

Choix automatique de l'espace couleur pour le suivi robuste d'objets

Florence LAGUZET¹, Michèle GOUIFFÈS², Lionel LACASSAGNE³, Daniel ETIEMBLE¹

¹Laboratoire de Recherche en Informatique CNRS UMR 8623, Université de Paris-Sud, 91405 ORSAY, France

²Institut d'Électronique Fondamentale CNRS UMR 8622, Université de Paris-Sud, 91405 ORSAY cedex, France

³CEA, LIST, Gif sur Yvette F-91191 Cedex, France

florence.laguzet@lri.fr, michele.gouiffes@u-psud.fr
lionel.lacassagne@cea.fr, daniel.etiemble@lri.fr

Résumé – Ce papier présente un algorithme sélectionnant automatiquement l'espace couleur le plus approprié afin d'améliorer la qualité du suivi. Huit espaces couleurs sont testés et l'algorithme de suivi considéré est Mean-Shift (MS). La procédure de sélection est basée sur la maximisation de critères basés sur : 1) la discrimination entre la cible à suivre et l'arrière plan, 2) les poids utilisés dans le suivi avec MS. Les temps de conversion des différents espaces couleurs considérés sont également étudiés. Des expériences sur séquences réelles montrent l'impact de la sélection de l'espace couleur sur les performances du suivi ainsi que la pertinence des critères de sélection proposés.

Abstract – This paper introduces an algorithm to automatically 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 procedure is based on the maximization of a criterion jointly based on: 1) on the discrimination between the target and the background, 2) the weights involved in the MS tracking. The time costs of the color conversions are also discussed. Experiments on real sequences show the impact of the color space on the tracking performances and the relevancy of the proposed selection criterion.

1 Introduction

Les procédures de suivi d'objets, largement utilisées dans la vidéo surveillance, la robotique ou les jeux (*i.e* dans des environnements dynamiques et changeants) peuvent souffrir de plusieurs problèmes comme le déplacement non rigide de la cible, des changements géométriques et/ou photométriques de l'objet ou du fond, d'occlusions. Même si la représentation de la cible est soigneusement construite (avec des histogrammes, sa forme, des vecteurs de directions, les patch de couleurs) afin de séparer au mieux l'objet du fond, les performances dépendent fortement du modèle de représentation couleur.

Les méthodes basées sur l'utilisation de *kernels*, par exemple le trackeur Mean Shift (MS) [2], consistent à représenter la cible avec une statistique globale basée sur la couleur et/ou la texture pour être robuste, dans une certaine mesure, aux changements non rigide de l'objet. Toutefois, cette bonne résistance aux déformations de la cible est contre-balançée par une mauvaise séparabilité entre l'objet et le fond. Cela a conduit à plusieurs contributions : utilisation de plusieurs fonctions de calcul de similarité entre histogrammes [11], classification de l'objet et/ou de l'arrière plan [8], représentation de la cible [4] ou encore association avec des approches locales [1]. Dans [6], plusieurs histogrammes issues de plusieurs vues de l'objets sont utilisés afin d'améliorer la modélisation de la cible.

Le clustering de couleur est un domaine où la sélection de l'espace de représentation couleur a été étudiée [10]. Cependant, peu de recherche ont été menées sur l'impact de ce dernier

sur les performances des méthodes de suivi. *RGB* ou *HSV* sont souvent les espaces choisis : *RGB* car il ne nécessite aucune conversion après l'acquisition des images et *HSV* de par la robustesse de la teinte (*Hue*) contre les changements d'illumination [3] et car il est proche de la vision humaine.

Peu de travaux ont été menés concernant la sélection automatique de l'espace de représentation couleur. La plupart des études sont réalisées pour les couleurs de peau [9, 5], dont la modélisation a été minutieusement étudiée.

Une méthode plus générique a retenu notre attention dans [7]. Cet article définit et compare deux différents critères afin de sélectionner les N meilleures combinaisons linéaires des composantes de *RGB*. Le premier critère, basé sur le ratio de deux classes de variances de la fonction de log-vraisemblance, évalue les caractéristiques de discriminabilité entre la cible et son environnement. Le second critère réduit les potentiels distractions produits par d'autres objets apparaissant dans le voisinage de la cible et ayant les mêmes couleurs. Il s'appuie sur la différence entre les pics de log vraisemblance de la cible et du distracteur respectivement.

Puisque les N premières caractéristiques de couleurs sont choisies indépendamment dans [7], ils peuvent transmettre des informations redondantes. De plus, $N = 3$ et, sauf dans certains cas particuliers, chaque composante couleur fournissent des informations complémentaires. Dans notre travail, le choix est effectué directement vers des espaces de représentation couleur existants. Alors que dans [7] les couleurs sont considé-

rées séparément, nous utilisons des histogrammes 3D quantifiés afin de permettre une meilleure discrimination en préservant l'étroite relation existant entre les composantes couleurs.

L'article est structurée comme suit. La Section 2 explique les principes du suivi Mean-Shift. Ensuite, la Section 3 compare les espaces couleurs choisis et introduit notre méthode de sélection de l'espace couleur à utiliser. Pour conclure, la Section 4 valide la pertinence de la méthode proposée en comparant la robustesse de notre technique dans des conditions réelles.

2 Procédure de suivi Mean-Shift

Par souci de concision, les explications détaillées de l'algorithme de suivi MS ne sont pas rappelées ici. Plus d'informations sont disponibles dans [2].

Tout d'abord, la représentation colorimétrique et spatiale de la cible sont détaillées. Ensuite, la procédure de suivi est expliquée.

Représentation colorimétrique de la cible. La cible à suivre est généralement représentée par une boîte englobante \mathcal{W} , résultant d'un algorithme amont tel que l'analyse de mouvement, la stéréovision ou encore la reconnaissance de forme. Une fois détecté, le modèle de la cible $\hat{\mathbf{q}}$ dans l'image initiale 0 est un histogramme 3D à m -bin. La cible candidate à l'image k , appelée $\hat{\mathbf{p}}(\mathbf{x}^k)$, possède la boîte englobante appelée \mathcal{W}^k centrée sur le pixel \mathbf{x}^k . La similarité entre le modèle de la cible à sa position initiale et la cible candidate à la position \mathbf{x}^k est calculée comme une similarité inter-histogrammes. La distance de Bhattacharyya de l'algorithme original [2] a été choisie :

$$\rho(\mathbf{x}^k) = \rho(\hat{\mathbf{p}}(\mathbf{x}^k), \hat{\mathbf{q}}) = \sum_{u=1}^m \sqrt{\hat{\mathbf{p}}_u(\mathbf{x}^k) \cdot \hat{\mathbf{q}}_u} \quad (1)$$

La localisation du candidat qui maximise (1) est trouvée en effectuant une optimisation basée sur la descente de gradient. Généralement, les histogrammes sont quantifiés afin de réduire le cout de calcul et ainsi permettre l'exécution en temps réel. De plus, lorsque les histogrammes sont presque vides, cela peut fausser le calcul de similarité par exemple lorsque la distribution des couleurs est modifiée après un changement photométrique. Dans la section 3, un critère de choix de l'espace est défini afin de choisir un espace couleur offrant une grande variété de couleurs.

Représentation spatiale de la cible. La qualité du suivi Mean-Shift peut pâtir des occlusions partielles et de la mauvaise séparation entre l'objet à suivre et l'arrière plan. Pour répondre à ces problèmes, chaque pixel de \mathcal{W} est pondéré par un noyau isotrope $K(\mathbf{x})$ qui donne une plus grande importance à la partie centrale de \mathcal{W} où l'objet est le plus susceptible d'être situé. De plus, $K(\mathbf{x})$ fournit un noyau de lissage (le noyau choisi ici est celui d'Epanechnikov) pour la minimisation à base de gradient (1). L'histogramme du modèle est calculé de la manière suivante :

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

où C est le coefficient de normalisation comme dans [2]. Si une couleur apparaît dans les alentours de l'objet ou dans le fond, cette couleur n'est pas pertinente pour aider au suivi car cela augmente la ressemblance entre l'objet et le fond. Afin de mieux réduire l'influence de ce dernier dans l'histogramme du modèle $\hat{\mathbf{q}}_u^0$, les couleurs appartenant au fond sont soustraites de l'histogramme en utilisant la méthode de la log vraisemblance présentée dans [1].

La procédure Mean-Shift. Considérons le modèle de la cible $\hat{\mathbf{q}}_u$ et la localisation de l'objet \mathbf{x}_{k-1} dans l'image précédente $k-1$. Le suivi consiste à trouver, dans chaque image, la localisation de l'objet candidat \mathbf{x}_k qui maximise la similarité (1) avec le modèle. La distance de Bhattacharyya est développée en série de Taylor comme dans [2] afin d'appliquer l'optimisation à base de gradient. Ci-après sont présentées les différentes étapes de l'algorithme :

1. Initialement, l'objet est considéré comme immobile d'où sa localisation estimée, notée \mathbf{x}_0 , est telle que $\mathbf{x}_0 = \mathbf{x}_{k-1}$. L'histogramme candidat est calculé à l'emplacement $\hat{\mathbf{p}}_u^k(\mathbf{x}_0)$, tout comme la similarité $\rho[\hat{\mathbf{p}}^k(\mathbf{x}_0), \hat{\mathbf{q}}^0]$.
2. La nouvelle position du candidat \mathbf{x}^k est calculée comme suit :

$$\mathbf{x}^k = \frac{\sum_{i \in \mathcal{W}} \mathbf{x}_i w_i g\left(\left\|\frac{\mathbf{x}_0 - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i \in \mathcal{W}} w_i g\left(\left\|\frac{\mathbf{x}_0 - \mathbf{x}_i}{h}\right\|^2\right)} \quad \text{avec } g(x) = -k'(x) \quad (3)$$

avec comme définition des poids dérivés du développement de Taylor :

$$w_i = \sum_u \sqrt{\frac{\hat{\mathbf{q}}_u^0}{\hat{\mathbf{p}}_u^k(\mathbf{x}^k)}} \delta(\mathbf{c}_i - \mathbf{u}) \quad (4)$$

3. **Tant que** $\rho[\hat{\mathbf{p}}^k(\mathbf{x}^k), \hat{\mathbf{q}}^0] < \rho[\hat{\mathbf{p}}^k(\mathbf{x}^0), \hat{\mathbf{q}}^0]$
Faire $\mathbf{x}^k = 0.5(\mathbf{x}^k + \mathbf{x}^0)$
4. **Si** $\|\mathbf{x}^k - \mathbf{x}^0\| < \epsilon$ **Alors** stop,
Sinon $\mathbf{x}^0 \leftarrow \mathbf{x}^k$ et retourner à l'étape 2.

Le changement d'échelle est géré de façon similaire à [2].

3 Choix de l'espace couleur

Dans cette partie, l'accent est mis sur la représentation couleur de la cible en termes de taille de l'histogramme et d'impact de l'espace couleur sur le suivi. Ensuite, les critères de sélection sont expliqués.

Représentation de l'histogramme. Comme les histogrammes sont le point central de l'algorithme MS, leur taille a un impact évident sur le temps d'exécution. Afin de réduire la quantité de données à traiter, la dynamique de l'image (initialement de 256) est quantifiée linéairement par q . Ce dernier doit être choisi en faisant le compromis entre précision et temps d'exécution.

Impact de l'espace couleur sur le suivi. L'histogramme de la cible contient les probabilités, pour chaque couleur, d'appartenir à l'objet. Si une couleur est présente dans l'objet mais

aussi dans l'arrière plan, elle n'est pas pertinente pour la définition de l'objet, d'où la nécessité d'y attribuer moins d'importance. Afin de construire cette distribution de probabilité, la méthode de soustraction de l'arrière plan par log vraisemblance est utilisée [1].

Lorsque la quantification est effectuée, la distinction entre couleurs est rendue plus difficile. De plus, la soustraction du fond risque d'éliminer certaines couleurs représentatives de l'objet lorsque ces dernières apparaissent dans l'arrière plan. Dans certains cas, ces deux problèmes peuvent faire échouer le suivi. En effectuant la conversion couleur approprié avant les étapes de quantification et de soustraction du fond doit permettre une meilleure distinction des couleurs entre l'objet et l'arrière plan et ainsi améliorer la qualité du suivi.

Une deuxième avantage de la conversion couleur est la possibilité d'apporter une meilleure robustesse du suivi face aux changements d'illumination. Plusieurs invariants couleurs ont été étudiés et proposés dans la littérature [3]. Certains d'entre eux ont été testés dans les présents travaux.

Sélection de l'espace couleur. Soit \hat{q} le modèle cible et sa boîte englobante \mathcal{W} centrée sur la position initiale x^0 . Admettons que la soustraction du fond (SF) a été effectuée. D'autre part, le modèle du candidat $\hat{p}(x^0)$ est calculé à la même position x^0 sans utilisation de SF.

La distance ρ de (1) entre ces deux distributions \hat{q} (avec SF) et \hat{p} (sans SF) est élevée lorsque l'espace couleur choisi entraîne une bonne discrimination entre la cible et le fond. Néanmoins, il est nécessaire de vérifier que la SF ne réduit pas trop fortement la consistance de l'objet c'est à dire le nombre de poids w_i non nuls impliqués dans le suivi (voir équation 3).

Ainsi, le critère de sélection proposé prend en compte deux mesures : la distance inter-histogrammes ρ et de la somme des poids w_i , appelée S_w , dans la fenêtre d'intérêt \mathcal{W} . Ces deux paramètres sont normalisés entre 0 et 1 par l'utilisation des coefficients α_1 et α_2 appropriés¹.

$$C_0 = \alpha_1 \rho + \alpha_2 S_w \quad (5)$$

Ainsi, l'espace couleur sélectionné est celui dont le critère C_0 est le maximum. Néanmoins, lorsque la cible est presque homogène colorimétriquement, avec des couleurs proches de celles de l'arrière plan, le critère C_0 est élevé. Pourtant, le nombre de couleurs subsistant de la soustraction du fond devient trop faible pour représenter l'objet. Il est alors nécessaire de prendre en compte la répartition des couleurs dans l'histogramme afin de s'assurer que la SF n'élimine pas une quantité trop élevée de couleurs dans la cible, sans quoi la procédure de suivi risque d'échouer. Pour prendre en compte cette contrainte, le critère de sélection doit faire intervenir le nombre de couleurs caractéristiques de la cible, une fois la SF effectuée. Dans ce but, le pourcentage de pixels R_T n'apparaissant que dans la cible est calculé :

$$R_T = 1 - \text{card}\{T \cap B\} / \text{card}\{T\} \quad (6)$$

1. Ils sont respectivement l'inverse de la distance maximum et du maximum des sommes de poids.

TABLE 1 – Citycam : valeurs du critère \mathcal{C} pour chaque espace couleur

	hsv	rgb	lab	NL1	otha	xyz	ycbcr	yuv
ρ	0.09	0.02	0.23	0.15	0.24	0.22	0.02	0.02
w_i	0.09	0.03	0.24	0.15	0.23	0.22	0.02	0.02
R_T	0.15	0.09	0.06	0.01	0.09	0.09	0.07	0.06
\mathcal{C}	0.13	0.01	0.32	0.03	0.52	0.45	0.00	0.00

où T et B expriment l'ensemble de couleurs constituant respectivement la cible et l'arrière plan. Le critère est ensuite défini tel que :

$$\mathcal{C} = C_0 R_T \quad (7)$$

4 Expériences

Afin d'évaluer la décision de l'algorithme de sélection, la procédure Mean-Shift a été exécutée avec chacun des espaces couleur. Ce test est illustré sur deux séquences de suivi de piéton² avec pour facteur de quantification $q = 16$. Afin d'évaluer la qualité du suivi les figures Fig.1 and Fig.2 montrent, pour les 8 espaces couleur, 4 imagerie centrées sur la cible suivie. Pour les deux séquences, on montre les valeurs des critères de choix de l'espace couleur (voir section 3) calculés dans la 1ère image de la séquences (voir Tab.1 et 4).

Séquence Citycam. Cette séquence présente une personne traversant une route sur un passage piéton avant de passer près d'un arbre. Ce dernier se comporte comme un distracteur dans la plupart des espaces couleurs. Dans la 1ère image, le critère (visible dans tab.1) a permis de sélectionner dans l'ordre : l'espace couleur de Otha, suivi par xyz et Lab. *RGB, HSV, YCbCr* et *YUV* ne sont pas performants : en effet, ils ne permettent pas le suivi de la cible tout au long de la séquence (Fig.1). Ces résultats sont cohérents avec la faible réponse obtenue par ces espaces en sortie de la fonction de sélection (Tab.1). Pour les trois meilleurs espaces couleur mis en évidence par notre méthode de sélection (Otha, *XYZ, Lab*), le suivi s'est déroulé correctement et la localisation de la personne suivie est similaire pour les trois séquences résultats. L'utilisation de l'espace couleur *NL1* permet également de suivre la cible mais les résultats ont montré plusieurs *décrochages* du suivi. Pour chaque image de la séquence, la distance inter-histogrammes ρ et la somme des poids w_i sont enregistrés. 4 valeurs sont visibles dans les tableaux 2 et 3. Ces résultats montrent que l'espace couleur sélectionné par notre méthode donne la meilleure valeur de ρ tout au long de la séquence (Tab.2). En ce qui concerne la somme des poids pour l'espace Otha, elle est la meilleure dans la plupart des images de la séquence (Tab.3).

Séquence Pets³. Cette séquence est difficile pour le suivi car la cible suivie est petite, donc peu définie, et elle passe devant une file de voitures de différentes couleurs, ce qui engendre une modification continue du fond de la scène.

2. L'aperçu des séquences ainsi que les images au format ppm peuvent être trouvées à l'adresse : <http://www.lri.fr/~flaguzet/sequences.html>

3. The sequence is available here : <http://i21www.ira.uka.de/image-sequences/>



FIGURE 1 – Résultat du suivi pour les images 100, 260, 300, 340



FIGURE 2 – Résultat du suivi pour les images 237, 287, 337, 377

TABLE 2 – Citycam : valeurs de ρ pour chaque espace couleur

Pic	hsv	lab	nl1	otha	rgb	xyz	ycbcr	yuv
100	0.24	0.55	0.33	0.65	0.07	0.60	0.06	0.06
180	0.23	0.51	0.25	0.59	0.06	0.56	0.06	0.05
260	0.23	0.50	0.22	0.60	0.09	0.53	0.08	0.07
300	0.28	0.50	0.23	0.51	0.06	0.46	0.10	0.12

D’après le tableau 4, notre critère a permis de sélectionner HSV suivi d’assez loin par YUV. En s’intéressant aux résultats du suivi montrés Fig.2, on observe que la cible est perdue par tous les espaces couleur excepté HSV. De plus, on remarque sur la deuxième ligne (image 287) que la cible est perdue excepté pour HSV et YUV. Ces observations confirment les résultats fournis par notre heuristique de sélection.

5 Conclusion et Perspectives

Cet article a proposé une méthode de sélection automatique de l’espace couleur à utiliser pour le suivi d’un objet. Le critère proposé prend en compte une mesure de discrimination couleur entre l’objet et le fond, ainsi que la consistance de l’objet après soustraction du fond. Dans de futurs travaux, la procédure de sélection pourrait être effectuée de manière périodique car l’arrière plan change tout au long de la séquence. Le temps de calcul pourrait également être mieux inclus dans le critère de sélection.

TABLE 3 – Citycam : valeurs de S_w pour chaque espace couleur

Pic	hsv	lab	nl1	otha	rgb	xyz	ycbcr	yuv
100	305	575	317	693	82	660	68	71
180	332	590	246	671	75	690	58	65
260	396	529	190	656	172	603	125	101
300	543	470	174	629	80	519	165	183

TABLE 4 – Pets : valeurs du critère C pour chaque espace couleur

	hsv	rgb	lab	NL1	otha	xyz	ycbcr	yuv
ρ	0,22	0,07	0,02	0,14	0,14	0,07	0,16	0,19
w_i	0,24	0,07	0,02	0,13	0,12	0,07	0,18	0,18
R_T	0,24	0,13	0,00	0,06	0,09	0,13	0,03	0,14
C	1,22	0,06	0,00	0,10	0,15	0,06	0,08	0,46

Références

- [1] R. V. Babu and A. Makur. Kernel-based spatial-color modeling for fast moving object tracking. In *ICASSP (1)'07*, pages 901–904, 2007.
- [2] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on PAMI*, 25 :564–577, 2003.
- [3] T. Gevers and A. Smeulders. Color-based object recognition. *Pattern Recognition*, 32(3) :453–464, 1999.
- [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.
- [5] 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, ICICIC '09*, pages 326–329, Washington, DC, USA, 2009.
- [6] I. Leichter, M. Lindenbaum, and E. Rivlin. Mean shift tracking with multiple reference color histogram. In *CVIU (114)'10*, pages 400–408, 2010.
- [7] R. T. Collin, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *IEEE Transactions on PAMI*, 2005.
- [8] S. Rastegar, M. Bandarabadi, Y. Toopchi, and S. Ghoreishi. Kernel based object tracking using metric distance transform and rvm classifier. In *AJBAS (3)'09*, pages 2778–2790, 2009.
- [9] H. Stern and B. Efron. 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, FGR '02*, pages 249–, Washington, DC, USA, 2002.
- [10] 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)*, volume 2, pages 800–803, Vancouver (Canada), 2000.
- [11] C. Yang, R. Duraiswami, and L. Davis. Efficient mean-shift tracking via a new similarity measure. In *IEEE Computer Society*, pages 176–183, 2005.