PROJECTION-HISTOGRAMS FOR MEAN-SHIFT TRACKING

Michèle Gouiffès[1] *and Florence Laguzet*[2] *and Lionel Lacassagne*[1]

(1) IEF CNRS UMR 8622, (2) LRI CNRS UMR 8623
University of Paris 11, 91405 ORSAY cedex
michele.gouiffes@u-psud.fr

ABSTRACT

This paper proposes an extension to the mean shift tracking. We introduce XY projection-histograms which, more than providing statistical information about the target to track, embeds information about the spatial arrangement of pixels. This approach, without any increase in complexity, provides a better robustness and quality of the tracking. That is asserted by the experiments performed on several sequences showing vehicle and pedestrians in various contexts.

Keywords Mean Shift Tracking, Color, Projection Histograms.

1. INTRODUCTION

Visual tracking is a common task, on which many crucial applications rely heavily on: traffic analysis and control, security monitoring, driving assistance, industrial control. Tracking encounters various difficulties, such as the clutter of the environment, the non-rigid motion, the photometric and geometric variations, the partial occlusions.

Local tracking methods, for example template matching [6, 9] or SSD tracker, take comprehensively the spatial information into account. Although time-effective, they usually fail when non-rigid objects are considered. Global approaches, mean-shift [2] to begin with, represent the target with a global statistical representation, mainly based on color or texture. A large number of extensions has been proposed, they differ mainly by the statistical distribution and on the similarity function [12]. Some authors have improved the procedure either by introducing an object/background classification [8] or by combining mean-shift with local approaches [3] or with particle filtering, in order to deal with severe occlusions .

Unfortunately, classical histograms are not always discriminative, since they do not preserve spatial information. Our works focus on the spatio-colorimetric representation of the object, in order to enhance the discrimination ability of the histogram. Some authors have addressed that issue by proposing the spatiogram [4] and the correlogram [13]. In the former method, each bin of the histogram is weighed by the mean and covariance of the locations of the corresponding pixels. In the latter, color correlations are considered for several directions. Differently in [11, 10], the object bounding box is spatially divided into regions or segments, to be processed separately. More recently, some new kernel methods [7, 5] use the covariance matrix of features, which is a compact spatio-colorimetric representation of the target.

In this paper, we introduce a simple spatio-colorimetric representation of the object, based on six projection-histograms [1], one histogram per color channel and spatial directions (x and y in this work). They are compared to the classical RGB 3D histogram on a few road sequences involving cars and pedestrians.The expected benefit is a gain in robustness and accuracy of the tracking.

The continuation of the paper is structured as follows. Section 2 introduces the color projection-histograms. Then, Section 3 explains the principles of the mean-shift tracker. To conclude, Section 4 asserts the relevance of the proposed method by comparing the robustness of our technique.

2. THE COLOR PROJECTION-HISTOGRAMS

Let be a trichromatic image and $\boldsymbol{c}_i = ({c^1}_i, {c^2}_i, {c^3}_i)$ the color components of a pixel i of coordinates $\boldsymbol{p}_i = (x_i, y_i)$ in that image. We quantize the color dynamics so that N colors are considered (instead of 256 generally), and the new components are $n_i = c_i \times 256/N$. Now let us consider a region of interest $\mathcal{W}$, of dimensions $h_x \times h_y$, which has to be tracked during an image sequence.

The projection-histograms are illustrated on Fig. 1. For each direction x (rows) and y (columns), three 2D histograms are computed, one for each channel. They are noted H_x and H_y and are indexed by the color channel c, and their dimensions (m, n). Their size is $N \times M$, where N is the number of colors considered for each color channel. Each axis x and y is divided into M sections. We note x' and y' the coordinates after an uniform quantization such that $x' = x \times M/h_x$ and $y' = y \times M/h_y$. The projection-histograms are finally given as:

$$H_x(c, n, m) = \sum_{\boldsymbol{p}_i \in \mathcal{W}} \delta(x'_i - m)\delta(n_i - n) \quad (1)$$

$$H_y(c, n, m) = \sum_{\boldsymbol{p}_i \in \mathcal{W}} \delta(y'_i - m)\delta(n_i - n) \quad (2)$$

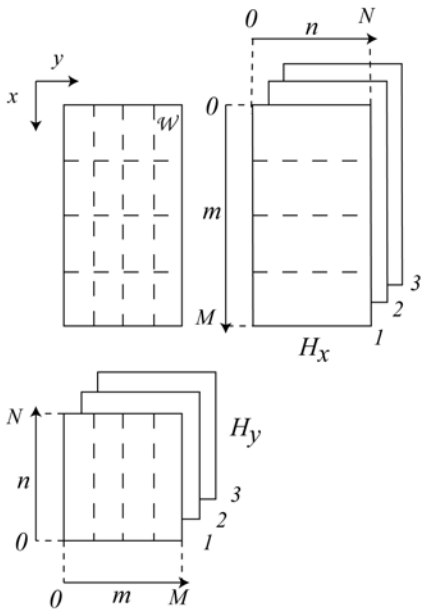

Figure 1: Illustration of the XY projection-histograms. For each object $\mathcal{W}$, six histograms are computed, for two directions and three color channels.

.

3. MEAN SHIFT PROCEDURE

3.1. Spatio-Colorimetric representation of the target

The target to track is generally represented by its bounding box $\mathcal{W}$, resulting from a downstream algorithm such as motion analysis, stereovision or pattern recognition. Once detected, the target model in initial frame 0, is described by the set of six projection-histograms:

$$\mathcal{H}^0 = \{H_y, H_x\}^0_{c=1..3} \quad \sum_{n=0}^{N}\sum_{m=0}^{M} H_{\cdot}{}^0(c,n,m) = 1 \ \forall c \qquad (3)$$

The target candidate in frame k, has a bounding box called $\mathcal{W}^k$ centered on $\boldsymbol{p}^k$. It is described as :

$$\mathcal{H}^k(\boldsymbol{p}^k) = \{H_y, H_x\}^k_{c=1..3}(\boldsymbol{p}^k) \quad \sum_{n=0}^{N}\sum_{m=0}^{M} H_{\cdot}{}^k(c,n,m) = 1 \ \forall c \qquad (4)$$

The similarity between the target model at initial location and the target candidate at location $\boldsymbol{p}^k$ is computed as the sum of the marginal similarities between each corresponding projection histogram. Similarly to the initial mean shift (MS) algorithm, the Bhattacharyya similarity is chosen :

$$\rho(\boldsymbol{p}^k) = \sum_{c=1}^{3}\sum_{n=0}^{N}\sum_{m=0}^{M}\left(\sqrt{H_y^0(c,n,m)H_y^k(c,n,m)(\boldsymbol{p}^k)}+ \sqrt{H_x^0(c,n,m)H_x^k(c,n,m)(\boldsymbol{p}^k)}\right) \qquad (5)$$

The candidate location which maximizes (5) is found by proceeding a gradient-based optimization.

3.2. Spatial representation of the target

Mean-shift can suffer from partial occlusions and ill-separation between the object and the background. To solve those issues, each pixel of $\mathcal{W}$ is weighted by an isotropic kernel $K(\boldsymbol{p})$ which allocates a higher relevance to the central part of $\mathcal{W}$, where the object is the most likely to be (compared to background or occluding objects). In addition, $K(\boldsymbol{p})$ provides a finite smoothing kernel for the gradient-based minimization (5). We chose the Epanechnikov kernel [2]. In addition, in order to better reduce the contribution of the background in the reference histogram $\mathcal{H}^0$, the colors belonging to the background are subtracted from the histogram using the log-likelihood ratio of foreground/background as in [3]. Unlike most classical color representation, our projection-histograms are likely to be affected by the spatial changes during the time.Therefore, it has to be updated frequently. In our experiments, it is updated in each frame.

3.3. Mean Shift procedure

Considering a given target model $\mathcal{H}^0$ and the previous location of the object $\boldsymbol{p}_{k-1}$ in previous frame $k-1$, the tracking consists in finding in each frame the candidate location $\boldsymbol{p}_k$ which maximizes the similarity (5) to the model. The Bhattacharyya distance 5 is expanded in Taylor series as in [2] in order to allow gradient based optimization. Here are the stages of the algorithm:

1. Initially, the object is assumed to be motionless so that the initial estimate location, called $\boldsymbol{p}_0$, is such that $\boldsymbol{p}_0 = \boldsymbol{p}_{k-1}$. The new histograms are computed at that location $\mathcal{H}^k(\boldsymbol{p}_0)$, as well as the similarity $\rho[\mathcal{H}^k(\boldsymbol{p}_0), \mathcal{H}^0]$.
2. The new candidate location $\boldsymbol{p}^k$ is computed:
$$\boldsymbol{p}^k = \frac{\sum_{i\in\mathcal{W}} \boldsymbol{p}_i\, w_i\, g\left(\left\|\frac{\boldsymbol{p}_0 - \boldsymbol{p}_i}{h}^2\right\|\right)}{\sum_{i\in\mathcal{W}} \boldsymbol{p}_i\, w_i\, g\left(\left\|\frac{\boldsymbol{p}_0 - \boldsymbol{p}_i}{h}^2\right\|\right)} \text{ with } g(x) = -k'(x) \qquad (6)$$
with the following definition of the weights derived from the Taylor expansion:
$$w_i = \sum_{c=1}^{3}\sum_{n=0}^{N}\sum_{m=0}^{M}\left(\sqrt{\frac{H_x^0(c,n,m)}{H_x^k(c,n,m)}}\delta(x_i'-m)\delta(n_i-n)+ \sqrt{\frac{H_y^0(c,n,m)}{H_y^k(c,n,m)}}\delta(y_i'-m)\delta(n_i-n)\right) \qquad (7)$$
3. while $\rho[\mathcal{H}^0, \mathcal{H}(\boldsymbol{p}^k)] < \rho[\mathcal{H}^0, \mathcal{H}(\boldsymbol{p}^0)]$ do $\boldsymbol{p}^k = 0.5(\boldsymbol{p}^k + \boldsymbol{p}^0)$
4. if $\left\|\boldsymbol{p}^k - \boldsymbol{p}^0\right\| < \epsilon$ then STOP, otherwise $\boldsymbol{p}^0 \leftarrow \boldsymbol{p}^k$ and go to step 2.

Scale change. The scale change of h is managed in a similar fashion as [2], i.e considering previous size h^{k-1}, and an offset $\Delta h = 0.1h^{k-1}$. The optimal size h_{opt} is chosen as the one with maximizes the Bhattacharyya similarity

Figure 2: The five sequences used in the experiments.

among three sizes : h^{k-1}(no scale change), $h^{k-1} + \Delta h$ (larger), $h^{k-1} - \Delta h$ (smaller), then the new size is given by:

$$h = \gamma h_{opt} + (1 - \gamma)h^{k-1}$$

with $\gamma = 0.1$ in our experiments.

Loss of the target. The object tracked is considered to be lost when the final Bhattacharyya coefficient is higher than a threshold T_{out}.

4. EXPERIMENTS

We experiment the robustness and accuracy of the projection-histograms versus the 3D color histogram on 5 sequences showing vehicles or pedestrians. The first and last frames of these sequences are shown on Fig. 2 on first and second row, with the projection MS tracking results. In each sequence, the target is selected manually. We call $\boldsymbol{p}_1$ the up left corner and $\boldsymbol{p}_2$ the bottom right corner:

1. *Car 1*: Sequence of dataset 5, testing, camera1 of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance 2001 (PETS). The selected sequence goes from frame 0 to 490. The images are of size 576×768 and the coordinates of the selected target are $\boldsymbol{p}_1 = (410, 13)$ and $\boldsymbol{p}_2 = (505, 51)$. Note that we pick one frame over 10, which makes the tracking more difficult.
2. *Pedestrian 1*: Sequence of dataset 3, testing, camera1 of the PETS01 with coordinates $\boldsymbol{p}_1 = (410, 13)$ and $\boldsymbol{p}_2 = (505, 51)$, tracked from frames 1415 to 1636.
3. *Pedestrian 2*: Sequence of dataset 1, from frame 1345 to 1475, with coordinates $\boldsymbol{p}_1 = (415, 492)$ and $\boldsymbol{p}_2 = (510, 630)$. we track a couple of pedestrians.
4. *Car 2:* That sequence `dtneu_schnee`[1] shows a street view under falling neve (image size 576×768). We consider frame 0 to 166, $\boldsymbol{p}_1 = (212, 320)$ and $\boldsymbol{p}_2 = (250, 360)$.

[1]That sequence is available on the web on http://i21www.ira.uka.de/image_sequences/

Table 1: Increase of CPU times of the proposed method compared to 3D histograms.

Sequence	% CPU time
Car 1	-32
Pedestrian 1	-11
Pedestrian 2	-14
Car 2	-5
Pedestrian 3	-10

5. *Pedestrian 3* : that sequence *walkstraight* (images size 240×320) comes from the INRIA-IRISA (Rennes). We analyze frames 30 to 108 and the initial coordinates are $\boldsymbol{p}_1 = (67, 263)$ and $\boldsymbol{p}_2 = (225, 307)$

In the experiments, the values of our parameters are $N = 8$ and $M = 8$. The size of the color RGB cube is $8 \times 8 \times 8$. Fig. 3 to 7 show the picture of the target objects, with the classical MS (first row) and our projection MS (second row). In *Car 1* of fig.3, the classical MS loses the target, contrary to the proposed approach. In the subsequent sequences, Fig. 4 to Fig. 7 show that the classical MS tracker usually center the target on the predominant color of the object (for example the blue pants in 7).

The computation of the projection-histograms is not more time consuming, and is not significantly more complicated than the RGB cube. Because the executing times depend on the target size, the number of iterations, etc, we choose to compare in table 1 the executing times for each sequence in a relative way. The values represent the time increase of the projection MS relatively to the classical MS (in $\%$). In each case, our algorithm is less time consuming, from -32 $\%$ % to -5%. In general, the projection histograms are more representative of the target, therefore they need a lower number of iterations to converge.

Figure 3: The *Car 1* sequence. The classical MS tracker (1^{st} row) fails. Our MS tracks the car during the whole sequence.

Figure 4: The *Pedestrian 1* sequence.

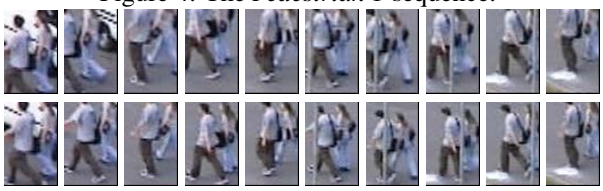

Figure 5: The *Pedestrian 2* sequence.

Figure 6: The *Car 2* sequence.

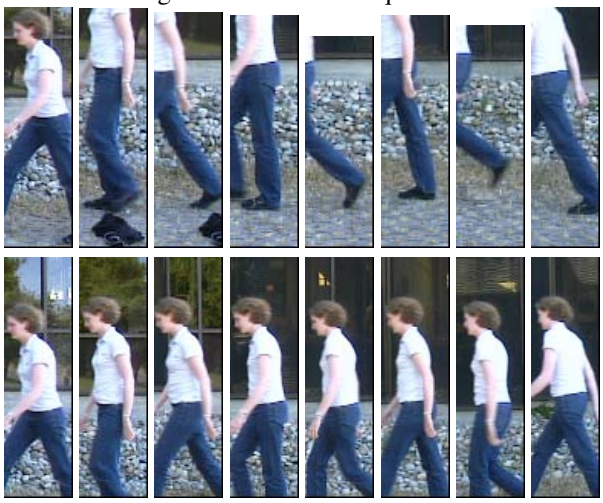

Figure 7: The *Pedestrian 3* sequence.

5. CONCLUSION

The paper proposed an extension to the mean-shift classical tracker by using a set of six color projection-histograms. Although simple, these data structures are more informative and discriminative than classical 3D histograms, since the spatial arrangement is better preserved. Undoubtedly, using 3D histogram instead of 1D classical should be more efficient for a better discrimination ability. In our work, we show that the integration of spatial information in marginal color histograms is finally more satisfying. The tracking results are improved in terms of robustness and in terms of quality, with no increase of the computation times. In our future works, it could be interesting to compare our technique with the use of more advanced spatio-colorimetric representations, the spatiogram to begin with.

6. REFERENCES

[1] *2D-Object Tracking Based on Projection-Histograms*, London, UK, 1998. Springer-Verlag.

[2] D. Comaniciu and. Kernel-based object tracking. *IEEE Trans. PAMI*, 25(5):564–577, 2003.

[3] R. V. Babu, P. Pérez, and P. Bouthemy. Robust tracking with motion estimation and local kernel-based color modeling. *IVC*, 25, 2007.

[4] S.T. Birchfield and S. Rangarajan. Spatiograms versus histograms for region-based tracking. In *Computer Vision and Pattern Recognition*, pages 1158–1163, 2005.

[5] P. Karasev, J. malcom, and A. Tannenbaum. Kernel-based high dimensional histogram estimation for visual tracking. In *IEEE ICIP*, October 2008.

[6] B.D. Lucas and T. Kanade. An iterative image registration technique. In *Internation Joint Conf. on A.I.*, pages 674–679, August 1981.

[7] F. Porikli, O.Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *IEEE Computer Vision and Pattern Recognition*, pages 728–735, 2006.

[8] S. Rastegar, M. Bandarabadi, Y. Toopchi, and S. Ghoreishi. Kernel based object tracking using metric distance transform and svm classifier. *Aus. Jour. of Basic and Applied Science*, 3(3):2778–2790, 2009.

[9] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report CMU-CS-91-132, April 1991.

[10] F. Wang, S. Yu, and J. Yang. Robust and efficient fragments-based tracking using mean-shift. *International Journal of Electronics and Communications*, 2009.

[11] D. Xu, Y. Wang, and J. An. Applying a new spatial color histogram in mean-shift based tracking algorithm. In *New Zealand Conference on Image and Vision Computing*, 2005.

[12] C. Yang, R. Duraiswami, and L. Davis. Efficient mean-shift tracking via a new similarity measure. In *Proceedings of the 2005 IEEE CVPR '05, vol. 1*, pages 176–183, Washington, DC, USA, 2005.

[13] Q. Zhao and H. tao. A motion observable representation using color correlogram and its application to tracking. *Computer Vision and Image Understanding*, 113:273–290, 2009.