On the Relation between Probabilistic Inference and Fuzzy Sets in Visual Scene Analysis

Ulrich Hillenbrand
Institute of Robotics and Mechatronics, German Aerospace Center
Oberpfaffenhofen, 82234 Wessling, Germany
Email: Ulrich.Hillenbrand@dlr.de

Abstract

Strict probabilistic inference is a difficult and costly procedure, and generally unfeasible in practice for interesting cases. It requires knowledge, storage, and computational handling of usually very complicated probability-density functions of the data. Independence assumptions commonly made to alleviate these problems are often wrong and may lead to unsatisfactory results. By contrast, working with fuzzy sets in data space is simple, while the underlying assumptions have remained largely obscure. Here I derive from probabilistic principles a fuzzy-set-type formulation of visual scene interpretation. The argument is focused on making explicit the conditions for reasoning with fuzzy sets and how their membership function should be constructed. It turns out that the conditions may be fulfilled to a good approximation in some cases of visual scene analysis.

Keywords: scene analysis, segmentation, probabilistic inference, Bayesian inference, fuzzy set

Published as *Pattern Recognition Letters* **25** (2004), pp. 1691–1699.

1 Introduction

Vision poses problems that involve dealing with uncertainties of various kinds. In visual systems, scene interpretation is only accomplished through multiple stages of processing, where intermediate interpretations (features) usually feed to the next level as facts about the scene. Thus, uncertainty is first introduced at the imaging stage through sensor noise, or when one is not fully in control or aware of the lighting conditions or the actual scene

constraints. Uncertainty is further increased at every level of processing where implicit assumptions may not hold and intermediate interpretations may go wrong.

One formalism that attempts to model uncertainties is fuzzy logic. Concepts or attributes that are intrinsically fuzzy, or crisp concepts that cannot be matched accurately are described by fuzzy sets. Fuzzy sets, in turn, are defined by membership functions that take values between 0 and 1 on the domain of discourse, representing the degree of membership of an element to the modeled concept. Membership values for different concepts or attributes are combined variously in so-called fuzzy rules, by application of which the desired inference is achieved.

In computer vision, fuzzy techniques have been applied to all levels of processing, from basic preprocessing to matching with object models to scene interpretation; see Keller (1997), Tizhoosh (1998), and Walker (1998) for brief overviews. Nonetheless, the formalism has never quite made it into the mainstream of the field. One of the reasons may be the vagueness of the assumptions that underlie a fuzzy inference algorithm. Indeed, fuzzy logic itself does not even ascribe a clear empirical meaning to its fuzzy sets (Zadeh, 1978) that would be required for an analysis of such assumptions. What does, e.g., a membership value of 0.3 tell us about an attribute of objects observed in the real world, beyond the vague statement that these objects may or may not have the attribute in question? In fact, the choice of membership functions and of the rules for their combination remain heuristic and are primarily left to an 'expert'.

Another formalism that deals with uncertainty of all kinds is probability theory. Its great advantage over fuzzy logic is its clear empirical grounding by the central limit theorem. In other words, we know what it means for a large number of trials, if an event occurs with probability $p \in [0, 1]$ or, more generally, a quantity has a certain expectation value.

There have been attempts to link fuzzy logic to probability theory by interpreting values of membership functions as probabilities (Hisdal, 1988; Mabuchi 1992, 1997). It is clear, however, that such a step will leave no room for an independent formalism of fuzzy logic, as all meaningful operations will follow from the theory of probability alone.

Strict probabilistic inference, on the other hand, is rarely feasible in practice. It is usually impossible to know, store, or computationally handle the joint probability densities of high-dimensional data that constitute the object models in interesting cases. To alleviate this problem, a number of statistical independencies are commonly assumed. Sometimes, however, such assumptions are terribly wrong and may lead to unacceptable results.

In this article, I derive a simple variant of reasoning with fuzzy sets, starting out from a probabilistic formulation. Some aspects of the argument are rather specific for visual-type data. If one identifies the resulting formalism with standard operations on fuzzy sets, a certain probabilistic interpretation of membership-function values is implied. Moreover, the analysis makes the assumptions explicit that underlie the use of fuzzy-set techniques. It will turn out that these assumptions may be reasonable in some cases of visual scene interpretation.

The main part of the article is contained in Sec. 2. There I shall lay out the arguments that lead from probabilistic inference to fuzzy-set-type reasoning. In Sec. 3, I will touch upon practical issues of learning and implementation. An example application with some real data is given in Sec. 4. Section 5 concludes with a summary of the most important points.

2 From Probabilistic Inference to Reasoning with Fuzzy Sets

This section will reduce the problem of probabilistic inference to its solution by simple fuzzy-set-based reasoning in two steps. First, it is argued that probabilistic considerations may, under appropriate conditions, lead to maximizing the amount of data that are explained in terms of object models. Second, it is shown that the explained amount can be maximized by optimizing the match between fuzzy sets and the data.

In order to avoid potential mathematical trouble with probability measures, all random variables will be treated as taking finitely many discrete values. This is perfectly adequate for data represented in a computer and does not entail any practical limitations. To keep the notational load in this article to a minimum and to support ease of reading, I will not introduce symbols for random variables but only for the values they take. Moreover, I will denote all probabilities by the letter p or, where I want to refer to the concept of a membership function, by μ and specify each type of probability by its index and arguments.

2.1 Probabilistic Inference by Explaining the Maximum Amount of Data

Suppose we obtain by some visual processing (feature extraction) or by more direct measurement N data points $D = \{d_1, d_2, \ldots, d_N\} \subseteq \Delta$ in a data space Δ. Since we are here concerned with vision, the data space will usually be

the image plane or three-dimensional (3D) Euclidean space as a model of physical space. Examples of data are points of high contrast or special texture in an image (edge points, corner points, etc.), 3D points computed from such image locations by a stereo algorithm, or 3D points measured by a range scanner. Different types of data points (differently oriented edge points etc.) may be represented in a data space extended by dimensions for type labeling or parameterization.

We seek an interpretation of the scene in terms of n (not necessarily different) objects with parameters $\Omega = (\omega_1, \omega_2, \ldots, \omega_n)$, where ω_i denotes the pose (position and orientation) of object i in physical space. Central to the following argument is the notion of a segmentation.

Definition 1: *A* segmentation σ *of the data space* Δ *is a sequence of* $n+1$ *subsets of* Δ*,*

$$\sigma = (\Delta_0, \Delta_1, \ldots, \Delta_n) \quad \textit{with} \quad \cup_{i=0}^{n} \Delta_i = \Delta \ , \quad \Delta_i \cap \Delta_j = \emptyset \quad \textit{for } i \neq j. \tag{1}$$

The subset Δ_i *is the segment of data space that contains* all *the data originating* exclusively *from the object with pose* ω_i*,* $i = 1, 2, \ldots, n$*;* Δ_0 *is the rest of the data space and contains the background data.*

This definition makes sense for a generic visual representation. In fact, it is the characteristic nature of visual-type data that objects in the real world map onto segments in data space in a *non-mixing* manner, as long as we do not deal with transparency in 2D images; see Fig. 1 for an illustration. In other words, each point in data space can, in principle, be assigned to a unique source. This is in contrast to, e.g., auditory data. Also excluded are representations of visual data that discard positional information, as in Srivastava et al. (2002) and Wahl et al. (2003).

The probability of obtaining the data D given a segmentation σ of the data space and object parameters Ω thus is

$$p(D|\sigma, \Omega) = \frac{N!}{\prod_{i=0}^{n} |D \cap \Delta_i|!} \prod_{i=0}^{n} \left[p_i(D \cap \Delta_i|\Omega)\, p_i(\Omega)^{|D \cap \Delta_i|} \right] \ , \tag{2}$$

with

$$\sum_{i=0}^{n} p_i(\Omega) = 1 \ . \tag{3}$$

Here $p_i(\Omega)$ is the probability for a data point to originate from object $i = 1, 2, \ldots, n$; $p_0(\Omega)$ is the probability for a data point to not originate from any of the objects, i.e., to originate from the background $i = 0$; $D \cap \Delta_i$ are the data points in segment Δ_i and $|D \cap \Delta_i|$ is their number; $p_i(D \cap \Delta_i|\Omega)$

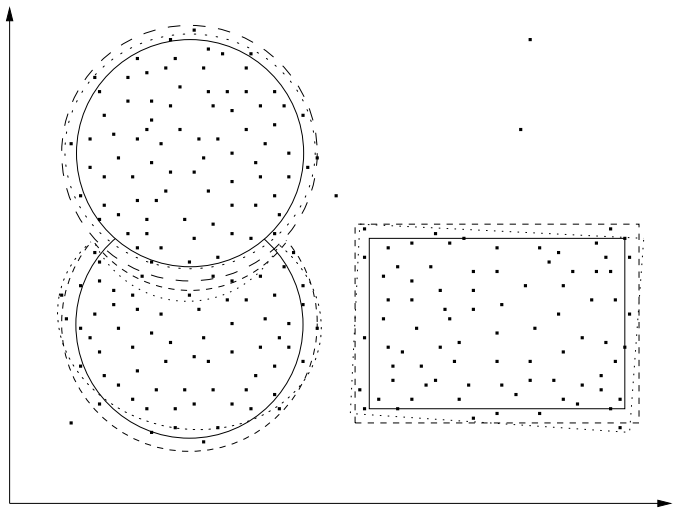

Figure 1: Illustration of segmentations of the data space. Three objects produce, by some visual processing, a number of data points (dots) in a 2D data space, which may be the image plane. Given the object poses, their possible segments (dashed and dotted lines for two examples) in the data space are close to ideal images of the object geometry (solid lines) under the visual mapping, taking occlusions into account.

is the probability of the data $D \cap \Delta_i$ given that they originate from object $i = 1, 2, \ldots, n$ or from the background $i = 0$, respectively. Note that, because of possible occlusions, the probabilities of object and background data are generally conditioned on all object parameters Ω. Equation (2) describes the generative model of the data D for a given number N of data points and for n given objects with parameters Ω and data-space segments σ.

The probability of a scene interpretation Ω given the data D is

$$p(\Omega|D) = \sum_{\sigma} p(\sigma, \Omega|D) = \sum_{\sigma} p(\sigma|\Omega)\, p(\Omega)\, p(D|\sigma, \Omega) \Big/ p(D) \ , \qquad (4)$$

where the sum is running over all possible segmentations σ of Δ. It is well known that maximizing this probability with respect to Ω yields the interpretation with the lowest probability of error (Bayes error). The prior probability $p(\Omega)$ of pose parameters of the n objects is zero for inconsistent combinations of object poses, that is, poses that would lead to intersections of objects. Henceforth, we will understand the poses Ω as being constrained to the set of consistent interpretations. The prior probability $p(D)$ of the data does not enter in maximizing $p(\Omega|D)$ with respect to Ω.

Because of high-order correlations between data points, however, it is usually impossible to obtain, store, or computationally handle the joint probabilities $p_i(D \cap \Delta_i|\Omega)$ of object and background data. Therefore, it is worthwhile to see whether we can get along with less knowledge. Indeed, we here want to remain agnostic as to the probability of distributions of the data over the n objects.

Let us now assume that the background probability is very much smaller than the object probabilities[1], i.e.,

$$p_0(\Omega) \ll p_i(\Omega) \quad \text{for } i = 1, 2, \ldots, n \text{ and all } \Omega. \tag{5}$$

In such cases, probable scene interpretations will be strongly constrained by assigning only a small fraction of the data to the background. To see this more clearly, we factor the generative model (2) as

$$\begin{aligned} p(D|\sigma,\Omega) \;\; = \;\; & \frac{N!}{|D\cap\Delta_0|!\,(N-|D\cap\Delta_0|)!}\, p_0(\Omega)^{|D\cap\Delta_0|}\, [1-p_0(\Omega)]^{N-|D\cap\Delta_0|} \\ & \times \frac{(N-|D\cap\Delta_0|)!}{[1-p_0(\Omega)]^{N-|D\cap\Delta_0|}}\, \frac{\prod_{i=1}^n p_i(\Omega)^{|D\cap\Delta_i|}}{\prod_{i=1}^n |D\cap\Delta_i|!} \\ & \times \prod_{i=0}^{n} p_i(D\cap\Delta_i|\Omega)\ . \end{aligned} \tag{6}$$

The factor on the first line is the probability of obtaining $|D \cap \Delta_0|$ points from the background; the one on the second line essentially means that very unequal distributions of data over the n objects are unprobable; the one on the third line describes the probability of data distributions within each object and the background. As $p_0(\Omega) \to 0$, the first factor becomes zero for every distribution of data D with a non-vanishing background component; see Fig. 2. If we are lucky, the assignment of the minimal number of data points to the background is quite unique such that the second and third factors may not effectively constrain scene interpretation in terms of Ω much further; cf. Fig. 1. In particular, we may then ignore the complicated probabilities $p_i(D \cap \Delta_i|\Omega)$ – and we need to do just this to arrive at a fuzzy-set formulation below.

Plugging only the first line of Eq. (6) into Eq. (4), we are left with

[1] Note that the reverse situation is analogous, with the role of data points being taken by data 'holes'.

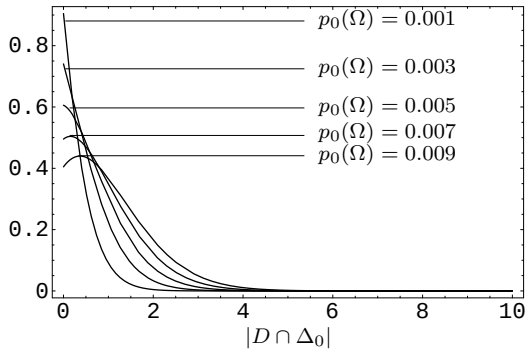

Figure 2: Plots of the probability of the number $|D \cap \Delta_0|$ of background-data points [binomial distribution; cf. first line of Eq. (6)]. The total number of data points is $N = 100$, the background probabilities $p_0(\Omega)$ are as indicated. The curves are interpolated for each fixed $p_0(\Omega)$ for graphical clarity.

maximizing

$$p(\Omega|D) \approx \sum_{\sigma} p(\sigma|\Omega) \times \frac{N!}{|D \cap \Delta_0|!\,(N - |D \cap \Delta_0|)!}\, p_0(\Omega)^{|D \cap \Delta_0|}\, [1 - p_0(\Omega)]^{N - |D \cap \Delta_0|} \quad (7)$$

with respect to the scene interpretation Ω. It is now time to take a closer look at the conditional probability $p(\sigma|\Omega)$ of segmentations σ. For given pose parameters Ω, rigid objects have a certain known extension in physical space and therefore, by visual mapping, a rather sharp extension in data space[2]; cf. Fig. 1. In other words, there is only a small uncertainty as to the segment in data space where data points may map that originate from a known object. Hence, the number $|D \cap \Delta_0|$ of background-data points will usually vary only little, compared to the total number of data points, over the support of $p(\sigma|\Omega)$. If this is true, maximizing (7) is roughly equivalent to minimizing the mean number

$$\langle |D \cap \Delta_0| \rangle_{\Omega} = \sum_{\sigma} p(\sigma|\Omega)\, |D \cap \Delta_0| \quad (8)$$

of background-data points, or to maximizing the mean number

$$M(\Omega, D) := N - \langle |D \cap \Delta_0| \rangle_{\Omega} \quad (9)$$

[2]For non-rigid objects, Ω has to be extended to include the additional degrees of freedom.

of object-data points. We have thus arrived at the problem of maximizing the amount of data that are explained in terms of the objects.

2.2 Explaining Data with Fuzzy Sets

We need to calculate the mean number $M(\Omega, D)$ of object-data points as a function of the scene interpretation Ω. Let $p(m|\Omega, D)$ be the probability for m data points from D to originate from the objects with parameters Ω, hence,

$$M(\Omega, D) = \sum_{m=0}^{N} p(m|\Omega, D)\, m \ . \tag{10}$$

In terms of probabilities of segmentations, the probability of having m object-data points is

$$p(m|\Omega, D) = \sum_{\sigma} p(\sigma|\Omega)\, \delta_{|\Delta_0|, N-m} \ , \tag{11}$$

where, as usually, $\delta_{i,j}$ denotes the Kronecker delta, i.e., the components of the (here N-dimensional) unit matrix.

There is an independence property for the data points that will often hold to a very good approximation. Knowing the positions and types of the objects is usually sufficient for estimating for each data point *individually* how probably it originates from one of the objects; what is the case for other data points does not have a strong influence. In other words, the information provided by the object types and positions on the causal origin of a given data point is much larger than the information provided by causal origins of other data points. Let $\mu(\Omega, d)$ be the probability for a data point $d \in D$ to originate from any of the objects with parameters Ω. We may then formally define as follows.

Definition 2: *The conditional* independence of origin *for data points D is the property*

$$p(m|\Omega, D) = \\ \mu(\Omega, d)\, p(m-1|\Omega, D\backslash\{d\}) + [1 - \mu(\Omega, d)]\ p(m|\Omega, D\backslash\{d\}) \ , \tag{12}$$

for all $d \in D$.

Let us compare the independence of origin as an assumption for our data points to a different, widely employed independence assumption. It is usually impossible to obtain, store, or computationally handle the joint probabilities $p_i(D \cap \Delta_i|\Omega)$, i.e., the models of the object data, for many

data points. The common procedure then is to assume the data points for each given object to be conditionally independent, i.e., $p_i(D \cap \Delta_i|\Omega) = \prod_{d\in D\cap\Delta_i} p_i(d|\Omega)$. How dramatically wrong this assumption generally is, however, becomes obvious by considering an extreme case. A distribution of $|D\cap\Delta_i|$ data points that all lie at the maximum of the one-point probability $p_i(d|\Omega)$ is taken as being more probable – often *much* more probable – than the most probable distribution according to the real probability $p_i(D\cap\Delta_i|\Omega)$. What makes the independence assumption (12) a lot more plausible than the latter is the fact that the origin of a data point as introduced here is just a binary alternative – a data point either originates from the objects or the background – with the set D of data points appearing in the *condition* rather than in the space of alternatives. Phrased shortly, we have to decide on much less given much more information.

Situations in which the property (12) does not hold include those when the method for data acquisition produces groups of data points rather than individual points. In that case, it may be a better choice to describe those groups as a single point in a different data space.

We are now prepared to formulate the final step towards fuzzy-set-based reasoning.

Assertion: *If independence of origin in the sense of Eq. (12) holds for the data, the mean number of object-data points is given by*

$$M(\Omega, D) = \sum_{i=1}^{n} \sum_{j=1}^{N} \mu_i(\Omega, d_j) \ , \tag{13}$$

where $\mu_i(\Omega, d_j)$ denotes the probability for data point d_j to originate from object i.

Proof: Substituting Eq. (12) for $d = d_1$ into Eq. (10) and remapping the summation index of the first sum ($m - 1 \mapsto m$) yields

$$\begin{aligned} M(\Omega, D) &= \mu(\Omega, d_1) \sum_{m=0}^{N-1} p(m|\Omega, D\backslash\{d_1\})\,(m+1) \\ &\quad + [1 - \mu(\Omega, d_1)] \sum_{m=0}^{N-1} p(m|\Omega, D\backslash\{d_1\})\, m \\ &= \mu(\Omega, d_1) + \sum_{m=0}^{N-1} p(m|\Omega, D\backslash\{d_1\})\, m \ . \end{aligned} \tag{14}$$

Comparing to Eq. (10), we see that the remaining summation term is just the original mean value with the set D of data points reduced to $D\backslash\{d_1\}$.

By induction it is now evident that the whole expression is simply

$$M(\Omega, D) = \sum_{j=1}^{N} \mu(\Omega, d_j) \ . \tag{15}$$

Let $\mu_i(\Omega, d)$ be the probability for a data point $d \in D$ to originate from object i. Because of possible occlusion by other objects, this probability generally depends on all the object parameters Ω. The probability $\mu(\Omega, d_j)$ that d_j originates from any of the n objects [cf. Eq. (15)] can be written as

$$\mu(\Omega, d_j) = \sum_{i=1}^{n} \mu_i(\Omega, d_j) \prod_{\substack{k=1 \\ k \neq i}}^{n} [1 - \mu_k(\Omega, d_j)] \ . \tag{16}$$

It is now important to realize that products $\mu_i(\Omega, d_j)\, \mu_k(\Omega, d_j) \ldots$ with $i \neq k$ are negligible: given a consistent interpretation Ω, at most one object can have caused each data point with a significant probability, neglecting a small fraction of the data that may fall suspiciously close to more than one object image. This is again a consequence of the visual mapping process; cf. Fig. 1. Thus the expression (16) simplifies to

$$\mu(\Omega, d_j) \approx \sum_{i=1}^{n} \mu_i(\Omega, d_j) \ . \tag{17}$$

Substituting Eq. (17) into Eq. (15), we arrive at Eq. (13). q.e.d.

The functions $\mu_i(\Omega, d)$ $(i = 1, 2, \ldots, n)$ are similar in spirit to the *membership functions* of fuzzy logic, albeit with a well-defined probabilistic meaning and rationale. They effectively describe the object segments Δ_i as fuzzy subsets of the data space Δ, weighting a location $d \in \Delta$ by the probability for a given data point at d to originate from object i in a scene Ω. On a more abstract level, this probabilistic interpretation of a fuzzy set is similar to what has been proposed by Mabuchi (1992, 1997). Equation (13) can now be seen to describe the cardinality of the intersection of a union of fuzzy sets $M_i(\Omega)$, one for each object sought in the scene, with the data set D,

$$M(\Omega, D) = |[\cup_{i=1}^{n} M_i(\Omega)] \cap D| \ . \tag{18}$$

The condition for consistent interpretations Ω reads

$$M_i(\Omega) \cap M_j(\Omega) = \emptyset \quad \text{for } i \neq j, \tag{19}$$

in the fuzzy-set formulation.

It turns out that we can calculate the mean number $M(\Omega, D)$ of data points that are explained through a scene interpretation Ω by evaluating Eq. (13) or, equivalently, Eq. (18), provided independence of origin holds in the sense of Eq. (12). Under some further conditions as discussed in Sec. 2.1, visual scene interpretation may then proceed by maximizing the function (13) or (18) with respect to the object poses Ω under the consistency constraint. The best interpretation of a scene is hence the one that maximizes the overlap of fuzzy object sets with the data.

3 Practical Issues

The great advantage of maximizing expression (13) rather than (4) with respect to Ω is that (13) does not depend on the complicated joined probabilities $p_i(D \cap \Delta_i|\Omega)$ of many data points $D \cap \Delta_i$. The question now arises, how to obtain the probabilistic membership functions $\mu_i(\Omega, d)$ for the objects $i = 1, 2, \ldots, n$. Their relation to the one-point probabilities $p_i(d|\Omega)$ of object data is

$$\mu_i(\Omega, d) = \frac{p_i(d|\Omega)\, p_i(\Omega)}{p_i(d|\Omega)\, p_i(\Omega) + p_0(d|\Omega)\, p_0(\Omega)} \ . \tag{20}$$

Taking into account our fundamental assumption (5) and neglecting the spatial variation of the background-data probability, we can write

$$\mu_i(\Omega, d) = \frac{p_i(d|\Omega)}{p_i(d|\Omega) + \epsilon_i} \ , \tag{21}$$

where $0 < \epsilon_i \ll 1$ is a constant that depends on the amount of background data. This parameter has to be adjusted for best performance, although its exact value will usually not be critical. For a vanishing background, $\epsilon_i \to 0$ and the fuzzy object sets become crisp sets with membership functions

$$\mu_i(\Omega, d) = \begin{cases} 1 & \text{if } p_i(d|\Omega) > 0, \\ 0 & \text{else.} \end{cases} \tag{22}$$

The one-point probabilities $p_i(d|\Omega)$ of object data, in turn, can be learned as histograms over adequate regions of data space by straightforward sampling of data points.

In practice, the dependence on the object poses Ω will often be reasonably approximated by taking only rotations of object i itself into account and perhaps setting to zero regions that are roughly occluded under each

hypothetical scene interpretation Ω. The membership functions $\mu_i(\Omega, d)$ are then object-view templates that can be efficiently translated in data space by, e.g., some variant of generalized Hough transform (Ballard, 1981; Samal & Edwards, 1997; Bonnet, 2002).

The most severe restriction on using the fuzzy-set approach to scene interpretation lies in the condition (5) on the background probability $p_0(\Omega)$: the smaller $p_0(\Omega)$, the more reliable the fuzzy-set result will be. Obviously, the only way to ensuring this, if possible, is to select features that are rather restricted to the objects sought. For instance, if one wants to analyze a scene of objects standing on a plane table, it is a good idea to use geometric features that are somehow related to high surface curvature.

Up to now, we have kept fixed the number n of objects employed in interpreting a scene. It is clear, however, that more data may be explained by more objects. If we let the number of object models grow purely based on the amount of explained data, we will almost certainly run into an unrealistic proliferation of objects in the interpretation. In fact, this is just the classic case of over-fitting data with a too complex model and, as always, this needs regularization. In order to not assign an object to noise and artifacts not accounted for in the modeled membership functions, we have to evaluate what is gained by each object we may include in an interpretation. An object should hence be included only if the number of data points it additionally explains exceeds an object-specific threshold.

4 An Application

For an example application of the discussed principles, take a look at Fig. 3 and Hillenbrand et al. (2004). The visual task was to analyze table-top scenes of bottles and glasses for an experimental service robot to manipulate the objects in a sensible way. In particular, the robot should unscrew the cap from a bottle, grasp the bottle, and pour drinks into appropriate glasses. Hence, the objects had to be localized and identified by the visual system.

The task was accomplished by matching probabilistic fuzzy sets to stereo-data points of the scene, that is, by maximizing (13) with respect to the positions $\Omega = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$ of n objects on the table. The number n of objects was determined from a threshold criterion as described in Sec. 3. For maximum simplicity, only one model $\mu_i(\omega_i, d)$, with $\omega_i = (x, y) = \text{const.}$, was sampled from training data for each object i. For computing a scene interpretation, these object models were translated in the discrete x-y-space in the style of a generalized Hough transform (Ballard, 1981; Samal & Ed-

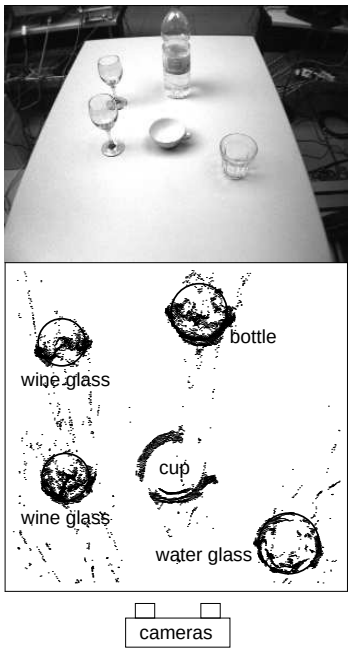

Figure 3: Example scene, its stereo data, and its interpretation. The scene features a bottle, two wine glasses (one partially occluding the other), a water glass, and a cup on a table, as seen in the top image taken by one of the cameras. The bottom image shows the stereo-data points, viewed directly from top, superimposed with the result of the scene interpretation. The circles outline the geometric extension of the objects as detected by fuzzy-set matching. Data points within or close to an object border are assigned to that object. The system did not have a model for cups, so the cup data are ignored. The total processing time for this scene was around 5.7 seconds on a Pentium 4 CPU at 2.4 GHz.

wards, 1997; Bonnet, 2002) to find their best match to the data.

Stereo data were computed by a straightforward correspondence search between three camera views, based on minimizing the sum of absolute differences over square patches of LoG-filtered (Laplacian-of-Gaussian filtered) images [the Triclops SDK, Point Grey Research Inc.; see www.ptgrey.com and Okutomi & Kanade (1993)]. We set a high threshold for valid regions of the LoG-filtered images, such that only regions of high scene contrast were taken into account. As a result, mainly surface creases, sharp bends, and depth discontinuities contributed any data points. The output from stereo processing thus was a sparse representation of the scene by rather few 3D points (around 20000), outlining the objects on the table. This kind of representation ensures the low background probability required for the fuzzy-set approach.

In agreement with the presented arguments, the resulting performance was quite robust for challenging objects such as glasses under partial occlusion and over large variations of lighting (daylight, fluorescent light, spotlight). This robustness derives partly from the fact that probabilistic fuzzy sets do not imply strong assumptions about the distribution of data points.

5 Conclusion

In this article, I have shown that probabilistic inference of a visual scene interpretation can be reduced to optimizing the match of probabilistic fuzzy sets to the data, if two critical conditions are satisfied. First, the amount of background data is a sufficiently informative quantity. This will be true, if the prior probability for a data point to originate from the background is small enough. Second, given data points originate from given objects or the background in a statistically independent fashion. This will usually hold to a good approximation, if the imaging procedure does not systematically produce groups of data points. Otherwise, a representation of such groups as single points in a different data space may be a solution.

It is important to realize that the independence of origin required for data points is a more plausible assumption than statistical independence of the data points as such for given objects. In this regard, the probabilistic fuzzy-set approach is less presumptuous than many other statistically inspired methods for visual scene analysis.

Some more assumptions have been made along the way from probabilistic inference to fuzzy-set matching. However, they can be regarded as approximately valid for the problem of visual scene analysis that I have been considering. They are, in fact, closely related to the nature of visual data, that is, data produced by some kind of one-to-one mapping of points from physical space into a 2D or 3D data space. The arguments presented here are, therefore, of a genuine visual character and cover only a very specific use of fuzzy techniques. Theoretical justification for a more general application of fuzzy logic would still be required.

References

Ballard, D. H., 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111–122.

Bonnet, N., 2002. An unsupervised generalized Hough transform for natural shapes. *Pattern Recognition*, 35:1193–1196.

Hillenbrand, U., Brunner, B., Borst, Ch., and Hirzinger, G., 2004. The Robutler: a vision-controlled hand-arm system for manipulating bottles and glasses. In *Proc. Intern. Symp. on Robotics*.

Hisdal, E., 1988. Are grades of membership probabilities? *Fuzzy Sets and Systems*, 25:325–348.

Keller, J. M., 1997. Fuzzy set theory in computer vision: a prospectus. *Fuzzy Sets and Systems*, 90:177–182.

Mabuchi, S., 1992. An interpretation of membership functions and the properties of general probabilistic operators as fuzzy set operators – Part I: case of type 1 fuzzy sets. *Fuzzy Sets and Systems*, 49:271–283.

Mabuchi, S., 1997. Supplement to "An interpretation of membership functions and the properties of general probabilistic operators as fuzzy set operators – Part I". *Fuzzy Sets and Systems*, 89:69–76.

Okutomi, M. and Kanade, T., 1993. A multiple-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:353–363.

Samal, A. and Edwards, J., 1997. Generalized Hough transform for natural shapes. *Pattern Recognition Letters*, 18:473–480.

Srivastava, A., Liu, X., and Grenander, U., 2002. Analytical image models and their applications. In *Proc. Europ. Conf. Computer Vision*, Lecture Notes in Computer Science, volume 2350, pages 37–51.

Tizhoosh, H. R., 1998. Fuzzy image processing: potentials and state of the art. In *Proc. Intern. Conf. Soft Computing*, volume 1, pages 321–324.

Wahl, E., Hillenbrand, U., and Hirzinger, G., 2003. Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification. In *Proc. Intern. Conf. 3-D Digital Imaging and Modeling*, pages 474–481.

Walker, E. L., 1998. Perspectives on fuzzy systems in computer vision. In *Proc. Ann. Conf. North American Fuzzy Information Processing Society*, pages 296–300.

Zadeh, L. A., 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28.