COLOR CONNECTEDNESS DEGREE FOR MEAN-SHIFT TRACKING

Michèle Gouiffès and Florence Laguzet and Lionel Lacassagne

Institut d'Electronique Fondamentale CNRS UMR 8622,
University of Paris 11, 91405 ORSAY cedex

ABSTRACT

This paper proposes an extension to the mean shift tracking. We introduce the color connectedness degrees (CCD) which, more than providing statistical information about the target to track, embeds information about the amount of connectedness of the color intervals which constitute the target. This approach, with a small increase of 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.

Index Terms— 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 [1, 2] 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 [3] 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 [4]. Some authors have improved the procedure either by introducing an objet/background classification [5] or by combining mean-shift with local approaches [6]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 objet, in order to enhance the discrimination ability of the histogram. Some authors have addressed that issue by proposing the spatiogram [7] and the correlogram [8]. 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 [9, 10], the object bounding box is spatially divided into regions or segments, to be processed separately. More recently, some new kernel methods [11, 12] 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 objet, based on the color connectedness degree (CCD). Initially designed for spatio-colorimetric classification in [13], that feature expresses the amount of connectedness of intervals of trichromatic colors.

The 3D CCD histogram is compared to the classical RGB 3D histogram on a few road sequences envolving 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 connectedness degree. 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 3D COLOR CONNECTEDNESS DEGREE

Undoubtedly, using 3D histogram instead of 1D, is necessary for a better discrimination ability. Indeed, two similar sets of 1D histograms can correspond to two different sets of colors. Let be a trichromatic image of components $\boldsymbol{c} = (c^1, c^2, c^3)$ and note $\boldsymbol{c}_i = ({c^1}_i, {c^2}_i, {c^3}_i)$ the color components of a pixel i of location $\boldsymbol{p}_i$. A color interval of size s^3, the origin of which is the color $\boldsymbol{c}_i$, is defined as:

$$ {I_i}^s = [c^1, c^1_i + s][c^2_i, c^2_i + s][{c^3}_i, {c^3}_i + s]. $$

The first order probability $\mathcal{P}_1(I^s_i)$ is the probability that a pixel of color $\boldsymbol{c}_a$ belongs to the cubic interval I^s_i. It is computed as the sum of the first order probabilities $\mathcal{P}_1(\boldsymbol{c}_a)$ of the components $\boldsymbol{c}$ belonging to I^s_i:

$$ \mathcal{P}_1({I_i}^s) = \sum_{\boldsymbol{c}_a \in I^s_i} \mathcal{P}_1(\boldsymbol{c}_a) \qquad (1) $$

The density of probabilities $\mathcal{P}_1({I_i}^s)$ is nothing but the classical 3D histogram, which bins have a size s. Now, we define

the co-occurrence probability of two colors $(\boldsymbol{c}_a, \boldsymbol{c}_b)$ as

$$\mathcal{P}_{cc}(\boldsymbol{c}_a, \boldsymbol{c}_b) = \frac{1}{8} \sum_{\boldsymbol{c}_a \in \mathcal{N}(\boldsymbol{c}_b)} \mathcal{P}_{oc}(\boldsymbol{c}_a, \boldsymbol{c}_b) \tag{2}$$

where $\mathcal{P}_{oc}(\boldsymbol{c}_a, \boldsymbol{c}_b)$ is the probability that $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$ are the colors of two neighbor pixels in the sense of 8-connectedness, the neighborhood being noted as $\mathcal{N}$. The second order probability $\mathcal{P}_2({I_i}^s)$ of the color interval ${I_i}^s$ is computed as the sum of the co-occurrence probabilities of all color couples $(\boldsymbol{c}_a, \boldsymbol{c}_b)$ belonging to ${I_i}^s$:

$$\mathcal{P}_2({I_i}^s) = \sum_{\boldsymbol{c}_a \in {I_i}^s} \sum_{\boldsymbol{c}_b \in {I_i}^s} \mathcal{P}_{cc}(\boldsymbol{c}_a, \boldsymbol{c}_b) \tag{3}$$

Therefore, the connectedness degree of a color cubic interval $\mathcal{D}({I_i}^s)$ is given as:

$$\mathcal{D}({I_i}^s) = \frac{\mathcal{P}_2({I_i}^s)}{\mathcal{P}_1({I_i}^s)} \tag{4}$$

This color connectedness degree is higher when the interval ${I_i}^s$ corresponds to connected components in the image, *i.e* to a meaningful class in the sense of connectedness. It is maximum when all colors of ${I_i}^s$ belong to a same connected component. The more there are regions, the lower the CCD. Then, contrary to the correlogram or to the color histogram, a small (but perhaps salient) homogeneous region can have a high CCD. Fig.1(a) to 1(c) show three synthetic images of size 16×16 with 4 equiprobable colors. The CCD are quite different ((a): 160, (b): 80, (c): 12), it is higher when a larger number of homogeneous pixels are connected.

(a) (b) (c)

Fig. 1. Illustration of the Color Connectedness Degree. (a) to (c) show 3 color images of size 16×16 with identical first order probabilities; the four colors are equiprobable. The CCD are significantly different from an image to the other: (a) 160, (b) 80, (c) 12.

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 defined as the CCD histogram:

$${H^0}_{\boldsymbol{u}} = \{\mathcal{D}(I^s_{\boldsymbol{u}})\} \quad \sum_{I^s_{\boldsymbol{u}}} {H^0}_{\boldsymbol{u}} = 1 \tag{5}$$

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

$${H^k}_{\boldsymbol{u}}(\boldsymbol{p}^k) = \{\mathcal{D}(I^s_{\boldsymbol{u}})\} \quad \sum_{I^s_{\boldsymbol{u}}} {H^k}_{\boldsymbol{u}} = 1 \tag{6}$$

The similarity between the target model at initial location and the target candidate at location $\boldsymbol{p}^k$ is computed as the similarity between those two CCD densities. Similarly to the initial mean shift algorithm, the Bhattacharyya similarity is chosen :

$$\rho(\boldsymbol{p}^k) = \rho[H^0_{\boldsymbol{u}}, H^k_{\boldsymbol{u}}(\boldsymbol{p}^k)] = \sqrt{H^0_{\boldsymbol{u}} H^k_{\boldsymbol{u}}(\boldsymbol{p}^k)} \tag{7}$$

It has the advantage not to produce singularities when density values are null. The candidate location which maximizes (7) 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 object / background. To solve those issues, each pixel of $\mathcal{W}$ is weighted by an isotropic kernel $K(\boldsymbol{p})$ which affects 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 (7). The target histogram is then computed as:

$$H^0_{\boldsymbol{u}}(\boldsymbol{p}^k) = \sum_{\boldsymbol{p}_i \in \mathcal{W}} K(\boldsymbol{p}_i) \mathcal{D}(I^s_i) \delta(\boldsymbol{c}_i - \boldsymbol{u}) \tag{8}$$

We chose the Epanechnikov kernel [3]. In addition, in order to better reduce the contribution of the background in the reference histogramm $H^0_{\boldsymbol{u}}$, the colors belonging to the bacground are subtracted from the histogram using the log-likelihood ratio of foreground/background as in [6]. In our paper, the target model is not updated during the sequences.

3.3. Mean Shift procedure

Considering a given target model $H_{\boldsymbol{u}}$ 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 (7) to the model. The Bhattacharyya distance is expanded in Taylor series as in [3] 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 CCD are computed at that location $H^k_{\boldsymbol{u}}(\boldsymbol{p}_0)$, as well as the similarity $\rho[H^k(\boldsymbol{p}_0), H^0]$.
2. The new candidate location $\boldsymbol{p}^k$ is computed:

$$\boldsymbol{p}^k = \frac{\sum_{i \in \mathcal{W}} \boldsymbol{p}_i \, w_i \, g\left(\left\| \frac{\boldsymbol{p}_0 - \boldsymbol{p}_i}{h} \right\|^2 \right)}{\sum_{i \in \mathcal{W}} \boldsymbol{p}_i \, w_i \, g\left(\left\| \frac{\boldsymbol{p}_0 - \boldsymbol{p}_i}{h} \right\|^2 \right)} \text{ with } g(x) = -k'(x) \tag{9}$$

Fig. 2. Sequences used in the experiments.

with the following definition of the weights derived from the Taylor expansion:

$$w_i = \sum_{\boldsymbol{u}} \sqrt{\frac{H^0_{\boldsymbol{u}}}{H^k_{\boldsymbol{u}}(\boldsymbol{p}^k)}} \delta(\boldsymbol{c}_i - \boldsymbol{u}) \tag{10}$$

3. while $\rho[H(\boldsymbol{p}^k), H^0] < \rho[H(\boldsymbol{p}^0), H^0]$ do $\boldsymbol{p}^k = 0.5(\boldsymbol{p}^k + \boldsymbol{p}^0)$
4. if $\|\boldsymbol{p}^k - \boldsymbol{p}^0\| < \epsilon$ 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 [3], 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 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 higer than a threshold T_{out}.

4. EXPERIMENTS

We experiment the robustness and accuracy of the CCD histogram versus 3D color histogram on 5 sequences showing vehicles or pedestrians. The first and last frames of these sequences are shown on Fig. 2 on firest and second row, with the CCD 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 $\boldsymbol{p}_2 = (250, 360)$.
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 size of the color intervals is $s = 32$, so the size of the RGB and CCD histograms is $8 \times 8 \times 8$. Fig. 3 to 7 show the picture of the objects tracked, with the classical MS (first row) and CCD MS (second row). In *Car 1* of fig.3, the classical MS loses the target, contrary to the CCD MS. 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 CCD histogram is obviously more time consuming than the RGB classical one. Therefore our tracking algorithm is generally more time consuming. However, that is not so significative. Since times depend on the target size, number of iterations, etc, we compare the relative time increase of the CCD MS to the classical MS, in table 1, in $\%$. In most cases our algorithm is more time consuming, from 20 % to 32%. However, it is more efficient on the *Car 2* sequence. Indeed, since the CCD histogram is more representative of the target, it needs a lower number of iterations to converge.

5. CONCLUSION

The paper proposed an extension to the mean-shift classical tracker by introducing the histogram of color connectedness degrees. That feature is high for the colors which correspond to connected components in the target to track, independtly from the size of the connected component. It proves to be

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

Fig. 3. The *Car 1* sequence. The classical MS tracker (1^{st} row) fails. The CCD MS trackes the car during the whole sequence.

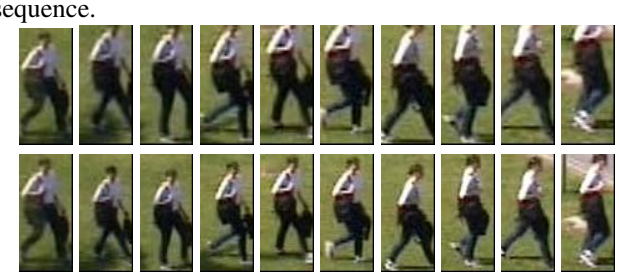

Fig. 4. The *Pedestrian 1* sequence.

Fig. 5. The *Pedestrian 2* sequence.

Fig. 6. The *Car 2* sequence.

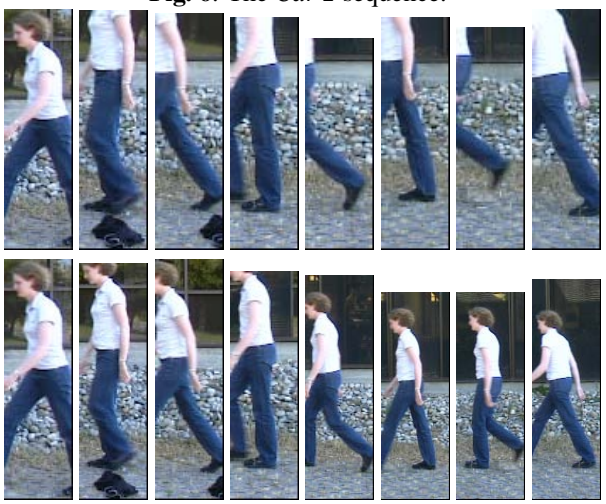

Fig. 7. The *Pedestrian 3* sequence.

Table 1. Increase of CPU times of the CCD histograms compared to the color 3D histograms.

Sequence	% CPU time
Car 1	+32
Pedestrian 1	+29
Pedestrian 2	30
Car 2	-30
Padestrian 3	+20

more informative and discriminative compared to the classical 3D histogram. Therefore the tracking results are improved in terms of robustness and in terms of quality, with a low increase of the computation times.

6. REFERENCES

[1] B.D. Lucas and T. Kanade, "An iterative image registration technique," in *Internation Joint Conf. on A.I.*, August 1981, pp. 674–679.

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

[3] D. Comaniciu and, "Kernel-based object tracking.," *IEEE Trans. PAMI*, vol. 25, no. 5, pp. 564–577, 2003.

[4] 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*, Washington, DC, USA, 2005, pp. 176–183.

[5] 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*, vol. 3, no. 3, pp. 2778–2790, 2009.

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

[7] S.T. Birchfield and S. Rangarajan, "Spatiograms versus histograms for region-based tracking.," in *Computer Vision and Pattern Recognition*, 2005, pp. 1158–1163.

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

[9] 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.

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

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

[13] M. Fontaine, L. Macaire, and J-G. Postaire, "Unsupervised segmentation based on connectivity analysis," in *International Conference on Pattern Recognition*, Barcelona, Spain, 2000, vol. 1, pp. 600–603.