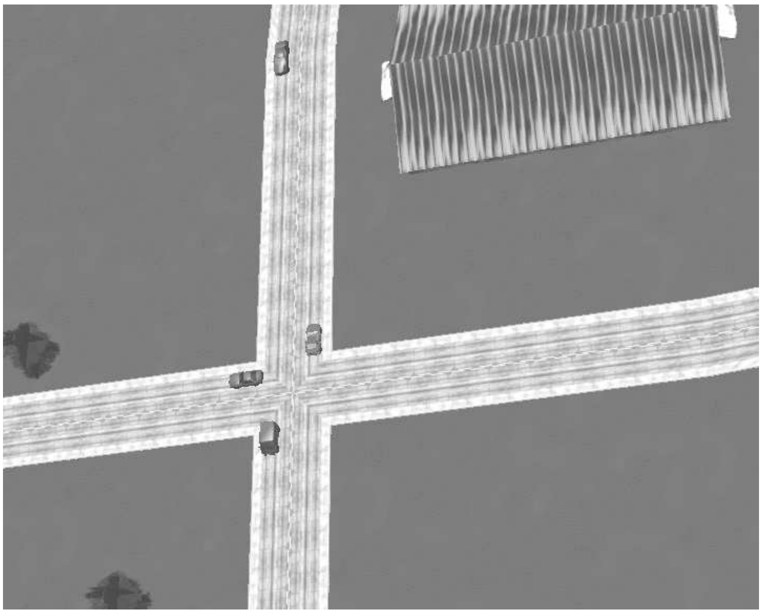

Fig. 5. The simulated scenario for our examples.

In our first example, the deliberative system decides to follow 'Van-B', which is described as a red van. The SIM sends the prototypical signature of a red van to the vision module. Since all the four cars in the image are red, and they have fairly similar shapes, the vision module returns the observed signatures of all the four cars to the SIM. These signatures are then matched against the desired signature by our routines, resulting in the following degrees of matching:

ID	Color	Shape	Overall
66	1.0	0.58	0.58
67	1.0	0.38	0.38
68	1.0	1.0	1.0
69	1.0	0.0	0.0

The ID is a label assigned by the vision system to each car found in the image. The degree of matching for the color is obtained by combining the individual degrees of hue, saturation, and value; in our case, this will be 1.0 for all the cars as they are all red. The degree of matching for the shape is the combination of the individual degrees of matching of length, width, and area. The overall degree is the Łukasiewicz combination of the color and shape degrees. In this case, car 68 is correctly[2] identified as the best candidate, and an anchor to that car is thus returned to the deliberation system.

In the second example, the deliberative system is interested in 'Car-D', a red small Mercedes. The SIM sends the corresponding prototypical signature to the vision module, and again gets the signatures of all the four cars in the image as an answer. In this case however, the helicopter is at a long distance from the crossing and it views the crossing at an inclination of about 30 degrees. By applying our fuzzy signature matching routine, we obtain the following degrees:

ID	Color	Shape	Overall
66	1.0	0.65	0.65
67	1.0	0.84	0.84
68	1.0	0.0	0.0
69	1.0	0.97	0.97

Cars 66, 67 and 69 match the desired description to some degree, while car 68 can safely be excluded. The SIM tries to improve the quality of the data by asking the vision module to zoom on each one of cars 66, 67, and 69 in turn. Using the observed signatures after zooming, the SIM then obtains the new degrees of matching:

ID	Color	Shape	Overall
66	1.0	0.30	0.30
67	1.0	0.70	0.70
69	1.0	0.21	0.21

The closer view results in a smaller segmentation error, since the scale factor is smaller, and hence in more narrow fuzzy sets. As a consequence, all the degrees of matching have decreased with respect to the previous observation. What

[2] This verification was done manually off-line.

matters here, however, is the relative magnitude of the degrees obtained from comparable observations, that is, those which are collected in the above table. The SIM sends the identifiers of each of the car to the reactive executor together with their degrees of matching. These degrees allow the reactive executor to select car 67 as the best candidate.

The reactive executor now has the option to try to further improve its choice by commanding the helicopter to fly over car 67 and take another measurement from above the car — the best observation conditions for the vision system. If we do this, we finally obtain a degree of matching of 1.00 for car 67. Note that this degree could as well have dropped, thus indicating that car 67 was not really the car that we wanted. In this case, the reactive executor could have requested the SIM to go back to cars 66 and 69 to get more accurate views.

6 Conclusions

Anchoring symbols to the physical objects that they are meant to denote requires the ability to integrate symbolic and numeric data under uncertainty. Although anchoring is rarely identified as a clearly separated process, we believe that this process must be present in any embedded symbolic system, including most of the current autonomous robots. In this paper, we have considered an instance of the anchoring problem in which we link the car identifiers used at the decision-making level to the perceptual data provided by a vision system. We have shown that explicitly representing and reasoning about the uncertainty in this problem improves the results of the anchoring process.

Our experimental results show that our technique is adequate to handle the ambiguities that arise when integrating uncertain perceptual data and symbolic representations. In particular, the use of fuzzy signature matching definitely improves our ability to discriminate among perceptually similar objects in difficult situations (e.g., perspective distortion). Moreover, degrees of matching allow us to exclude unlikely candidates, and to rank the likely ones by their similarity to the intended description. Finally, degrees of matching can help in decision making; for example, when these degrees indicate a large amount of anchoring ambiguity, the system may decide to engage in active information gathering such as zooming and getting closer to the object in order to obtain better information. It should be noted that the fuzzy logic techniques that we use in our system are not novel. The main novelties of our work are the explicit use of the notion of *anchoring* to integrate symbols and sensor data, and the extension of this notion to take uncertainty into account. Further insights on the notion of anchoring in autonomous agents are presented in [4].

The work reported in this paper is still in progress, and many aspects need to be further developed. First, we need to study more sophisticated forms of aggregation of the individual degrees of matching of different features into an overall degree. Second, we plan to include features of a different nature into the matching process, like the observed position and velocity of the cars. Finally, until now we have only performed experiments in simulation. At the current stage of development of the WITAS project, the vision system takes as input the video frames produced by a 3D simulator. Although this configuration results in some amount of noise and uncertainty in the extracted features, we are aware that a real validation of our technique will only be possible when we have access to the real data from an embarked camera.

Acknowledgments

Discussions with Thord Andersson, Dimiter Driankov, and Klas Nordberg have been greatly beneficial. This work is part of the WITAS project, which is fully sponsored by the Knut and Alice Wallenberg Foundation.

References

1. H. Bandemer and W. Naether, editors. *Fuzzy data analysis.* Kluwer Academic, 1992.
2. I. Bloch. Information combination operators for data fusion: A comparative review with classification. *IEEE Trans. on Systems, Man, and Cybernetics*, A-26(1):52–67, 1996.
3. S. Coradeschi, L. Karlsson, and K. Nordberg. Integration of vision and decision-making in an autonomous airborne vehicle for traffic surveillance. In H. I. Christiansen, editor, *Computer Vision Systems*, volume 1542 of *LNCS*, pages 216–230, Berlin, Germany, 1999. Springer.
4. S. Coradeschi and A. Saffiotti. Anchoring symbolic object descriptions to sensor data: problem statement. *Linköping Electronic Articles in Computer and Information Science*, 1 (2) July 1, 1999. http://www.ep.liu.se/ea/cis/1999/002/.
5. G. Klir and T. Folger. *Fuzzy sets, uncertainty, and information.* Prentice-Hall, 1988.
6. D. Kortenkamp, P. Bonasso, and R. Murphy, editors. *Artificial Intelligence and Mobile Robots.* MIT Press, Boston, MA, 1998.
7. E. H. Ruspini. On the semantics of fuzzy logic. *Int. J. of Approximate Reasoning*, 5:45–88, 1991.
8. A. Saffiotti. Pick-up what? In C. Bäckström and E. Sandewall, editors, *Current Trends in AI Planning — Proc. of EWSP '93*, pages 166–177. IOS Press, Amsterdam, NL, 1994.
9. S. Weber. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy sets and systems*, 11:115–134, 1983.
10. WITAS web page: http://www.ida.liu.se/ext/witas/.
11. L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.