14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

TRACKING SYSTEM USING CAMSHIFT AND FEATURE POINTS

GANOUN Ali[*,**], *OULD-DRIS Nouar*[*], *and CANALS Raphaël*[*]

[*]Laboratory of Electronics, Signals and Images (LESI), University of Orléans, FRANCE
[**] Faculty of Engineering, EE Department, University of Garyounis, LIBYA

{Ali.Ganoun, Nouar.Ould-Driss, Raphael.Canals}@univ-orleans.fr

ABSTRACT

In this paper, we present a new object tracking approach in grey-level image sequences. The method is based on the analysis of the histogram of distribution of the image and the feature points of the object. Among the essential contributions of this work, we can quote a better modelling of the object to track and the consideration of the target appearance changes during the sequence. Our approach is a prolongation of the CamShift algorithm (Continuously Adaptive MeanShift) applications. The goal is to widen the field of application of this algorithm in order to adapt it to grey-level image sequences presenting strong modifications of shape, luminosity and grey-level values.

1. INTRODUCTION

In the literature, many methods have been developed to solve the object tracking problem. That begins with simple approaches based on the visual primitives which are detected and tracked in images by employing correlation, and does not finish with hybrid approaches which process the edge or region data, or the movement of the object in order to track it in the images sequence [1], [2].

Another robust and nonparametric technique is proposed in the library of computer vision "Intel Corporation, 2001" [3]: it implements the CamShift algorithm which uses a one-dimensional histogram to track an object with known hue in colour images sequences. The difficulty emerges when one wishes to employ this algorithm to track objects without a priori knowledge nor training phase. In [4], it was then stated that the use of a three-dimensional histogram solves the problem and leads to a target localization improvement. The histogram back-projection permits to obtain a probability distribution image which is processed by the iterative CamShift algorithm in order to find the maximum of the distribution and thus the object centre.

In this work, we present a method to track object in grey-level images sequences. The general principle of our approach lies in the fact that we calculate the displacement of the target centre between two successive images thanks to the matching of feature points. We apply then the CamShift algorithm in order to determine the final object position.

After this introduction, section 2 briefly presents the CamShift algorithm. In section 3, we describe our tracking system. Our target model is presented in section 4, and the obtained results are shown in section 5. We finish by a conclusion and propose some orientations for future research.

2. THE ALGORITHM OF CAMSHIFT

The principle of the CamShift algorithm is given in [3] and [5]. Each image of the sequence is converted into a probability distribution image relatively to the histogram of the object to be tracked. From this image, the centre and the size of the object are determined by using the CamShift algorithm. These new centre and size are employed to place the research window in the next image. This process is then repeated for a continuous target tracking in the video sequence.

The algorithm of CamShift thus employs a 2D probability distribution image produced from a back-projection of the target histogram with the image to process.

The CamShift algorithm calls upon the MeanShift one to calculate the target centre in the probability distribution image, but also the orientation of the principal axis and dimensions of the probability distribution [6]. It is a matter of finding a rectangle presenting the same moments as those measured on the probability image. These parameters are given from the first and second moments [4]

3. TRACKING SYSTEM

The global system of tracking we have implemented is primarily based on the CamShift method, but also on the feature points matching (*FP*) between two successive images for the preliminary prediction of the target displacement. The principal steps of this algorithm are stated as follows:

1- Calculation of the target model (histogram and *FP*).
2- Calculation of the temporary target displacement by the *FP* matching between image I_t of the sequence at instant t and image I_{t+1} at instant t+1.
3- Determination of a reduced search zone positioned on the centre C_{temp} calculated at step 2.
4- Calculation of the probability image in the search zone.
5- Application of the MeanShift algorithm to determine the new target centre.
6- Actualization of the target model.

4. TARGET MODEL

The target model is based on the model of the histogram and the model of the feature points. Rather than to calculate these models in the rectangle surrounding the target and defined by the operator, we apply a Canny edge detection [7] coupled to a basic contour closing algorithm and carry out these processsing inside the region defined by the external edges. In this case, in order to limit artefacts concerning the background edge detection, the operator must define a rectangle as near as possible of the target, nay a rectangle in the target.
The histogram is now computed in the region defining the target. In fact our histogram is not sufficient for a good modelling of the target [4]. Indeed, if certain pixels in the image present the same grey levels that those belonging to the target, it is necessary that the weight of these grey levels in the model is reduced to prevent that the MeanShift algorithm provides a bad target localization. To solve this problem, a ratio histogram is implemented.
In order to be able to predict the target displacement, we calculate another model based on the feature points (*FP*). We can define a *FP* as a point in the image where significant texture changes occur, such as corners or T-junctions. Several detectors of *FP* were developed since about twenty years. In our work, we use the Harris detector which defines the *FP* as being the local extremes of the cornerness image R(x,y) defined in [8], [9], [10].
The object model consists of a set of *FP* being inside this object and each *FP* is characterized by a vector containing certain attributes. We note the *FP* of an object Q_t in an image I_t by FP_i, $i \in \{1,\ldots,N\}$ where N is the number of *FP*. Each FP_i is thus characterized by a vector noted V_i such as:

$$V_i(t) = \left\{ I, I_x, I_y, R, \mu, \upsilon, x_x, y_c \right\}$$

where I is the grey level in FP_i, I_x and I_y are the first order derivative Gaussian in x and y, R is the cornerness utilized for the *FP* detection, μ and ν are respectively the average and the variance of the grey levels in a 5x5 window around FP_i, and x_c and y_c are the distances between FP_i and the centre of the object Q_t.
The vector of the *FP* set modelling the object Q_t to track is given by:

$$V^{Q_t}(t) = \left\{ V_i(t) \middle| i = 1, \ldots, N \right\}$$

Once the object Q_t in the current image I_t is characterized by a *FP* set, it is necessary to find the corresponding model in the next image I_{t+1}. The basic principle is to determine, for each FP_i, $i \in \{1,\ldots,N\}$, the best matching FP_j, $j \in \{1,\ldots,M\}$. To find j, we compare the vector $V_i(t)$ and each vector $V_j(t+1)$, as illustrated in figure 1.
FP_j in image I_{t+1} corresponds to FP_i in image I_t if j is the index *k* which gives the minimal Mahalanobis distance $D_M(i,k)$, for $k \in \{1,\ldots,j,\ldots,M\}$. The Mahalanobis distance thus enables us to correctly pair all *FP* in one image to the next one. Possibly, a few *FP* can not have pairings because some of them disappear and others appear during the sequence.
The Mahalanobis distance is given by:

$$D_M(i,j) = \sqrt{(V_{ij} - \gamma)^t C^{-1} (V_{ij} - \gamma)}$$

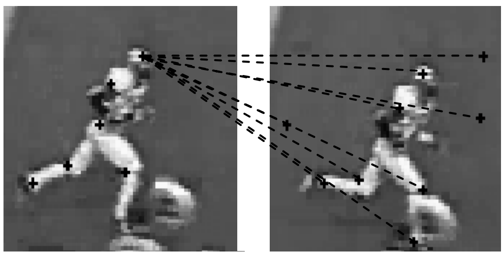
where $V_{ij} = V_i(t) - V_j(t+1)$. γ and C are respectively the mean values vector and the covariance matrix.
A good pairing of *FP* enables us to calculate the displacement of the object centre between two successive images. It is thus a question of considering the displacement of each correctly paired *FP* and of calculating the mean displacement in x and y. Initially, the displacement of all paired *FP* is calculated. We suppose that all *FP* appreciably covered the same distance and we calculate the displacement $(\Delta X_g, \Delta Y_g)$ of each *FP*, with $g \in \{1,\ldots,h\}$, h being the number of paired *FP* of image I_t to image I_{t+1}.

$$\Delta X_g = X_g(t) - X_g(t+1)$$
$$\Delta Y_g = Y_g(t) - Y_g(t+1)$$

$X_g(t)$, $Y_g(t)$, $X_g(t+1)$ and $Y_g(t+1)$ are the respective *FP* abscissas and ordinates in images I_t and I_{t+1}.
Then we calculate a more correct estimate of displacement $(\Delta mX, \Delta mY)$:

$$\Delta mX = mean(\Delta X_g)$$
$$\Delta mY = mean(\Delta Y_g)$$

In order to consider a more precise but always temporary displacement of the object centre, we compute again the displacement without taking into account some *FP* presenting a displacement much larger than the average one calculated previously. The object centre in image I_t being noted C, it is enough to add the estimated displacement $(\Delta mX, \Delta mY)$ to the centre C in order to find the new object centre in image I_{t+1}. The *FP* matching from an image to the other one presents often some computation errors, owing to the fact that some *FP* can be badly paired. It is the reason we can centre the search zone on this new temporary centre C_{temp}, in order to build the probability distribution image in a reduced search window.



a) Image I_t b) Image I_{t+1}

Figure 1- Pairing by optimal matching of each FP_i in image I_t with all the FP_j in image I_{t+1}

The back-projection of the target histogram with the reduced research window enables us to obtain the probability distribution image. We apply the MeanShift algorithm on this probability image to determine the final centre of the required object. The final centre of the object Q_{t+1}, its orientation and its dimensions are calculated as explained in [4].
In order to consider the various appearance and shape changes of the target, we update its model for each image of the sequence with weighting coefficients computed from the preceding histogram. Indeed, the histogram updating permits to maintain up to date the model, and to mitigate the problem of errors which can accumulate progressively during the sequence. The histogram used for the target model is a one-dimensional weighed and filtered one, such as it was presented in [4] and [9].

5. RESULTS

We have applied our method on various video sequences. The images in figures 2, 3 and 4 illustrate the obtained results.
In Fig.2, we track an American football player in a video sequence of 320x240 8-bit images. The algorithm tracks the player as it moves from one frame to the next one in approximately 0.16 second on a 2GHz PC in Matlab.
In Fig.3, the algorithm tracks a woman while managing occlusion with a cyclist, in about 0.50 second. During the occlusion, we note a shift of the target window as a result of the mass centre displacement. After the passage of the cyclist, the algorithm correctly finds again the object to be tracked. During this occlusion, we can observe a drifting problem when updating the histogram, that is normal since a part of the target is masked. Thanks to the feature points, we can nevertheless maintain the search zone on the visible target part. In the same way, when some feature points disappear because of noise or occlusion, no pairing or bad pairing can be realized, but as all the feature points are used to predicate the new target centre, the problem is largely minimized, especially when their number is high. Moreover, the Camshift algorithm displaces then this centre in order to obtain the best matching. So there is a real cooperation between the histogram model and the feature point one.
Figure 4 demonstrates the tracking of a woman in a 576x720 image sequence with occlusion. The algorithm takes about 0.34 second to track correctly the woman, even when the woman walks behind the road sign. Indeed the occlusion is less important here than in the previous example, so the mass centre displacement is insignificant as the major part of the target is visible.

6. CONCLUSION

We have proposed in this paper a new object tracking approach in grey-level images sequences, based at the same time on the CamShift algorithm and on the feature points of the target. In this approach, the target is characterized by two different models: its one-dimensional histogram and its feature points. It is thus a question of estimating, initially, the target displacement between two successive images thanks to the matching of the feature points detected in the two images. This evaluated displacement is only temporary and allows us to gain in computational time by reducing the size of the search window. In the second time we calculate in this window the probability distribution image from which we determine the position of the object.
Execution times obtained under Matlab allow us to envisage a real-time implementation of our method on an embedded system.
We can improve our method by memorizing the trajectory of the feature points and the mass centre. That should permit to limit the search zone for the matching of each feature point while avoiding some matching ambiguities, and thus should allow reducing the computational time as the combinatory complexity should be decreased. Moreover, as the feature points and the mass centre are correlated, we could manage in a better way the occlusion problem in case of feature point disappearance.

REFERENCES

[1] R. Canals, A. Roussel, J.-L. Famechon and S. Treuillet, “A Vision-Based Biprocessor-Oriented Target Tracking System”, IEEE Trans. on Industrial Electronics, Vol. 49, N° 2, pp. 500-506, 2002.
[2] S. Besson, M. Barlaud, and G. Aubert, “A 3-Step Algorithm Using Region-Based Active Contours For Video Objects Detection”, EURASIP Journal on Applied Signal Processing, pp. 572-581, issue 6, 2002.
[3] Intel Corporation, “Open Source Computer Vision Library Reference Manual”, 123456-001, 2001.
[4] G. John Allen, Y. D. Richard Xu and S. Jin Jesse, “Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces”, Inc. Australian Computer Society, vol.36, 2004.
[5] G. R. Bradski. “Computer vision face tracking for use in a perceptual user interface”, Intel Technology Journal, 2nd Quarter, 1998.
[6] Y. Cheng, “MeanShift, mode seeking, and clustering”, IEEE Trans. on PAMI, vol.17, pp.790-799, 1995.
[7] J. F. Canny, “A computational approach to edge detection”, IEEE Trans. on PAMI, pp. 679–698, 1986.
[8] J. Harris and M. Stephens, “A combined corner and edge detector. 4th ALVEY Vision Conference, pp. 147–151, 1988.
[9] V. Gouet and B. Lameyre, “SAP: A robust approach to track objects in video streams with snakes and points”, British Machine Vision Conference, pp.737–746, 2004.
[10] V. Gouet and N. Boujemaa, “About optimal use of color points of interest for content-based image retrieval”, Internal report, INRIA Rocquencourt, 2002.
[11] D. Comaniciu, V. Ramesh, and P. Meer “Kernel-Based Object Tracking”, IEEE Trans. on PAMI, vol. 25, n° 5, pp.564-575, 2003.

Figure 2 - Football player tracking. Images 1, 14 and 32 of the sequence

Figure 3 - Tracking of a woman in presence of occlusion.
Images 1, 8, 11, 12, 13 and 31 of the sequence

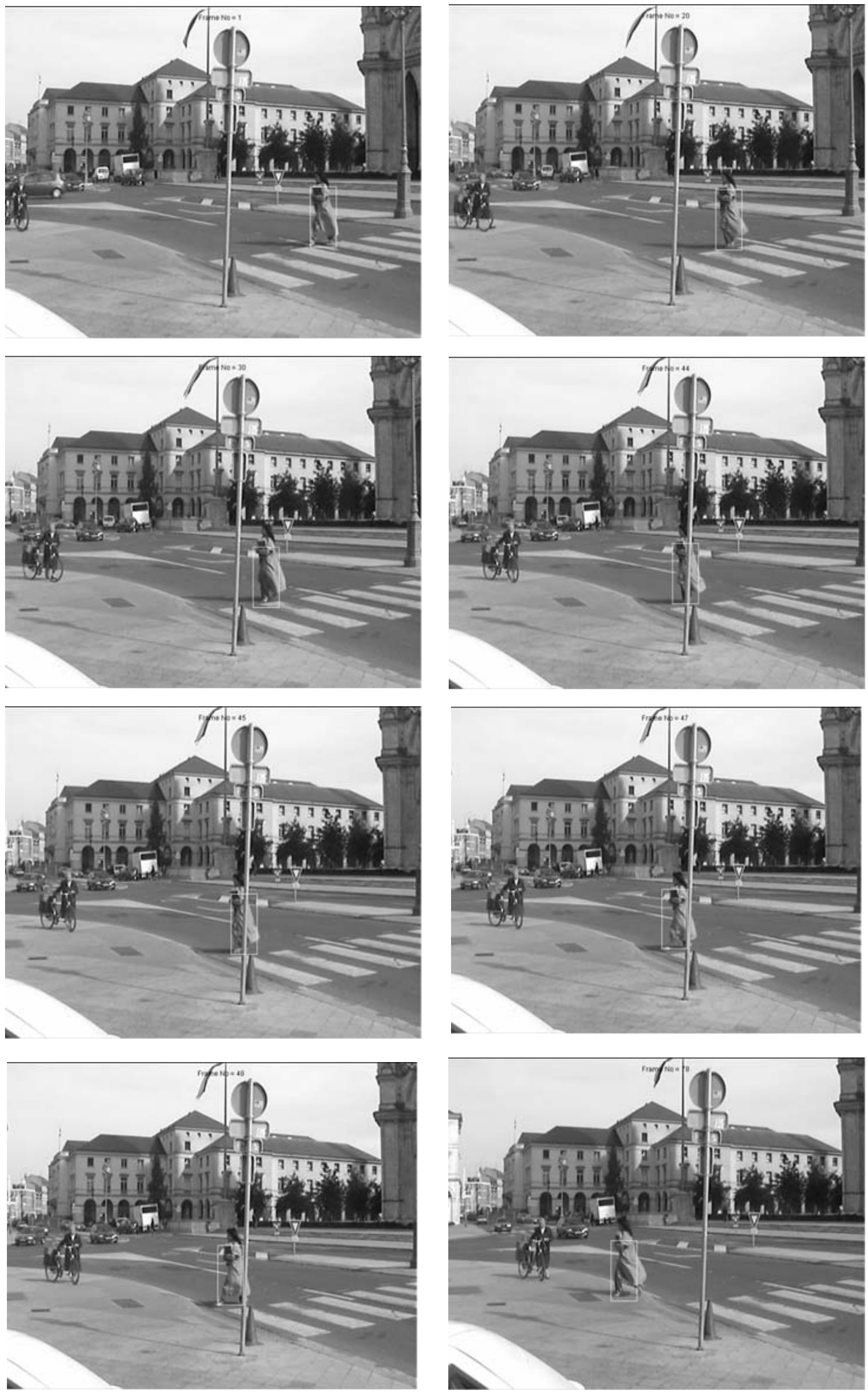


Figure 4 - Tracking of a woman in presence of occlusion.
Images 1, 20, 30, 44, 45, 47, 48 and 78 of the sequence