Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp225-230)

Combining Hausdorff Distance, HSV Histogram and Nonextensive Entropy for Object Tracking

PAULO S. RODRIGUES [1] AND GILSON A. GIRALDI [1] AND JASJIT S. SURI [2]

[1]LNCC–National Laboratory for Scientific Computing -
Av. Getulio Vargas, 333, 25651-070 Rio de Janeiro, RJ, Brazil

[2]Biomedical Research Institute, Idaho, ID, USA

Abstract. In Computational Vision, object tracking in a sequence of frames is one of the most important problems. Among the most used approaches there is the model-target one, which matches a model object against a candidate target region in a frame sequence. To accomplish this task, the Hausdorff distance has an attractiveness due to its simplicity of implementation and possibility of matching between two sets with different cardinality. Viewing images as non-extensive systems, we may apply the Tsallis Entropy (which works with only one parameter, called entropic parameter) to segment the frames in order to find the target object. In this work we propose a methodology which combines Hausdorff distance, Bayesian network, HSV histogram and Tsallis non-extensive entropy for objects recognition and tracking in a frame sequence. With this proposal, we reduce the Hausdorff's noise sensitive and the high parameter dependence of the tracking task. We apply our method in experiments with one object over a moving background in a sequence of 300 frames.

1 Introduction

In the context of multimedia applications the detection of objects under changes in illumination, directly affect the object tracking task. A fail due to changes in illumination in the detection of the object of interest may propagate the error through out the remainder frames, imposing a need for human intervention or system reinitialization. In this paper we focus on robust object tracking against changes in illumination conditions [13, 8].

In a simple way, the problem of object tracking can be posed as a correspondence problem between two regions. The general idea is to use a model, called Model-Object (MO), in a frame i and to find the correspondent Target-Object in the frame $i+1$. But, before achieving the correspondence, generally we need a segmentation process which may increase the computational complexity and may yield problems for parameter choice [13]. Among these methods, the one which uses the Tsallis entropy for segmentation is of special interest for our work [1]. It is based on the hypothesis that we can find out a threshold between the background and the foreground through the maximization of the image information (entropy). The main advantage (which we are interested) of this new theory is that it is based on a parameter, called q, which may be handled according to the image characteristics.

In this paper, we firstly segment the objects through the approach based on Tsallis entropy. Then, the task is handled through correspondence between segmented regions (objects in scenes) and a Model Object (MO). In this step we use the Hausdorff distance [13, 8] for region matching and a similarity measure between the HSV Histograms. Our main contribution is the combination of the Hausdorff distance and the Tsallis non-extensive entropy in order to reduce the Hausdorff's noise sensitive and the high parameter dependence of the tracking task. To validate our approach, we experimentally test it under hard illumination conditions.

The paper is organized as follows. The sections 2 and 3 present related works and the background of the used theory. Then, in section 4 we describe our algorithm which was tested in the experimental results presented in section 5. Finally, we end with conclusions and final comments (section 6).

2 Related Work

In this paper we focus on region-based object tracking by combining Hausdorff distance and Tsallis non-extensive entropy. Recently, H. Xu [13] and collaborators proposed a work for object detection with a combination of Hausdorff distance and segmentation based on Watershed approaches. Their method concerns of three steps. Firstly, the target objects are extracted with the use of only the two first frames. With the extracted object taken as a model, the second step of the method uses the Hausdorff distance for tracking the objects of interest in the remainder frames. The object found in the last frame is used as a new model to find the object in the current one. Then, the model is updated at each frame during the tracking task. This approach has the advantage of letting the system immune against possible error propagation. The final step uses a Watershed algorithm, not for a segmentation as usual, but for upper bounding the object boundary. This strategy avoids the influence of external regions, as well as noises, in the object segmentation. The main disadvantage, outlined by the authors, is that their proposed approach fails to separate regions with low contrast; then, it is sensitive to changes in illumination.

Another approach which uses Hausdorff distance for object tracking is that one proposed by N. Paragios and R. Deriche [8], which minimizes the interframe difference density function.

Recently, we introduce ongoing works [10, 11] using a Bayesian Network for parameter estimation in a framework for object recognition under an augmented reality environment. In these works, we used feature extraction to generate several candidate regions in a frame. Then, the proposed Bayesian Network was applied in order to choose, among all candidates, the target object.

3 Background

In this section, we present the two main theories underlying our proposed approach for object tracking: the Hausdorff distance and the Tsallis Nonextensive Entropy.

For the context of interest, given two point sets $A, B \subset \Re^2$, the Hausdorff distance between them can be formally defined as [3],[2]:

$$H(A,B) = max(h(A,B), h(B,A)) \tag{1}$$

where:

$$h(A,B) = max_{a \in A} min_{b \in B} \|a - b\|. \tag{2}$$

$\| \bullet \|$ represents some underlying norm defined in $\Re^2$. The same can be defined for $\Re^n$, $n > 2$. In general, it is used a L_p norm, usually the L_2 or Euclidean norm. The function $h(A,B)$ is called the direct Hausdorff distance from A to B. Intuitively, if $h(A,B) = d$, each point in A must be within a distance d of some point in B.

The maximum value for the Hausdorff distance of two regions A and B in an image with dimension $M \times N$ is the half of its diagonal, $\eta = \sqrt{M^2 + N^2}/2$. Therefore, to get the Hausdorff distance between 0 and 1, we normalize the equation (1) by η, and define the following extensible Haudorff distance:

$$H_e(A,B) = 1 - 2\frac{H(A,B)}{\eta} \tag{3}$$

The Hausdorff distance can be used for matching two data sets of different cardinalities, it is sensitive to small rotations and smooth deformations and is simple for implementation. These are the motivations for applying Hausdorf distance in this work.

The Tsallis entropy is defined as:

$$S_q = \frac{1 - \sum_{i=1}^{k} (p_i)^q}{q - 1} \tag{4}$$

where k is the total number of states of the system, p_i is the probability of finding the system in the state i, and the real number q is an entropic parameter that characterizes the degree of nonextensivity. This expression meets the Shannon entropy ($S = -\sum_i p_i \ln(p_i)$) in the limit $q \to 1$. The Tsallis entropy is nonextensive in the sense that for two statistically independent subsystems, A and B, the probability of the composed system will be:

$$S_q(A+B) = S_q(A) + S_q(B) + (1-q) \cdot S_q(A) \cdot S_q(B) \tag{5}$$

The concept of entropy has been used for image segmentation through thresholding [6][7]. The following method, proposed in [1], is such an example and will be used in our method described in the next section. For an image with k gray-levels, let $p_i = p_1; p_2; \ldots; p_k$ be the probability distribution of the levels. From this distribution we derive two probability distributions, one for the object (class A) and another for the background (class B). The probability distributions of the object and background classes, A and B, are given by

$$p_A : \frac{p_1}{p^A}, \frac{p_2}{p^A}, \ldots, \frac{p_t}{p^A} \tag{6}$$

$$p_B : \frac{p_{t+1}}{p^B}, \frac{p_{t+2}}{p^B}, \ldots, \frac{p_k}{p^B} \tag{7}$$

where $p^A = \sum_{i=1}^{t} p_i$ and $p^B = \sum_{i=t+1}^{k} p_i$.

The a priori Tsallis entropy for each distribution is defined as

$$S_t^A = \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p^A})^q}{q-1} \tag{8}$$

$$S_t^B = \frac{1 - \sum_{i=t+1}^{k} (\frac{p_i}{p^B})^q}{q-1} \tag{9}$$

The Tsallis entropy $S_q(t)$, of the composite system, is dependent upon the threshold value t for the foreground and background. If we suppose that A and B are statistically independent, its value can be obtained by equation (5):

$$S_q(t) = \frac{1 - \sum_{i=1}^{t} (p_A)^q}{q-1} + \frac{1 - \sum_{i=t+1}^{k} (p_B)^q}{q-1} +$$

$$(1-q).\frac{1 - \sum_{i=1}^{t} (p_A)^q}{q-1}.\frac{1 - \sum_{i=t+1}^{k} (p_B)^q}{q-1} \tag{10}$$

We maximize the information measure between the two classes (object and background). When $S_q(t)$ is maximized, the luminance level t is considered to be the optimum threshold value. This can be achieved with a cheap computational effort. Note that the value t which maximizes equation (5) depends on the parameter q also. This is an advantage because, by varying q it is possible to obtain a value of t adapted to current illumination conditions. Certainly, the range for searching q values is a key issue in this proposal. Fortunately, the experiments have shown that this range is small. This fact, makes the nonextensive segmentation adequate to applications which need a frame sequence segmentation at real time under changes in the illumination conditions.

4 The Proposed Tsallis-Hausdorff-HSV Algorithm for Tracking

In our work we proposed a framework which combines the nonextensive Tsallis entropy, Hausdorff distance and HSV Histogram for tracking objects in a frame sequence. We follow the well known idea of using a model, called Model-Object, to find the correspondent Target-Object in the next frame.

The segmentation algorithm of section 3 depends on the parameter q. By changing q we can adapt the measure of information in the image given by the Tsallis entropy (equation (4)). Thus, to get a computational feasible method we sample the q parameter to obtain a set $\{q_1, q_2, ..., q_m\}$ of values. Thus, given a frame f , the optimization of expression (10) will produce a sequence of optimum thresholds $\left\{t_{opt}^{q_1}, t_{opt}^{q_2}, ..., t_{opt}^{q_m}\right\}$ and, consequently, a set of segmented (binary) images $Ic = \{f_1, f_2, \ldots, f_m\}$, where f_i is the image segmented by using $t_{opt}^{q_i}$ as a threshold. Fortunately, the experiments have shown that a suitable small range of q can be found.

Each f_i may have several segmented regions — among foreground and background ones — which we call Target-Regions. Then, let $r_{i,k} \subset f_i$ be the k-th Target-Region of the binary image f_i. In this paper, we assume that the Model-Object MO was defined in the first frame through user interaction or some automatic technique [5, 11].

The object recognition in a scene is accomplished, in this work, with the combination of color and shape characteristics. For shape characteristic we use the Hausdorff distance throughout equation (3), and for color characteristics we used the probability distribution of color in the HSV System. Then, we use

$$H_c(MO, r_{jk}) = \frac{\sum_i \alpha_i \times v_{i,jk}}{\sqrt{\sum_i \alpha_i^2} \times \sqrt{\sum_i v_{i,jk}^2}} \tag{11}$$

as a similarity measure between the HSV Histograms of the Model-Object, MO, and the HSV Histograms of the region of interest r_j. In the equation (11), α_i and $v_{i,jk}$ stand for the bins of the histograms of the regions MO and r_{jk}, respectively.

Finally, we define the similarity measure, $S \in [0, 1]$, which uses color and shape between the Model-Object, MO, and the Target-Object, r_{jk}, as follows:

$$S(MO, r_{jk}) = 1 - [(1 - H_e(MO, r_{jk}))\times$$
$$(1 - H_c(MO, r_{jk}))] \tag{12}$$

This equation stands for the probability of the Target-Object, r_{jk}, occurs since the Model-Object, MO, has occurred with probability distribution given by H_e and H_c. Expression (12) is based on the equation defined in [9] and [12], for text information retrieval, and in [11] for image retrieval, where the probability of occurrence of a document A since document B has occurred, is defined as:

$$P(A|B) = 1 - [(1 - P_i(A|B)) \times (1 - P_j(A|B))\times$$

$$(1 - P_k(A|B))] \tag{13}$$

This equation, which has been defined from a Bayesian Theory [4], also can be interpreted as the probability of the document A be equals to document B. Each term i, j and k stands for a document characteristic (also called evidence). In the present paper,

Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp225-230)

we have adapted this equation by considering the evidences i, j and k, as object characteristics such as color and shape, and so, obtaining a similarity measure between the Model-Object and the Target-Object given by equation (12).

Our proposed Hausdorff-Tsallis algorithm for tracking can be summarized as follows: Given a sample set $\{q_1, q_2, ..., q_m\}$ and a frame f, we compute $I_c = \{f_1, f_2, \ldots, f_m\}$, the set of segmented (binary) frames obtained through the application of the Tsallis entropy algorithm over the frame f. Then, the Hausdorff distance is computed with equation (3) for all $r_{jk} \subset f_j$ and the color distance is computed with equation (11) for the MO. Finally, the MO is updated with the r_{jk} which minimizes the equation (12). With the new MO, the process is repeated for the next frame.

5 Experimental Results

In order to validate the proposed algorithm, several experiments were carried out. The Fig. 1 shows an example of the problem at hand. For simplification, the Fig. 1-(a) shows a unique object under a homogeneous background. We assume that before the frame capture, the Model Object' contour is a known data. The value of $q = 0.35$ was chosen since it generates good segmentation yielding the contour of Fig. 1-(b) (frame 1). However, in the Fig. 1-(c) (frame 7), the illumination conditions changed, and the parameter $q = 0.35$ is now inadequate yielding a wrong segmentation (as can be seen by the curves outlining the Target-Object). In this case, there will be error propagation to the remainder frames.

Even if there is an automatic mechanism avoiding error propagation — such as, for example, if an object data base is requested periodically to check if the current contour is according to the Model Object — the problem holds yet, due to the fixed value for q under changes in illumination.

The Fig. 2 shows a frame sequence with the same object of Fig. 1. However, we have applied different values for q, taken according to different scene conditions. Moreover we observe changes in the camera view to show the method's robustness against changes in the object scale. We carried out the complete experiment using our approach with 300 frames for the same scene of the Fig. 1. From this experiment, 4 frames are presented in Fig. 3 with the Target-Object inside a rectangle to indicate a positive finding. From these frames we can see strongly variation of illumination. However, we observe a correct recognition of the Target-Object. In this sequence, we can also see

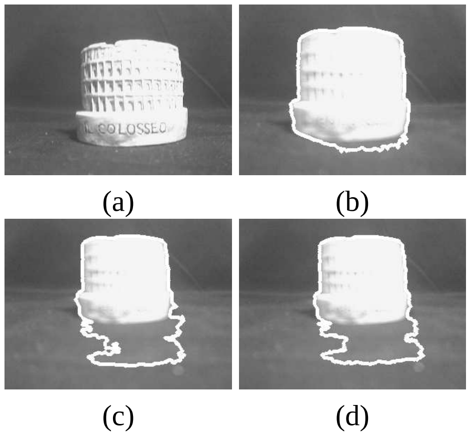

Figure 1: A frame sequence illustrating the problem at hand. (a) shows the image of Model-Object; (b)-(d) show some remainder frames under different illumination conditions. The white curves around the objects (b)-(d) bound the best matched regions. The parameter $q = 0.35$ is the same for all frames. Under changes in illumination conditions the parameter $q = 0.35$ becomes inadequate to segment the remainder Target-Regions.

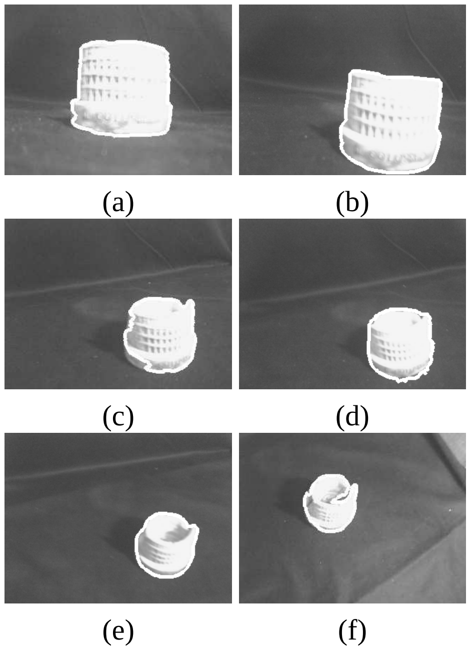

Figure 2: A frame sequence illustrating the application of the proposed model when changing q value. (a)-(f) shows the same object model of Fig. 1, under different illumination and viewer conditions.

a strongly variation in the object viewpoint, e.g: Figs. 3-(a), (c). In the performed experiments, the entropic parameter q had a range $0 \leq q \leq 1$.

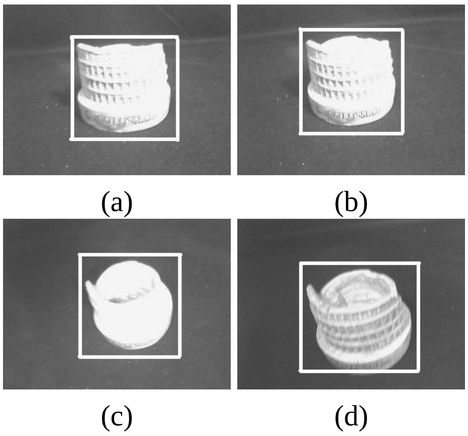

Figure 3: A sequence of frames for which the proposed method is applied. The rectangular box shows the area found with a target object. In this experiment there are changes in view and illumination.

In order to measure the precision of the proposed method, we may define the following equation:

$$Precision = \frac{A \cap B}{A \cup B} \tag{14}$$

where A and B are two sets of pixels. This equation captures the matching precision underlying both the position and shape of the regions. We can observe that, when the shape of A resembles the shape of B and also they have geometrical centers near each other, the obtained value tends to 1.0. Inversely, when they have no intersection, the precision tends to 0. This equation matches the shape and position at the same time.

In our experiment, we have used this equation to measure the precision of our proposed algorithm at each frame, where the Model-Object is matched against a manual segmentation, and the equation (14) is used to get the precision. The graphic frames $\times$ precision for 300 frames are presented in Fig. 4.

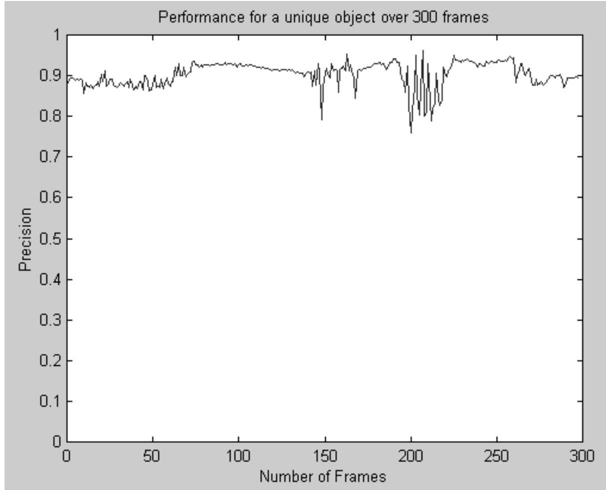

Figure 4: This graphic shows the frames $\times$ precision for 300 frames.

6 Conclusion

We presented a methodology for object tracking in a frame sequence. This methodology uses a combination of Hausdorff distance for shape matching and HSV histograms for color comparison. Also, it uses Tsallis nonextensive entropy for image segmentation.

The usefulness of Tsallis nonextensive entropy was demonstrated to set the segmentation according to changes in illumination. When comparing the segmentation using Tsallis entropy with other methods for segmentation that use spatial filters and/or mathematical morphology, the handle of only one parameter is one advantage, since the method is efficient and robust against changes in the illuminations conditions. The proposed model is easily extensible to others features and should be tested for real-time applications.

Acknowledgment

The authors are acknowledged to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), a Brazilian agency for scientific financing.

References

[1] M. P. Albuquerque, M. P. Albuquerque, I.A. Esquef, and A.R.G. Mello. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25:1059–1065, 2004.

[2] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Medical Imaging*, 15(9):850–863, 1993.

[3] D. P. Huttenlocher, J. J. Noh, and W. J. Rucklidge. Tracking non-rigid objects in complex scenes. Technical Report 1320, Dept. of Computer Science, Cornell Univ., 1992.

[4] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. Springer, 2001.

[5] C. Kim and J. Hwang. Fast and automatic video object segmentation and tracking for content-based applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(3):122–129, February 2002.

[6] C. H. Li and C. K. Lee. Minimum cross entropy thresholding. *Pattern Recognition*, 26:617–625, 1993.

[7] N. R. Pal. On minimum cross entropy thresholding. *Pattern Recognition*, 26:575–580, 1996.

[8] N. Paragios and R. Deriche. Geodesic active contours and level sets for detection and tracking of moving objects. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22:266–280, March 2000.

[9] B. Ribeiro-Neto and R. R. Muntz. A belief network model for IR. In *ACM Conference on Research and Development in Information Retrieval - SIGIR96*, pages 253–260, 1996.

[10] P. S. Rodrigues, D. M. C. Rodrigues, G. A. Giraldi, and R. L. S. Silva. A bayesian network model for object recognition environment using color features. In *Eletronic Proceedings of Sibgrapi'04*, 2004.

[11] P. S. Rodrigues, D. M. C. Rodrigues, G. A. Giraldi, and R. L. S. Silva. Object recognition using bayesian networks for augmented reality applications. In *Proceedings of VII Symposium on Virtual Reality (SVR'04)*, 2004.

[12] I. Silva, B. Ribeiro-Neto, P. P. Calado, E. S. Moura, and N. Ziviani. Link-based and content-based evidential information retrieval in a belief network model. *Proceedings of the 23rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 96–103, July 2000.

[13] H. Xu, A. Younis, and M. R. Kabuka. Automatic moving object extraction for content-based applications. *IEEE Transaction on Circuits and Systems for Video Technology*, 14(6):796–812, June 2004.