INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Fuzzy data for video image object tracking

F.A.S. Dias *N.J. Leite*

Technical Report - IC-05-024 - Relatório Técnico

September - 2005 - Setembro

Fuzzy data for video image object tracking

Fábio Augusto Salve Dias* Neucimar Jerônimo Leite†

Abstract

TIn this paper, we introduce a new method for object tracking which takes into account multiple target features and fuzzy logic for knowledge representation and information fusion. The knowledge about the target location is represented by fuzzy matrices, one matrix for each considered feature. Since the set of all included information belongs to the same domain, the result indicating this final location is given by a fusion of matrices, through fuzzy operators, which expresses the combination of all defined target features.

1 Introduction

Nowadays, the use of digital video images to store and/or analyse events is becoming very common. Locate a target in a video image has applications in many knowledge areas such as biological experiments, surveillance, human-machine interaction, and in any event in which the considered information can be visually observed. In this context, automatic tracking is very useful, offering a faster and more precise tool than manual approaches.

The results of tracking methods is highly dependent on the feature space used to model the target. The main idea of this work takes into account the use of various feature spaces to improve tracking results.

The most common feature spaces for tracking methods are color [18, 7, 4, 10], shape [20, 14, 24, 13], spatial relationship [16, 11], and motion [26, 2]. Most of these methods are based mainly on *one* of these features, thus ignoring other significant properties of the analysed scene. Therefore, the choice of a feature space is problem dependent, leading to the definition of constrained methods for tracking which often fail when the characteristics of the problem change slightly.

Further, these methods do not deal directly with aspects concerning uncertainties, inherent to any physical measurement. To reduce noise effect, most of the approaches preprocess the image taking into account linear or non-linear filtering techniques, for instance.

This work considers the problem of transforming the maximum available information about a scene into fuzzy matrices with the same dimensions as the original scene. These matrices can be seen as grayscale images representing the result of a feature comparison,

*Institute of Computing, University of Campinas, 13081-970 Campinas, SP. Research supported in part by FAPESP grant $04/02541-0$

†Institute of Computing, University of Campinas, 13081-970 Campinas, SP.

maped into a fuzzy domain in the range $[0, 1]$, where 1 indicates the best possible match. Fuzzy representations of some image descriptors, such as histogram, motion and shape, are used here to handle measures and significance uncertainties in a tracking procedure. As we will illustrate elsewhere, this approach yields a method easily adaptable to almost any concerned environment, in which the more complicated or unconstrained is the problem the more information about the scene will be requested.

The rest of this paper is organized as follows: Section 2 gives some brief concepts on fuzzy logic and introduces a new fuzzy histogram used in our image color distribution analysis. Section 3 describes a tracking model which is based on multiple feature spaces representation and data fusion. Section 4 illustrates some experimental results and, finally, some conclusions are drawn in Section 5.

2 Fuzzy Logic

Fuzzy logic was introduced in 1965 by Zadeh [25] as a more general alternative to Boolean logic. In Boolean logic, the available data are only *true* and *false*, often represented by *1* and *0*, respectively. This information does not allow any intermediary values in this range. In fuzzy logic, *all real* values between 0 and 1 are allowed, providing a flexible way to represent uncertain or incomplete knowledge.

2.1 Operations with fuzzy numbers

With a greater domain, fuzzy logic also provides a larger number of possible operations. In this work, we considered two basic classes of these operations [19].

- **t-norms:** an example of a t-norm, representing a more restrictive fuzzy operation, is the *and* operator which computes the minimum between two fuzzy values.
- **t-conorms:** an example of a t-conorm, representing a less restrictive fuzzy operation, is the *or* operator which computes the maximum between two fuzzy values.

2.2 Fuzzy Histogram

In this section, we consider the concept of fuzzy histograms used in the characterization of an image content, representation of uncertainties and reduction of quantization problems [9, 22, 21, 5].

The histogram is largely used to represent the color or intensity distribution of an image pixels. In [7], Comaniciu et al. use this information as the main feature of a tracking procedure. The matching of the target in a scene is done by searching the minimum Bhattacharyya distance between two histograms.

The common histogram representation does not deal with uncertainties, like noise. Usually, this problem can be solved by reducing the number of bins per histogram dimension. This solution leads to a coarse quantized information in which colors with different meanings (distant in the color-space representation) are often merged.

In this work, we consider each color as a measure of a Gaussian probability density with a mean value and a variance given as parameter. This variance, which can also be computed by statistical analysis of the input image, should reflect, for instance, measurement errors related to the camera or other used sensors.

Based on this model, a group of bins of the fuzzy histogram is updated by considering the Gaussian distribution function which models how much each bin changes, i.e., the distribution values represent *how much* a given data could be of each bin, thus acting as a fuzzy *membership function*.

Most of the fuzzy histograms described in literature [9, 22, 21] consider uncertainties by modeling the data as a triangular fuzzy membership function. The other basic operations, such as quantization, are identical to the common ones applied to the distribution data.

Unlike the work in [5], which also models the input data as a Gaussian function and defines a fuzzy aggregational quantized histogram whose query operations are given by a t-conorm of the bin values after the histogram quantization. In our model, the returned value of a query is the summation of all the distribution data associated with a histogram range of bins, thus defining a quantized histogram representation, centered at the corresponding queried data. This range is defined as an input parameter but can also be changed dynamically, yielding a multi-resolution fuzzy histogram model.

3 Video image tracking

This section introduces a new approach for video image tracking based on multiple feature spaces. Figure 1 shows a flowchart of this system model. The main steps of this approach are:

- **Target Definition:** The user provides a binary image representing a mask of the target in a given frame.
- **Feature extraction:** Each feature space of the method should have a corresponding model to be compared with. This step extracts and stores information about the interest target .
- **Specific feature analysis:** Each feature space analysis operation processes the original frame and returns a fuzzy matrix. A locus in this matrix represents one pixel in the original frame and its value corresponds to a similarity measure between the target and the considered region in the respective frame.
- **Data Fusion:** All information obtained from the specific feature analysis are merged into one final data fusion matrix to define the actual target location.

The next section discusses some aspects of the above last four steps.

3.1 Feature extraction

After definition of the target mask from the first frame of a sequence, this step in Figure 1 extracts specific information about the target, thus defining its model which will

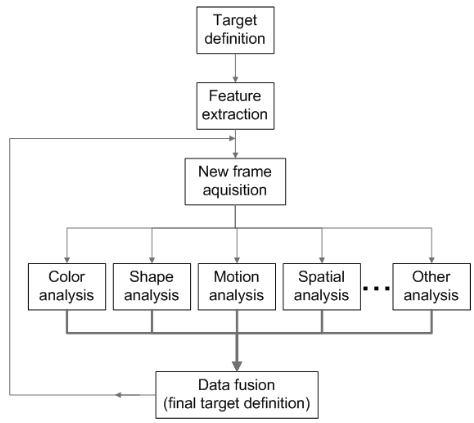


Figure 1: General flowchart of the proposed method.

be considered in the specific feature analysis step. Our system stores, for example, the fuzzy histogram representation (Section 2.2 of the target region and the Fourier descriptors of its shape. Although the target model can be modified by an updating schema along the sequence processing, this information was kept unchanged in this version of the work.

3.2 Specific feature analysis

This operation in Figure 1 is such that, given the original image and the appropriate target model representation, it returns a fuzzy matrix representing the result of their matching for each image pixel. Also, the methods for these features analyses should be independent from each other so that the deletion or insertion of a new feature space can be simplified. This also facilitates parallel implementations of the system model.

Initially, we consider the following feature spaces:

- **Color analysis:** By using the target fuzzy histogram built in the above step, we define a feature matrix according to equation 1. In this equation, I represents the original image, F the resulting fuzzy matrix, (x, y) are the image/matrix coordinates, and function $H(C)$ is the result of a query of the model fuzzy histogram based on a given color C.

$$F(x, y) = H(I(x, y)) \tag{1}$$

  Figure 2 shows an example of this color analysis.

  Note that this approach supports others chromaticity spaces, like xy, L*u*v*, HSV, etc. Furthermore, due to the method generality and scalability, its possible to have *many* of them combined simultaneously.

- **Shape analysis:** Shape is one of the main features included in many tracking systems. It is usually associated with background subtraction [18, 10, 23, 12] or other segmentation techniques, such as Watershed [3] or Mean Shift algorithm [6]. Here, we

(a) Original frame. The target is the left-most ball of the image.

(b) Color analysis result (the clearer the regions the higher the fuzzy values).

Figure 2: Color feature analysis example.

considered a well-known watershed-based segmentation method followed by a post-processing step to reduce oversegmentation effects [3], [15]. Note that any other segmentation procedure can be employed in this part of the work.

After proper segmentation, we compare the obtained image regions with the target model. This target can have more than one homogeneous region and its model should be given by a set of appropriate descriptors related to each of these regions. In this case, *any* shape descriptor comparison returning values into a finite range of real numbers can be used. Initially, we compare the Fourier descriptors of the shapes which are rotation invariant, and, with proper modifications, scale invariant [17], [8].

To compare sets of these descriptors, we use a cross-correlation function which takes two vectors as input and returns a value between -1 and 1, representing the degree of correlation (a zero value indicates no correlation, -1 and 1 represent inverse and direct correlation, respectively).

Let S be the set of Fourier descriptors of a target, S_i the vector of the Fourier descriptor of index i, R the vector of Fourier descriptors of a candidate region, and CC a cross-correlation function. The fuzzy value F associated with the candidate region is given by:

$$F = max|(CC(S_i, R))|, \forall S_i \in S \tag{2}$$

The resulting matrix based on this specific feature is defined by applying the above to all the segmented regions in the frame, and assigning the normalized value F to the corresponding region pixels.

By taking into account the zero-frequency terms of the Fourier descriptors, we obtain the size of the candidate regions which can be related to the target size by a normalized value assigned to each of these regions. This information constitutes a new shape feature analysis of the system model in Figure 1. Figure 3 shows an example of these operations for the frame in figure 2(a).

(a) Grayscale image representing the shape analysis fuzzy matrix for the image in figure 2(a) (the clearer the regions the higher the fuzzy values).

(b) Grayscale image representing the fuzzy matrix for the size of the image regions in figure 2(a).

Figure 3: Shape feature analysis example.

- **Motion analysis:** In many tracking problems, the original scene is recorded by still cameras, often under a constrained environment. In such a case, definitions of background subtraction and blob extraction techniques are very usual. Our system includes a *motion*-based feature analysis which can improve the tracking results of this kind of solution.

  This analysis takes into account differences between two consecutive frames, thus yielding non-zero valued pixels only on areas of moving objects. Note that the obtained difference image may not exactly define the target position since, depending on the processing frequency, a moving object may have significant values only around its boundaries (w.r.t the previous frame). To represent possible target locations in the current frame, we make a fusion of this information with the target location in the previous frame.

  To do this, we make the following considerations:

  - A zero-valued pixel obtained from the difference between the current and the previous frame represents the same information in both frames.
  - A non-zero-valued pixel in the difference image, not related to the target area in the previous frame, can be seen as a pixel to which the target is currently moving on.
  - A pixel with non-zero value in the difference image, associated with the target area in the previous frame, can be seen as a pixel from which this target is moving on.

Thus, the specific motion feature analysis matrix can be given by:

$$m(i,j) = |T_{previous}(i,j) - D(i,j)|, \forall (i,j) \in m, \tag{3}$$

where m is the resulting matrix, $T_{previous}$ is the fuzzy matrix representing the target location in the previous frame, and D is the normalized matrix corresponding to the difference between the previous and the current frames.

This analysis is very useful in cases of a scene containing still objects very similar to the moving target, as in the image in Figure 2(a). Figure 4 shows the result of such a feature analysis for this image.

Although the specific consideration of this analysis to still video cameras, this step does not compromise the performance of the whole method if non-stationary cameras are used. In such a case, the resulting matrix will have a few zero-valued locations whose feature contribution can be negligible depending on the employed fuzzy operator.

(a) Difference between two consecutive frames.

(b) The motion fuzzy matrix indicating the moving tennis ball location.

Figure 4: Motion analysis result for the image in figure 2(a).

- **Spatial analysis:** A simple spatial feature considered in this work is given by the Euclidean distance between the center of mass of a candidate region and the center of mass of the target in the previous frame. This normalized distance is complemented and assigned to each one of the segmented regions.

3.3 Data Fusion

To obtain the final matrix containing all relevant information about the actual frame and the target model, the specific feature analysis matrices should be combined into one according to some defined procedures. This data fusion is not difficult to accomplish since all the matrices values belong to the same domain, i.e., fuzzy numbers.

The choice of this fusion method should reflect certain characteristics of the video scene. For example, if a video contains many similar rigid-body objects, as in Figure 2(a), a more restrictive fusion should be used (e.g., based on a t-norm). This choice implies the definition of the final target area as the one having good matches in *all* analysed features. On the other hand, if the video sequence contains a highly deformable or variable color target, with very different objects, a less restrictive fusion operator should be used, for instance, a t-conorm. This choice implies the definition of the target area based on the image regions with good matches in at least *one* feature space.

It is easy to see that the t-conorm will not be very useful here since, to generate good results, the final matrix should have as *less* as possible non-zero regions conveying information about target locations. A t-conorm will return a matrix with few non-zero regions only if all other objects in the scene have different characteristics w.r.t the target. Figure 5 shows the data fusion by an *and* operator of all the feature matrices related to the frame in figure 2(a).

Figure 5: Result of the data fusion by an *and* fuzzy operator.

4 Other examples and results

Figures 6-8 shows some examples of tracking for a real outdoor scene having two persons wearing red t-shirts. In such a case, the tracked object is the leftmost person t-shirt. Note that, though we considered simple feature analysis methods in this work, the combination of different characteristics representing the target allowed a proper definition of its location along the analysed sequence.

All the videos in this work were processed by considering the RGB color space and a fuzzy histogram (Section 2.2) with 32 bins per dimension. For such a histogram, the variance of the considered distribution was 16 and the query operation was defined in a $13\times13\times13$ neighborhood of the queried data.

Figures 9 and 10 compare some aspects of our approach with a well-known color-based method, named CAMSHIFT [4]. This method takes into account the probability distribution of the objects during the tracking and is based on the Mean Shift approach [6]. In the used synthetic video, we have a still block and a moving ball whose color changes dynamically along the sequence. Figure 9 indicates the target location in a certain video frame for both methods. The ellipsis in Figures 9(c) and 10(c) are the output of the CAMSHIFT algorithm. Figure 10(c) illustrates how, in some cases, the CAMSHIFT algorithm can misclassify the target location as a function of some changes in the target features.

The CAMSHIFT algorithm is available on the OpenCV [1] environment. Our approach considered the above defined parameters and a fuzzy *or* operation between the motion and color matrices, followed by a fuzzy *and* with the shape analysis result.

(a) Original video frame. (b) Target location.

Figure 6: Method example.

(a) Original video frame. (b) Target location.

Figure 7: Method example.

5 Conclusion

This work introduced a tracking method based on multiple feature analyses and fuzzy knowledge representation. This method is more robust than others considering only one feature space, and is also quite scalable since any specific feature expressed in the considered domain can be easily incorporated to the system.

(a) Original video frame.

(b) Target location.

Figure 8: Method example.

We also developed a fuzzy histogram variation to deal with uncertainties and multi-resolution representations of an image content distribution.

As future works, we intend to improve aspects concerning data fusion, define new specific feature spaces based, for instance, on background mismatch analyses and increase the overall performance of the method, in terms of its high computational cost, by considering a lower processing rate and multi-resolution representations of the data.

Acknowledgments

The authors are grateful to WebMaps, Agroflow, PRONEX - SAI and **FAPESP** for supporting this work through the project number $04/02541 - 0$.

References

[1] Opencv: Open computer vision library, http://sourceforge.net/projects/opencvlibrary/.

[2] S. Avidan. Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1064 – 1072, August 2004.

[3] S. Beucher and F. Meyer. The morphological approach to segmentation: the watershed transformation. In E. Dougherty, editor, *Mathematical morphology in image processing*, volume 34 of *Optical Engineering*, chapter 12, pages 433–481. Marcel Dekker, New York, 1993.

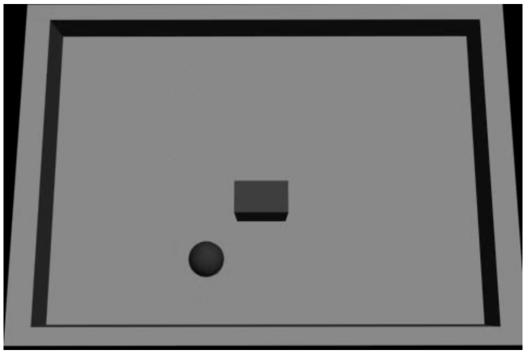
[4] G. Bradski. Real time face and object tracking as a component of a perceptual user interface. In *Fourth IEEE Workshop on Applications of Computer Vision*, pages 214 – 219, Oct 1998.

[5] N. C. Vertan. Embedding fuzzy logic in content based image retrieval. pages 85 – 89. North American Fuzzy Information Processing Society, July 2000.

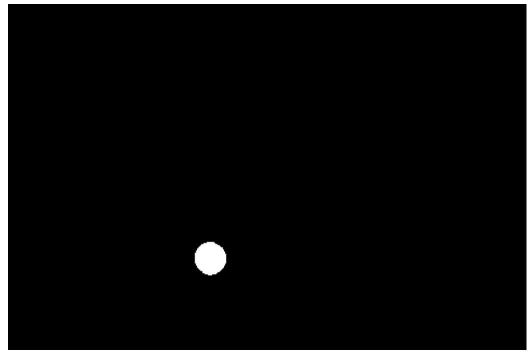
[6] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *ICCV (2)*, pages 1197–1203, 1999.

[7] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564 – 577, May 2003.

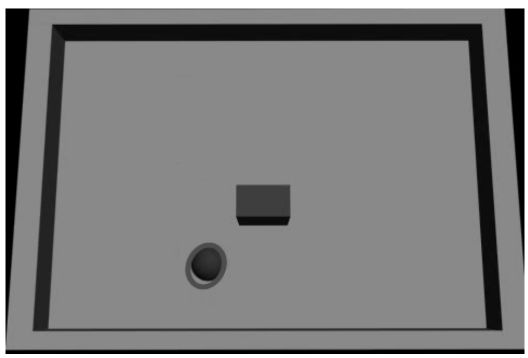
[8] L. da F. Costa and R. M. C. Jr. *Shape Analysis and Classification: Theory and Practice*. CRC Press, 2001.



(a) An original video frame.



(b) Target location according to the proposed method.



(c) Target location, indicated by the ellipsis in the figure, according to the CAMSHIFT algorithm.

Figure 9: Example of tracking by the CAMSHIFT and proposed method (before a collision of the two objects) in which a moving target changes dynamically in color.

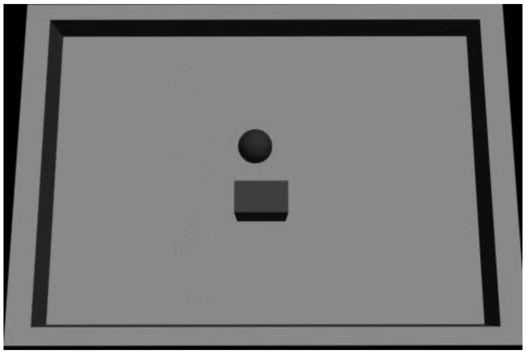
[9] A. Doulamis and N. Doulamis. Fuzzy histograms for efficient visual content representation: Application to content-based image retrieval. In *IEEE International Conference on Multimedia and Expo*, pages 1100–1003, 2001.

[10] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *in Proc. European Conference on Computer Vision*, pages 751–767, 2000.

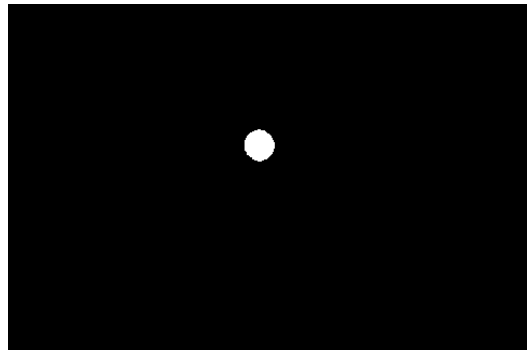
[11] P. Figueroa, N. Leite, and R. M. L. Barros. A flexible software for tracking of markers used in human motion analysis. *Computer Methods and Programs in Biomedicine*, 72:155–165, 2003.

[12] P. J. Figueroa, N. J. Leite, and R. M. L. Barros. Background recovering in outdoor image sequences: an example of soccer players segmentation. *Image and Vision Computing (to appear)*, 2005.

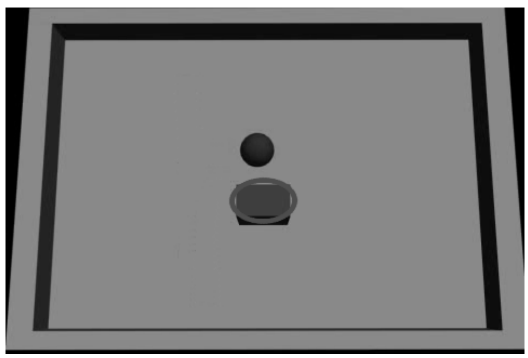
[13] C. Gentile, O. Camps, and M. Sznaier. Segmentation for robust tracking in the presence of severe occlusion. *IEEE Transactions on Image Processing*, 13(2):166–178, February 2004.



(a) Another original video frame.



(b) Target location according to the proposed method.



(c) Target location according to the CAMSHIFT algorithm.

Figure 10: Target location after a collision of the two objects in the scene.

[14] G. Hager and P. Belhumeur. Real-time tracking of image regions with changes in geometry and illumination. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 403–410, June 1996.

[15] K. Haris, S. Efstratiadis, N. Maglaveras, and A. Katsaggelos. Hybrid image segmentation using watershed and fast region merging. *IEEE Trans. Img. Proc.*, 7(12), Dec. 1998.

[16] G. M. I. Cohen. Detecting and tracking moving objects for video surveillance. In *IEEE Proc. Computer Vision and Pattern Recognition*, pages 23–25, Jun 1999.

[17] L. Keyes and A. C. Winstanley. Fourier descriptors as a general classification tool for topographic shapes. Jan. 01 1999.

[18] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *Computer Vision and Image Understanding: CVIU*, 80(1):42–56, 2000.

[19] H. Nguyen and E. Walker. *A First Course on Fuzzy Logic.* CRC Press, 1997.

[20] S. M. Nils T. Siebel. The application of colour filtering to real-time person tracking. In *Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance System*, pages 227–234, 2001.

[21] T. Runkler. Fuzzy histograms and fuzzy chi-squared tests for independence. In *IEEE International Conference on Fuzzy Systems*, pages 1361 – 1366, July 2004.
[22] H. B. S. Siggelkow. Improvement of histogram-based image retrieval and classification. In *16th International Conference on Pattern Recognition*, pages 367–370, Aug. 2002.
[23] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. Computer Vision and Pattern Recognition*, pages 246–252, 1999.
[24] A. Yilmaz, X. Li, and M. Shah. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1531 – 1536, November 2004.
[25] L. A. Zadeh. Fuzzy sets. *Inf. Control*, 8:338–353, 1965.
[26] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208 – 1221, September 2004.