AUTOMATIC COLOR SPACE SWITCHING FOR ROBUST TRACKING

Florence Laguzet[1] *and Michèle Gouiffès*[2] *and Lionel Lacassagne*[3] *and Daniel Etiemble*[1]

(1) Laboratoire de Recherche en Informatique CNRS UMR 8623
(2) Institut d'Electronique Fondamentale CNRS UMR 8622,
Université de Paris-Sud 11, 91405 ORSAY cedex
(3) CEA, LIST, Gif sur Yvette F-91191

ABSTRACT

This paper introduces an algorithm to automatically and continuously select the most appropriate color space to use in order to improve the performances of visual tracking. Eight color spaces are tested, and the Mean-Shift (MS) tracker is considered. The selection of the colorspace is made using an evaluation criterion based on the quality of the weights involved in the MS tracking, and implicitly on the good separability between the target and its close background. Experiments on real sequences show the impact of the color space on tracking performances and the relevancy of the proposed selection criterion.

Keywords: Kernel-based tracking, Mean-Shift algorithm, Colorspace selection.

1. INTRODUCTION

Visual tracking procedures, widely applied in video surveillance, robotics or games, *i.e* in dynamic and changing environments, can suffer from several issues among which: non-rigid motion of the target, geometric and photometric variations of target and/or background, occlusions. Even if the target representation is carefully carried out (with histogram, shape or direction vectors, color patches) in order to allow a satisfying separability with respect to the background, performances strongly depend on the color space model.

The kernel-based methods, for instance the Mean Shift (MS) tracking [1] represents the target with a global statistical representation based on color and/or texture, to provide, to some extent, robustness to non-rigid motion of the objects. However, its good resistance against tricky deformations of the target is counterbalanced by a poor separability with the background. This issue has lead to several contributions: comparison of different similarity functions [2], object and/or background classification [3], improvement of the target representation [4] or association with local approaches [5]. In [6], several histograms from different views are used to improve the modelization of the target.

While color clustering is a domain in which the selection of the best color space model (CSM) has been studied [7], few studies deal with their impact on tracking performances. RGB or HSV are generally chosen : RGB because no conversion is required after the image is captured; HSV because of the robustness of *Hue* against illumination changes [8] and its good understanding by the human perception.

Less works have been achieved concerning the automatic selection and switching of the color representation online during the tracking. Most studies are carried out for skin colors [9, 10] the modeling of which has been thoroughly studied.

A more generic method has kept our attention in [11]. The paper defines and compares two different criteria in order to select the N best linear combinations of RGB color features. The first criterion, based on the two-class variance ratio of the log-likelihood function evaluates the feature discriminability between the target and its background. The second criterion reduces the potential distraction produced by another object appearing in the neighborhood of the target and having similar colors. It relies on the difference between the peaks of log-likelihood of the target and the distractor respectively.

Differently from [11] where the criteria are used separately because of times costs, the present paper defines a heuristic which allows to select the most appropriate CSM by allowing: 1) a good *color separability* between the target and the background color distributions; 2) a good *object consistency*, *i.e* a good quality of the weights involved in the MS tracking.

In addition, a strategy is proposed to automatically run the CSM switching when the representation of the target has significantly changed since its last update.

Since the N color features are chosen independantly in [11], they might convey redundant information. In our work, the switch is directly achieved among existing CSMs. Therefore, $N = 3$ and, except in some peculiar cases, each color channel conveys a complementary piece of information. In addition, while in [11] the colors are considered separately, our contribution uses 3D quantized histograms in order to allow a better discrimination by preserving the tight relation between color channels.

The continuation of the paper is structured as follows. Section 2 explains and comments the main elements of the mean-shift procedure. Then, Section 3 discusses the classical CSM in terms of performances. The CSM switching proce-

dure is described in 4. To conclude, Section 5 asserts the relevance of the proposed method by comparing the robustness of our technique in real conditions.

2. MEAN SHIFT PROCEDURE

For concision purposes, detailed explanations on the MS tracker are not recalled here. More information is indeed available in the seminal paper [1]. First, the colorimetric and spatial representations of the target are detailed, then the tracking procedure is explained.

2.1. 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 $\widehat{\mathbf{q}}$ in the initial frame 0 is a 3D $\boldsymbol{m}$-bin histogram. The target candidate in frame k, called $\widehat{\mathbf{p}}(\boldsymbol{x}^k)$, has a bounding box noted $\mathcal{W}^k$ centered on the pixel $\boldsymbol{x}^k$. The similarity between the target model at initial location and the target candidate at location $\boldsymbol{x}^k$ is computed as the similarity between their respective color distributions. Similarly to the initial mean shift algorithm, the Bhattacharyya similarity is chosen:

$$\rho(\boldsymbol{x}^k) = \rho(\widehat{\mathbf{p}}(\boldsymbol{x}^k), \widehat{\mathbf{q}}) = \sum_{u=1}^{m} \sqrt{\widehat{\boldsymbol{p}_u}(\boldsymbol{x}^k).\widehat{\boldsymbol{q}_u}} \qquad (1)$$

The candidate location which maximizes (1) is found by proceeding a gradient-based optimization.

Note that the histogram is generally quantized in order to reduce the computational costs and to allow real-time execution. In addition, as defined by equation (1), the similarity measure between histograms is based on a bin-by-bin product. Therefore, when the histogram is nearly empty after quantization, the similarity measure might vanish as soon as the color distribution changes because of photometric changes for example. Such a phenomenon has to be detected in order to update the model or to switch the current CSM.

2.2. 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 $K(\boldsymbol{x})$ 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, $K(\boldsymbol{x})$ provides a finite smoothing kernel (Epanechnikov kernel is chosen here) for the gradient-based minimization (1). The target histogram is then computed as:

$$\widehat{\boldsymbol{p}_u}(\boldsymbol{x}^k) = C \sum_{\boldsymbol{x}_i \in \mathcal{W}} K(\boldsymbol{x}_i)\delta(\boldsymbol{c}_i - \boldsymbol{u}) \qquad (2)$$

where C is the classical normalization coefficient as in [1].

If a color appears on both the object and its vicinity or background, this color is not relevant for tracking because it reduces their separability. Therefore a lower confidence has to be granted to that color. In order to better reduce the contribution of the background in the reference model $\widehat{q}_u^0$, colors belonging to the background are subtracted from the histogram using the log-likelihood ratio of foreground/background as in [5]. A color $\boldsymbol{u}$ in the target is kept in the model when the following log-likelihood ratio is high enough:

$$L_{\boldsymbol{u}} = log\frac{\max(h_o(\boldsymbol{u}), \epsilon)}{\max(h_b(\boldsymbol{u}), \epsilon)} \qquad (3)$$

where h_o and h_b are the histograms of the object and background respectively and ϵ a small value put to avoid zero at denominator. This is an important part of the algorithm as some background colours are part of the target selection. Indeed, our automatic colorspace selection favors the color representations that produce a larger color distance between the background and the target.

2.3. Mean Shift procedure

Considering a given target model $\widehat{\boldsymbol{q}}_{\boldsymbol{u}}$ and the previous location of the object $\boldsymbol{x}_{k-1}$ in previous frame $k-1$, the tracking consists in finding in each frame the candidate location $\boldsymbol{x}_k$ which maximizes the similarity (1) to the model. The Bhattacharyya distance is expanded in Taylor series as in [1] 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 $\boldsymbol{x}_0$, is such that $\boldsymbol{x}_0 = \boldsymbol{x}_{k-1}$. The candidate histogram is computed at that location $\widehat{\boldsymbol{p}}_{\boldsymbol{u}}^k(\boldsymbol{x}_0)$, as well as the similarity $\rho[\widehat{\boldsymbol{p}}^k(\boldsymbol{x}_0), \widehat{\boldsymbol{q}}^0]$.
2. The new candidate location $\boldsymbol{p}^k$ is computed:

$$\boldsymbol{x}^k = \frac{\sum_{i\in\mathcal{W}} \boldsymbol{x}_i\, w_i\, g\left(\left\|\frac{\boldsymbol{x}_0 - \boldsymbol{x}_i}{h}^2\right\|\right)}{\sum_{i\in\mathcal{W}} w_i\, g\left(\left\|\frac{\boldsymbol{x}_0 - \boldsymbol{x}_i}{h}^2\right\|\right)} \text{ with } g(x) = -K'(x) \qquad (4)$$

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

$$w_i = \sum_{\boldsymbol{u}} \sqrt{\frac{\widehat{\boldsymbol{q}}_{\boldsymbol{u}}^0}{\widehat{\boldsymbol{p}}_{\boldsymbol{u}}^k(\boldsymbol{x}^k)}}\delta(\boldsymbol{c}_i - \boldsymbol{u}) \qquad (5)$$

3. while $\rho[\widehat{\boldsymbol{p}}^k(\boldsymbol{x}^k), \widehat{\boldsymbol{q}}^0] < \rho[\widehat{\boldsymbol{p}}^k(\boldsymbol{x}^0), \widehat{\boldsymbol{q}}^0]$ do $\boldsymbol{x}^k = 0.5(\boldsymbol{x}^k + \boldsymbol{x}^0)$
4. if $\left\|\boldsymbol{x}^k - \boldsymbol{x}^0\right\| < \epsilon$ then stop, otherwise $\boldsymbol{x}^0 \leftarrow \boldsymbol{x}^k$ and go to step 2.

Scale change of h is managed in a similar fashion as [1].

Note that equation (4) can produce an accurate location result when :

- the weights w_i are well defined in (5), *i.e.* the colorspace has to be chosen to be robust against photometric changes and/or the reference colorspace has to be updated in order to provide a good adaptability to appearence changes

- the number of significant weights (at least positive) is large enough in equation (4) in order to be relevant.

Therefore, the criterion defined for the CSM selection will take these remarks into account.

3. DISCUSSION ABOUT THE CSM

The eight color spaces chosen are well-known in the literature:
- Primary coordinates RGB and their L_1 normalization rgb, XYZ;
- Chrominance-luminance spaces Lab, HSV, YUV and YCbCr;
- Otha independant components.

In addition to the separability problem which is tackled in the next section, the performances of the color representation can be discussed beforehand in terms of dimensions of the color histogram and photometric robustness.

3.1. Dimensions of the histogram

As histograms are central in the algorithm, their dimension particularly affects the execution time. 3D histogram is used in our work. Indeed, 1D or 2D histograms require less resources but do not comprehensively represent the tight relation between color channels.

In order to reduce the amount of data to be processed, the image dynamics (initially 256) is quantized linearly with q steps. q has to be chosen depending on a trade-off between accuracy and execution times. In the current paper, $q = 16$.

3.2. Robustness against photometric changes

All color spaces which do not separate luminance and chrominance are very sensitive to illumination changes and their three components are likely to be impacted. Consequently, it also has a huge influence on the histogram distribution as well as on the distance function (1) which has to be computed at several steps of the algorithm. Note that the selection procedure described in next section does not encourage the invariant CSM (for instance rgb, HSV, YUV or $YCbCr$) more than any other representation. Indeed, the reference target is updated when needed in order to prevent from significant appearance changes. Our procedure favors the selection of the most discriminant CSM, even if less invariant to illumination changes, since it is likely to improve the tracking performances in most image sequences.

4. CSM SELECTION AND MODEL UPDATE

As evoked previously, several constraints have to be taken into account in the choice of a good color representation. When performing the quantization, it becomes obviously more difficult to distinguish between colors.Therefore, by using an appropriate color conversion before quantization and background subtraction, it is possible to improve the color distinction between target and its background and therefore to improve the MS tracking.

This section describes the criteria defined to determine a *good CSM for tracking*. The last section explains the procedure designed to automatically decide when to run the CSM switching and the model update.

4.1. Criteria of a good CSM for tracking

Two criteria have been studied for the selection of an appropriate CSM:
- good distinction between the object and its background;
- significant number of weights for an accurate tracking.

4.1.1. Separability between the target and the background

Since the Battacharrya distance is chosen as the similarity measure in the MS procedure, the same distance could be used to evaluate the similarity between the target and the background.

Consider $\widehat{\mathbf{q}}$ the target model and its bounding box $\mathcal{W}$ centred at initial pixel $\boldsymbol{x}^0$. Assume also that a log-likelihood ratio background subtraction (BS) has been performed on that model [5]. First of all the candidate model $\widehat{\mathbf{p}}(\boldsymbol{x}^0)$ can be computed at the same location $\boldsymbol{x}^0$, without any BS. As the target has not moved since the computation is made on the same picture, finding the best CSM consists in finding the color space which maximises the similarity (1) between the two histograms noted $\mathbf{q}_{\boldsymbol{u}}^0$ (with BS) and $\mathbf{p}_{\boldsymbol{u}}^0$ (without BS) computed in the same frame.

However, a good separability between the object and its background is not enough to guarantee a correct and accurate tracking. Indeed let us consider an object and a background containing different but close colors. After quantization and histogram normalization, the resulting colors can become perfectly similar leading to similar histograms.

In addition, the background subtraction can eliminate some representative colors of the target, when a similar color occurs in the background. Therefore, it might be important to evaluate the amount of color pixels which can actually be used in the tracking, after BS. That can be done by evaluating the quality and number of the weights w_i involved in the equation (5).

4.1.2. Amount of weights for the accurate computation of the new location

As noticed previously in section 2.3, the accuracy of the tracking in (4) relies on the accurate computation of the weights w_i

for the maximum number of the pixels belonging to the target. In other words, it is necessary to be sure that the BS does not eliminate a too large quantity of pixels from the target (*i.e* make all w_i vanish), otherwise the tracking procedure is likely to fail.

On the other hand, a too large number of positive weights w_i could pervert the result given by (4). Indeed, this phenomemon can occur when a distractor appears in the vicinity of the target. Note that a distractor is an object of which the color distribution is approximately similar to the target objet. Such a situation can also occur when the chosen CSM does not correctly separates the object colors from the background's.

As a consequence, a criterion has to be defined to evaluate the quality of the weights. To that aim, two histograms computed in the current frame: the histogram of the target computed with BS (noted $\mathbf{q}_{\boldsymbol{u}}^0$ as it is build the same way as the target model) and the histogram of the target without BS (noted $\mathbf{p}_{\boldsymbol{u}}^0$ as it is used as the candidate model). Then, the weights w_i are computed as in (5):

$$w_i = \sum_{\boldsymbol{u}} \sqrt{\frac{\widehat{\boldsymbol{q}}_{\boldsymbol{u}}^0}{\widehat{\boldsymbol{p}}_{\boldsymbol{u}}^0}} \delta(\boldsymbol{c}_i - \boldsymbol{u}) \tag{6}$$

Finally, the criterion is defined as the average computed in the window $\mathcal{W}$ around the target:

$$\mu_{w_i} = \frac{\sum_{\mathcal{W}} w_i}{N_w} \tag{7}$$

where N_w is the number of positive weights in $\mathcal{W}$. As the candidate $\mathbf{p}_{\boldsymbol{u}}^0$ is computed at the same location and in the same picture as the target $\mathbf{q}_{\boldsymbol{u}}^0$, the most discriminative CSM will provide the best μ_{w_i}.

4.2. Automatic selection of the CSM

The previous sections have presented the two criteria involved in the final selection criterion. As explained previously, relying on the background similarity or the color difference of the histograms can not guarantee the good tracking performances in some cases.
With support of several experiments, it was found that the best criterion to use is the average of weights described in equation (7). Indeed, it provides information about the separability between target/candidate while taking into account the resulting weights used in the MS optimization stage.

4.3. Switching procedure

In many existing works, the model is periodically updated arbitrarily every n frames. However, updating the model without any precaution can be risky. Indeed, unless the tracking procedure is perfectly robust, the update can occur at a bad time, for example when the target is partly occluded or when it brutally changes due to temporary illumination changes due to camera acquisition.
In the current work, a procedure is used to determine when the model needs to be updated or when the CSM needs to be changed.

In each frame, the following information about the target is stored: 1) the current distance ρ between the model and the current target; 2) the sum S_w of the corresponding weights w_i in $\mathcal{W}$; 3) the number of positive weights N_w.

The switching decision is described in Fig.1 (it has to be read from the left to the right). When a scale change occurs [1], the model is updated only when the current ρ becomes lower than the previous one in the previous frame. Otherwise, a procedure is run to decide whether the model has to be updated or not

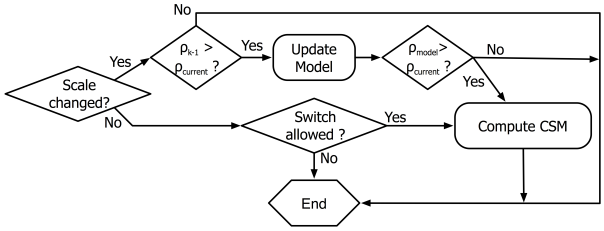


Fig. 1. The switching decision.

A switch is allowed when each of the three indicators ρ, N_w and S_w decreases under 70 % of the value they had at the previous update. Indeed, our experiments have shown that the use of only one criterion frequently leads to some unnecessary changes during the image sequence. In addition, the procedure checks that the ratio N_w/S_w remains under 1, *i.e* N_w and S_w have to decrease in a similar way.

When the switching is allowed, the new CSM is chosen by the criterion (7) and the model is updated. If the CSM selection leads to the same CSM as previously, the model is also updated.

5. EXPERIMENTS

In order to evaluate the decision of the selection algorithm, the Mean-Shift procedure was run with each of the height CSM and our automatic switch. The test was realized with three sequences of pedestrian [2] with quantization $q = 16$. For these sequences and for each of the nine tracking method, the results are displayed on thumbnails which represent the target (see Fig.2, 4 and 5). Concerning our method, the values of each criterion are shown for each CSM switch (see Tab.1, 2 and 3) and the best value is written in bold characters. Fig. 6 shows the last frame of each sequence in order to compare the behavior of the 9 trackings (each color window relates to one of the method).

[1] Let us recall that the scale change is achieved in a similar fashion as [1]

[2] Sequence overview and picture in ppm format can be found at http://www.lri.fr/ flaguzet/sequences.html

Citycam 2 - Pedestrian 1. This sequence shows a person who crosses the street. Pictures are not colourful, which makes MS tracking non-trivial. In addition, the tree acts as a distractor for most of the CSM (because of quantization). Note that none of the eight CSM are able to follow the target while our method performs correctly (see Fig.2 and Fig. 6). The values of our criterion allows to switch the CSM two times (excluding the first selection in frame 176) during the sequence, at picture 202 and 266 (Tab.1). Fig.3 shows the weights w_i computed at the call of the CSM switch. Obviously, our method choose one of the CSM for which the weights are well representative of the target. Indeed, most of the positive weights are correctly located on the target.

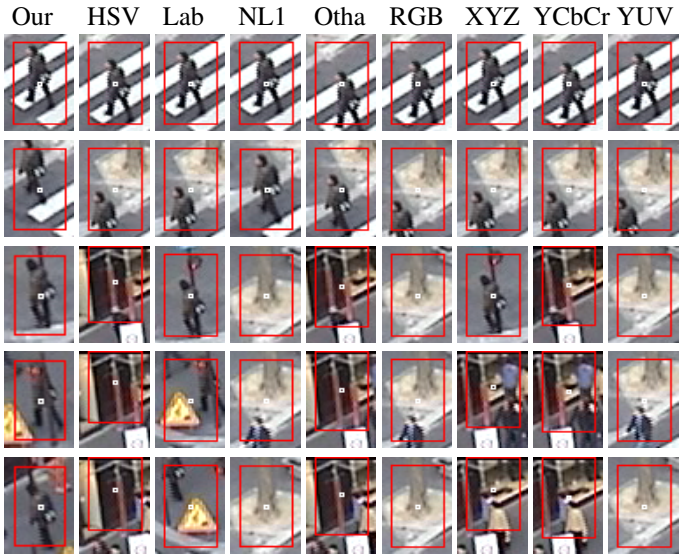


Fig. 2. Citycam 2 ped 1: tracking result *k*= 215, 260, 305, 350, 390

Table 1. Citycam 2 ped 1: criterion values at each CSM switch

	hsv	**lab**	**nl1**	**otha**	**rgb**	**xyz**	**ycbcr**	**yuv**
176	0.30	0.11	**2.52**	0.15	0.09	0.17	0.11	0.04
202	0.46	1.22	**2.20**	1.51	1.15	0.19	1.31	0.96
266	0.32	0.83	0.22	0.16	0.33	**1.20**	0.12	0.24

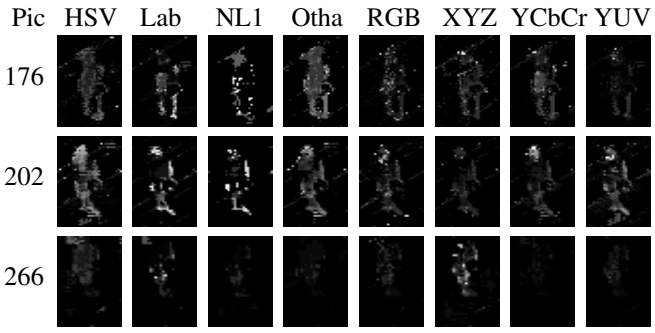


Fig. 3. Citycam 2 ped 1 : weights at CSM selection

Citycam 2 - Pedestrian 2. This sequence comes from the same video *Citycam 2*. The target is a pedestrian with RGB values close to the street colors. Since the target is poorly colored, almost half of the CSM failed for tracking the target (see Fig.4 and Fig. 6): it is the case for Otha, which loses it at the beginning, Lab and YCbCr which lose the target because of the same distractor. Finally, the rest of the CSM can be split into two classes depending on their results: 1) the CSM with a lack of precision in the position or scale: Nl1, RGB and YUV; 2) the CSM with a good accuracy: HSV, XYZ and our method.

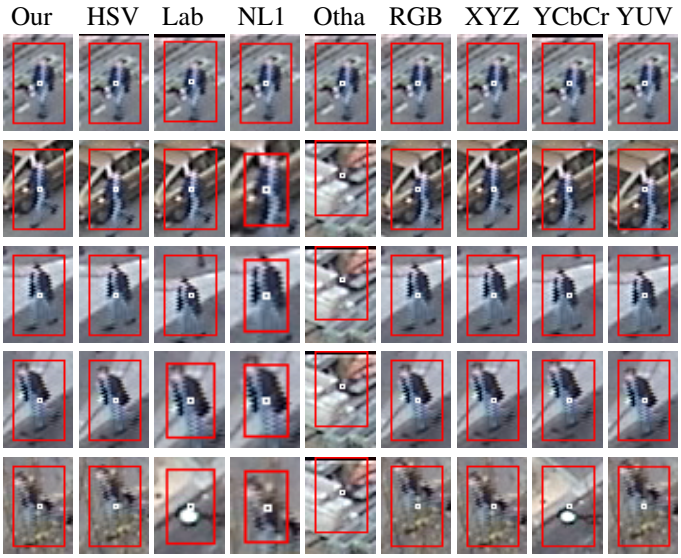


Fig. 4. Citycam 2 ped 2 : tracking result *k*= 180, 280, 380, 480, 545

Table 2. Citycam 2 ped 2: criterion values at each CSM switch.

	hsv	**lab**	**nl1**	**otha**	**rgb**	**xyz**	**ycbcr**	**yuv**
176	0.52	0.11	0.09	0.15	0.31	**1.09**	0.24	0.16
545	0.41	0.05	0.08	0.14	0.16	1.02	0.98	**1.20**
567	0.18	0.03	0.04	0.06	**0.19**	0.14	0.07	0.03

Pets 1. This sequence is known to be relatively difficult for tracking. First, the target is small, *i.e* of bad definition. Second, even if the camera is fixed, the background changes all along the sequence because the person walks in front of several cars of different colors, generating a lot of potential distractors.

Progressively, less and less CSM are able to follow the target (see Fig.5 and Fig.6) since they are distracted by the background composed of one or more colors which are in the target histogram (Fig.5). Our CSM switching method is automatically called twice (in addition to the first CSM selection) at pictures 472 and 523 (Tab.3). At the end, the tracking has failed for almost all of the CSMs except for YUV which focuses on the lower part of the body. Despite an accurate scale, our method provides an accurate position on the center of the target (see Fig.5).

Fig. 5. Pets1 : tracking result for pictures 350, 450, 550, 650

	hsv	lab	nl1	otha	rgb	xyz	ycbcr	yuv
118	0.83	0.13	0.69	0.55	0.37	0.80	0.79	**0.86**
472	1.11	0.19	**1.57**	1.16	1.02	1.17	1.12	0.98
523	1.13	**3.70**	0.23	1.24	0.27	1.04	1.11	1.17

Table 3. Pets 1: criterion values at each CSM switch.

6. CONCLUSION AND PERSPECTIVES

A procedure was proposed for automatically selecting and switching to a discriminative CSM in the context of MS tracking. Three non-trivial tracking sequences were introduced to demonstrate the good performance of our method.

In future works, the impact of each criterion will be studied to find a more accurate way to switch CSM and to better update the target model in order to prevent the false scale changes. Moreover, the quantization could also be determined automatically in order to decrease the computational time and complexity of the procedure.

ACKNOWLEDGMENTS : This research is supported by the European Project ITEA2 SPY[3].

7. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. on PAMI*, vol. 25, pp. 564–577, 2003.

[2] C.Yang, R.Duraiswami, and L.Davis, "Efficient mean-shift tracking via a new similarity measure," in *IEEE Computer Society*, 2005, pp. 176–183.

[3] S.Rastegar, M.Bandarabadi, Y.Toopchi, and S.Ghoreishi, "Kernel based object tracking using metric distance transform and rvm classifier.," in *AJBAS (3)'09*, 2009, pp. 2778–2790.

[4] M. Gouiffès, F. Laguzet, and L. Lacassagne, "Color connectedness degree for mean shift tracking," in *IEEE International Conference on Image Processing (ICIP'10)*, Istanbul, 2010.

[3] Surveillance imProved sYstem http://www.ppsl.asso.fr/spy.php

[5] R.Venkatesh Babu and A.Makur, "Kernel-based spatial-color modeling for fast moving object tracking.," in *ICASSP (1)'07*, 2007, pp. 901–904.

[6] I.Leichter, M.Lindenbaum, and E.Rivlin, "Mean shift tracking with multiple reference color histogram.," in *CVIU (114)'10*, 2010, pp. 400–408.

[7] N. Vandenbroucke, L.Macaire, and J.-G. Postaire, "Unsupervised color texture features extraction and selection for soccer images segmentation," in *IEEE International Conference on Image Processing (ICIP'00)*, Vancouver (Canada), 2000, vol. 2, pp. 800–803.

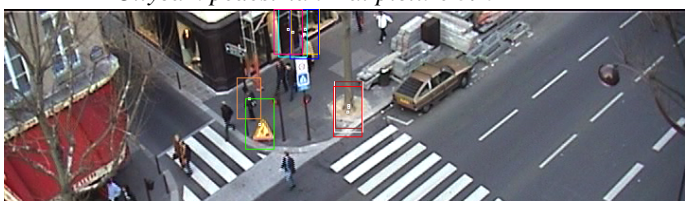
[8] T. Gevers and A.W.M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[9] H. Stern and B. Efros, "Adaptive color space switching for face tracking in multi-colored lighting environments," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, 2002, FGR '02, pp. 249–.

[10] D-Y Huang, W-C Hu, and M-H Hsu, "Adaptive skin color model switching for face tracking under varying illumination," in *Proceedings of the 2009 Fourth International Conference on Innovative Computing, Information and Control*, Washington, DC, USA, 2009, ICICIC '09, pp. 326–329.

[11] R.T.Collin, Y.Liu, and M.Leordeanu, "Online selection of discriminative tracking features," *IEEE Transactions on PAMI*, 2005.

Citycam pedestrian 1 at picture 390



Citycam pedestrian 2 at picture 545

Pets 1 pedestrian at picture 640

Each square represents final position for each CSM : HSV (deep pink), Lab (green), NL1 (brown), Otha (cyan), RGB (white) , XYZ (yellow), YCbCr (blue), Yuv (red) and our method (sandy brown).

Fig. 6. Overview of the tracking in the last picture of each sequence.