People Tracking using Color Control Points and Skin Color

K.Laaroussi
[1]LIIAN
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco
Laaroussi.khadija@yahoo.fr

A.Saaidi
[1]LIIAN, [2]LIMAO
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco.
[2]Faculty polydisciplinary of Taza, Sidi Mohamed Ben Abdellah University, B.P 1223, Taza, Morocco.
saaidi.abde@yahoo.fr

K.Satori[1]
[1]LIIAN
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco
khalidsatorim3i@yahoo.fr

***Abstract*—This paper presents a new approach to be used for people detection and tracking in image sequences based on color control points and skin color modeling. This method aims to track these people in complex situations such as players on a soccer field. Each person in the image is represented by several control points which are obtained using a color version of the Harris algorithm detector. Each control point is characterized by the local appearance which is a vector of local characteristics. Then we determine the rules that define the skin regions in three different color spaces such as RGB, HSV and YCbCr, and we apply these rules to our images to segment skin regions. Using a set of control points and skin regions allows us to track a person by matching control points based on the measure of ZNCC correlation «Zero mean Normalized Cross Correlation». The simulations and experimental results show the robustness of our algorithms in terms of stability and convergence. Performance is illustrated by some examples. Thus, our method fits well with the noise conditions.**

***Index Terms*—Color Control Points, Skin Color, Appearance Model, Matching, People Detection, People Tracking.**

I. INTRODUCTION

Detecting and tracking people in image sequences is a difficult task for several reasons namely the complex movements of people followed, the human body is highly articulated, the illumination of the scene changes can result in non-consistency pixel values representing a person, as well as image noise and phenomenon of occlusion (the person itself, with the other moving objects or objects in the background). However, the use of people tracking is pertinent in many applications of computer vision such as human-computer interaction, civil and military surveillance, video conferencing or analysis of human movement.

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. A large number of tracking methods have been proposed; in fact most of them are based on region segmentation [16, 17] or people tracking based on the subtraction of the background [1] which simplifies subsequent processing by locating regions of interest in the image from a model of the environment and an observation. We also find algorithms that are based on the extraction of contours with a particle filtering using color information or a filter Calman [2, 3, 18, 21] or tracking based on mean shift [13, 19]. Other algorithms are based on active contours [20] or normalized cross correlation and color space [22, 23], optical flow [24].

Or algorithms which using a variable search window (VSW) algorithm based on color and feature points is proposed [28].

Recently, Gouet and Lameyre [14, 25] presented a tracker that uses control points and Snacks but with images in grayscale and for a single object in the scene. However, to the best of our knowledge, there papers [26, 31] presents a new approach for tracking objects in complex situations using color interest points characterized both by local descriptors and by a geometric model.

In this work, we are interested in tracking people in image sequences, or more specifically to track a certain number of points detected on these people; the main purpose is to reduce the maximum information of an image in some control points and some skin regions. Thus, a person is represented by several control points which are obtained with a color version of the Harris algorithm detector [4, 33]. Each control point is characterized by its local appearance which is a vector of local characteristics. Then we determine the rules that define the skin regions in three different color spaces such as RGB, HSV and YCbCr, and we apply these rules to our images to segment skin regions. Using a set of control points allows us to track a person by matching control points from image to image based on the measure of ZNCC correlation «Zero mean Normalized Cross Correlation» [9, 34, 35].

doi:10.4304/jetwi.6.1.94-100

The main contribution of our method is to track multiple people in a color video sequence containing complex scenes where people can have different size. Indeed, the combination of these two primitives will develop a tracking method implementing the time parameter and quality of tracking. However, skin color is a primitive often used as a first estimate of location and segmentation to reduce the search area. The color information is clearly relevant to the presence of people in video since they are very specific and allow fast algorithms invariant to changes in orientation and scale. The strong point of this method is that it is possible to achieve a more robust tracking by using only one of the two primitives considered i.e. if one of the two failed, it is possible to keep track of a person using the other and vice versa.

Figure 1 explains our treatment procedure for detection and tracking people in image sequences:

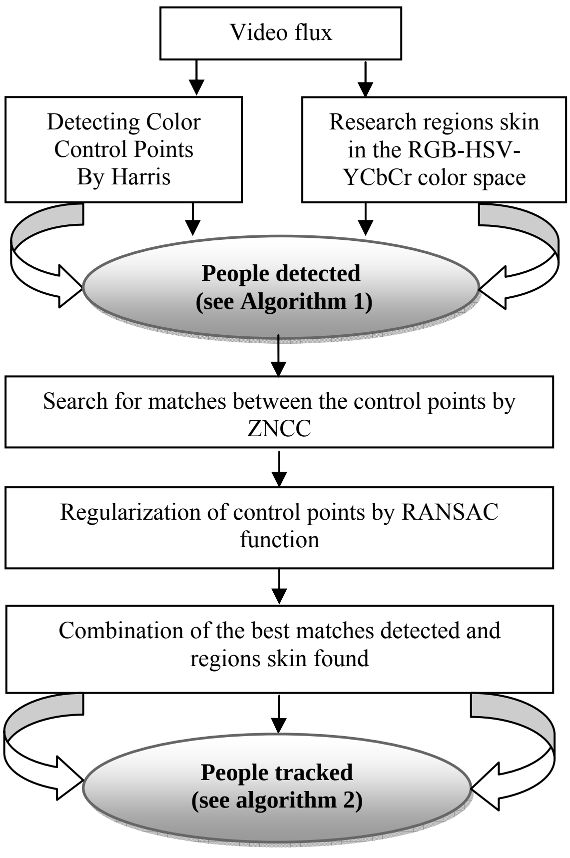


Fig1: Procedure for detection and tracking people.

This paper is organized as follows: in section 2 we describe the tracking tools, namely skin color detection, color version of Harris detector, and matching. In section 3, we present our proposed tracking system. In section 4, we show the results of our experiment. In section 5 we conclude with a summary.

II TRACKING TOOLS

2.1 Skin color Detection

The skin detection is to detect the pixels corresponding to human skin in a color image; including methods of detecting skin color are the explicit methods [5, 6, 12]. However, the skin color can be presented in different color spaces such as:

- **RGB Space:** This color space combines three color components (red, green and blue), it not need any model of skin and no color transformation [10, 15]. The threshold applied in this space is defined as: *(RGB*) is classified as a skin color pixel if:

(R>95) and (G>40) and (B>20) and (((max{R, G, B})-(min{R, G, B}))>15) and (ABS(R-G)>15) and (R>G) and (R>B)
(1)

- **YCbCr space:** In this space there is Y represents the luminance and Cb Cr represent the two components of chrominance (Blue and Red). This space separates the luminance and chrominance making YCbCr attractive for modeling the skin color [15]. The threshold in this area is as follows:

$$77 <= Cb <= 127$$
$$133 <= Cr <= 173 \qquad (2)$$

- **HSV space** *(Hue, Saturation, value*): Hue defines the dominant color in a region. The saturation of the color measurement of a region in proportion to its shine. The highlight of the use of this color space is the discrimination between the luminance and chrominance [15]. The Threshold in this region is as follows: [11]:

$$0 \le H \le 0.5$$
$$0.17 \le S \le 0.63 \qquad (3)$$

However, we can make a hybridization of two or more spaces to have good results.

2.2 Color Harris Detector

Control points are the locations of an image which contain the maximum information, there are involved in many applications, like stereovision, image retrieval or scene monitoring. Several control points detectors have been developed over the last two decades. Schmid and Mohr [7] compare the performance of several of them. The most popular control point detector is the Harris detector [8] with its adaptations [27, 29, 30]. While this detector only applies to grayscale images, Montesinos et al. [4] generalized it to color images. The control points produced by their detector are defined as the positive local extrema of the intermediate grayscale image.

$$R(x,y) = \det(M(x,y)) - k \cdot trace(M(x,y))^2 \qquad (4)$$

Where k is typically set to 0.04 and $M_{Color}(x,y)$ is the 2×2 matrix.

$$M_{Color} = \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_xR_y + G_xG_Y + B_xB_y \\ R_xR_y + G_xG_Y + B_xB_y & R_x^2 + G_x^2 + B_x^2 \end{bmatrix} \qquad (5)$$

With R, G and B are three components such as Red, Green, and Blue.

After the matrix calculation M_{Color} for each pixel, we eliminate the values that are lower to a single modifiable of response R and we make the extraction of the local maxima.

According to the comparisons made by Gouet and Boujmaa [32], the above detector appears to be the most stable among the popular color control points detectors with regard to illumination changes, noise with a Gaussian filter, rotation, and viewpoint changes.

2.3 Matching

Matching correlation ZNCC "Zero mean Normalized Cross Correlation" or centered normalized correlation [9, 34, 35] allows measuring the similarity of pixels between the two images by calculating the correlation between two windows. It provides a measure of the interval [-1, 1], the similarity between two points X and Y is high when ZNCC(X,Y) approaches to 1.

In the practice, we set a threshold S (in general S=0.5) such as the couples(X, Y) are considered the true correspondents if the following constraint is satisfied:

$$ZNCC\ (X, Y) > s \tag{6}$$

This function uses the following formula to measure the correlation between the points of images:

$$ZNCC\ (X, Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{X})^2 \sum_{i=1}^{n}(y_i - \overline{Y})^2}} \tag{7}$$

With: - $X = (x_i)_{i=1...n}$ and $Y = (y_i)_{i=1...n}$ the coordinates of points X and Y.

- $\overline{X}$ and $\overline{Y}$ are the averages of X and Y respectively such as:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

III TRACKING SYSTEM

The purpose of the proposed system is to track people in a color video sequence, for each image we apply the color Harris detector to extract the control points. Then we determine the rules that define the regions of the skin in three color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions. To keep track of people we are looking for connections between these points, calling the measure of correlation ZNCC.

The purpose of the proposed system is to track persons from frame to frame, for each image we use the Harris detector to extract color control points. Then we determine the rules that define the areas of the skin in three different color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions. To keep track of people we are looking for connections between these points, calling the measure of correlation ZNCC.

However, each control point is characterized by a vector containing some local attributes, it consists of three channels r, g and b plus Gaussian derivatives $r_x, g_x, b_x, r_y, g_y, b_y$, and we also add the local Cornerness R given by equation (4).

So, we characterize each control point j in frame n by the vector following characteristics:

$$V_j(n) = (r, g, b, r_x, g_x, b_x, r_y, g_y, b_y, R) \tag{8}$$

A person p in image n is then modeled by:

$$V^p(n) = \{V_j(n) / j = 1,...., P(n)\} \tag{9}$$

Where P(n) is the number of control points.

The basic problem is to find for each control point $i \in \{1,....,P(n-1)\}$ with $P(n-1)$ the number of control points detected in the image $n-1$, the best control point j which is corresponding with $j \in \{1,...., P(n)\}$ where $P(n)$ the number of control points for the image n. The search party is local and occurs in an area of research for the control point i is calculated all the correlation values on the corresponding epipolar line in the other image, the control point associated with the highest score is selected as appropriate (see figure 2).

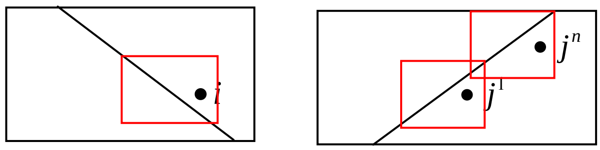


Fig2: Correlation windows

The degree of similarity between pairs $(i \quad j^1),\dots,(i \quad j^n)$ is calculated using a window centered at i and a window that moves on the points $j^1,\dots,j^n$, the robustness of these functions are classified according to their matching capabilities in the case of large displacement, change brightness, noise and low textured areas.

The performance of our tracking system is also closely linked with the results obtained in the detection of skin color. The segmentation procedure is done by scanning all the pixels of the image in search of those that follow the desired threshold in the space used. The pixel belonging to the range of desired threshold is 1 otherwise it is 0. This process produces a binary image highlighting segments with skin color (white) the other did not take the skin color black.

Our suggested method for tracking persons contains two algorithms, in the first; we propose a simple and effective method for detection of people in a sequence images based on the color control points detected by color Harris and skin color with the aim of eliminating the non- skin color regions. The second algorithm is to keep track of people detected by Algorithm 1 based on matching correlation ZNCC and the skin regions.

Algorithm 1: (People Detection)

Step1: *Image acquisition: Take a series of successive images by a digital camera or from a color video.*

Step2: *Apply color Harris detector for the first image* $n = 1$ *with the aim of identifying regions of interest detected on people.*

Step3: *Apply a skin detector in the YCbCr-RGB-HSV* *space to segment the image into skin regions and differentiate individuals with respect to other objects.*

Step4: *Apply a skin detector in the YCbCr-RGB-HSV space to segment the image into skin regions and differentiate people with respect to other objects.*

Step5: *Combination of the detected control points and regions skin found.*

Step6: *Show the detection result: people detected are surrounded by rectangles containing the found color control points and sections of skin.*

The result of Algorithm 1 allows us to initialize the algorithm 2, so the detection step serves only to initialize people tracking.

Algorithm 2: (People Tracking)

Step1: *Initialization :*
Image acquisition: Take a successive series of images containing people to track.
Apply Algorithm 1 for the first image $n = 1$ *and* $n = 2$.
Determine the characteristic vector for each detected point control $V^P(1)$ *and* $V^P(2)$.

Step2: *Apply Algorithm 1 for each image n and fill vectors* $V(n)$ *with appearance features.*

Step3: *Match the color control points detected by Harris using correlation measure ZNCC.*

Step4: *Regularization of control points detected by RANSAC function.*

Step5: *Combining the results of the best matches detected and regions skin found.*

Step6: *Show results of tracking by combining color control points and skin color, such as persons tracking is surrounded by rectangles.*

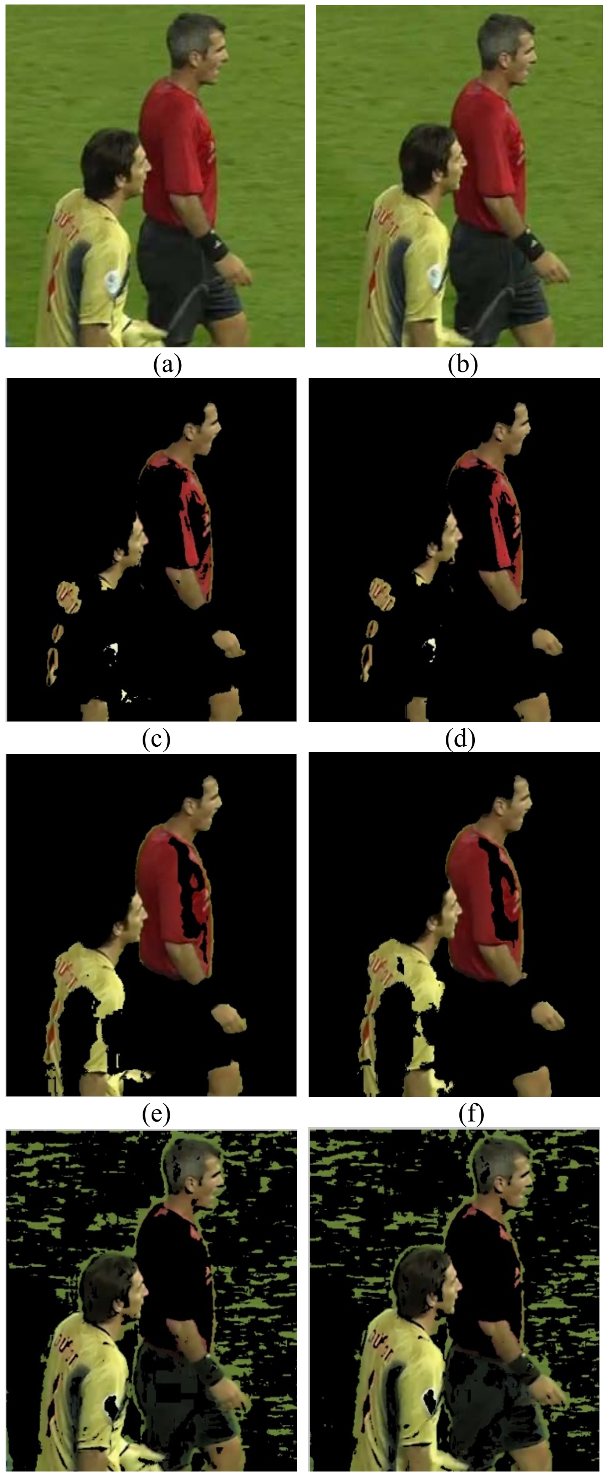
These two algorithms were tested on a large number of image sequences. The scenarios are varied, ranging from individual person walking alone to people in a crowd interacting (two or more people meet and walk side by side). The simulated results show that our algorithm is able to track more people, and therefore, detect and manage the interactions between them.

IV EXPERIMENTATION

To work on a real case with our method, we have chosen two successive images with 374×421 size taken from a football match video time of 4:13 min (two people who meet and walk side by side). First, we detected the control points by Harris algorithm [4, 33] and matched to each pair of images by measuring correlation ZNCC [9, 34, 35], at this stage several matches are false, where a regularization step is very important to keep just the best matches found by calling the function (RANSAC: Random Sample Consensus) [36]. Then we determine the rules that define the areas of the skin in the hybridization of these three color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions.

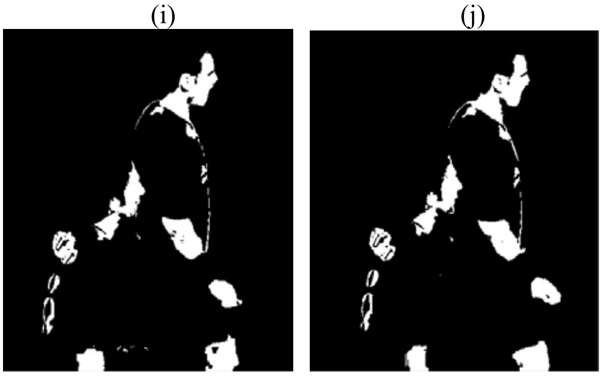
All algorithms used in this article have been implemented in Matlab 7 Intel Core (TM) 2 CPU 2.00GHz, Microsoft Windows 7 with 3GB memories.

Fig.3 shows an example for detecting skin color for these two images in the three spaces RGB, YCbCr and HSV.



(a) (b) (c) (d) (e) (f) (g) (h)

(i) (j)

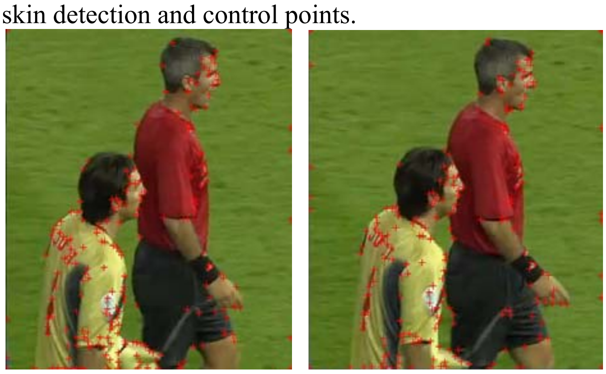


(k) (l)

Fig3: Example of skin detection. (a) and (b) two successive original image, (c) and (d) skin pixels in the RGB space, (e) and (f) skin pixels in the YCbCr space, (g) and (h) in the HSV space, (i) and (j) in the RGB-YCbCr-HSV space, (k) and (l) the binary images in the YCbCr-RGB-HSV space.

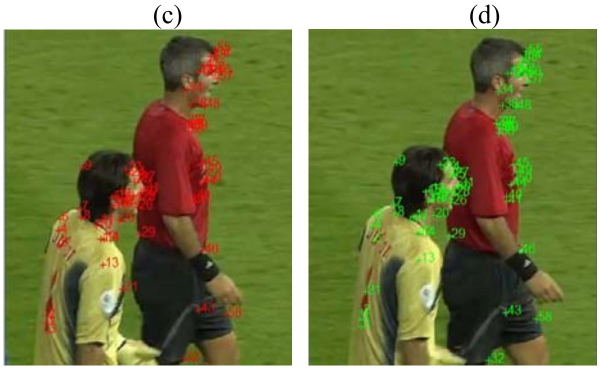
Based on these results, the detection of skin is based on the choice of the color space and the phase thresholding, and the hybridization of these three spaces RGB, YCbCr and HSV gives us a better result compared to others, we can nevertheless see a higher error rate for certain spaces (YCbCr, HSV). The execution time of the program is 1.039306 seconds for each image.

In Figure 4, we illustrate the result of tracking by applying the two algorithms 1 and 2. First of all, we detect the control points then we determine the skin regions. Tracking consists of matching control points detected on these two images and skin color. We can notice that our system is able to track these people, by detecting several skin areas, despite the existence of false skin detection and control points.

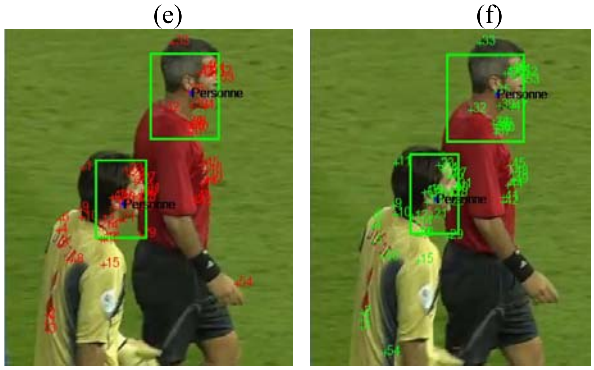


(a) (b)

(c) (d)



(e) (f)



(g) (h)

Fig4: Results of tracking using both algorithms 1 and 2. (a) and (b) the red and green points indicate control points detected by Harris detector, (c) and (d) indicate control points detected by correlation ZNCC, (e) and (f) regularization of control points detected by RANSAC function, (g) and (h) results of tracking based on the color control points and skin color in the RGB-YCbCr-HSV space such as persons tracking is surrounded by rectangles.

According to the results in Figure 4, the two people are correctly detected and tracked; the result of matching is best as pictures show (e) and (f) after the elimination of false matches by RANSAC function. The skin regions are specified and surrounded by boxes as the pictures show (g) and (h). The total running time of our system is 72.256917 seconds for the two images.

Our proposed tracking system has been tested on a large number of images from several snuff videos of different length sequences. The scenarios are varied, ranging from simple person walking alone several people interacting (two or more people meet and walk side by side). The experiment results show that our algorithm is able to detect and track multiple people in image sequences, and that this system is robust in the case of the existence of noise.

In summary, according to the simulations, only the proposed method can correctly track people, other methods are disturbed by noise or the presence of control points belong to the background and use of skin color allows increase the performance of our approach. By combining these two primitives will develop a tracking method implementing the time parameter and quality tracking.

V Conclusion

In this paper we have presented a new approach for tracking people in image sequences using color control points and skin color. A person is defined by a set of control point detected with version color Harris detector. Each control point is characterized by local appearance i.e. a vector of local characteristics more cornerness. Using a set of control points and regions skin allows us to track a person by making connections between these points from one image to another based on matching correlation ZNCC. The combination of control points with local descriptors has been successfully used for matching points and the color of the skin allows us to increase the performance of our proposed method, reducing areas of research.

References

[1] C. Stauffer and W.E.L.Grimson, Adaptive background mixture models for real-time tracking, Computer Vision and Pattern Recognition, pages 246–252, 1999.

[2] M.Isard and A. Blake. Condensation – conditional density propagation for visual tracking, International Journal of Computer Vision, pages 5–28, 1998.

[3] P.Li, T.Zhang, and A.E.C.Pece, Visual contour tracking based on particle filters, Image and Vision Computing, pages 111–123, 2003.

[4] P.Montesinos, V.Gouet, and R.Deriche, Differential invariants for color images, International Conference on Pattern Recognition, pages 838–840, 1998.

[5] M.HAMMAMI, Modèle de peau et application à la classification d'images et au filtrage des sites Web, L'ECOLE CENTRALE DE LYON, pages 69-80, 2005.

[6] F.Gianni, Suivi de parties du corps pour l'interprétation de gestes de communication à partir d'une séquence monoculaire, Université Paul Sabatier-Toulouse III, pages 7-10, 2010.

[7] C.Schmid, R. Mohr, and C.Bauckhage, Evaluation of interest point detectors, International Journal of Computer Vision, pages 151–172, 2000.

[8] J. Harris and M. Stephens, A combined corner and edge detector, 4th ALVEY Vision Conference, pages 147–151, 1988.

[9] A.Saaidi, H.Tairi et K.Satori, Fast Stereo Matching Using Rectification and Correlation Techniques, ISCCSP, Second International Symposium on Communications, Control and Signal Processing, Marrakech, Morrocco, March 2006.

[10] J.Kovac, P.Peer and F.Solina, 2D versus 3D color space face detection, 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, pages 449-454, 2003.

[11] S. Tsekeridou and I. Pitas, Facial feature extraction in frontal views using biometric analogies, Proc. of the IX European Signal Processing Conference, vol. I, pages 315-318, 1998.

[12] H.BOUIROUGA, S.ELFKIHI, A.JILBAB, D.ABOUTAJDINE, Skin detection in pornographic videos using threshold technique, Journal of Theoretical and Applied Information Technology, Vol.35 No.1, 15th January 2012.

[13] D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects using mean shift. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC, volume II, pages 142–149, June, 2000,.

[14] V. Gouet and B. Lameyre. SAP: A robust approach to track objects in video streams with snakes and points. British Machine Vision Conference, pages 737–746, 2004.

[15] V.VEZNEVETS, V.SAZONOV, A.ANDREEVA. A Survey on Pixel-Based Skin Color Detection Techniques. International Conference Graphicon, Moscow, Russia, 2003.

[16] M. Kim, J.G. Jeon, J.S. Kwak, M.H. Lee, and C. Ahn. Moving object segmentation in video sequences by user interaction and automatic object tracking. IVC, 19(5):245–260, April 2001

[17] S. Pateux. Tracking of video objects using a backward projection technique. In Visual Conference on Image Processing, volume 4067, pages 1107–1114, Australia, 2000.

[18] L.Peihua , W.Haijing , Object Tracking with Particle Filter Using Color Information, Third International Conference, MIRAGE, Rocquencourt, France, pages 534-541, 2007.

[19] F.Zheyi , L.Mo , L.Zhiwen. An Improved Video Target Tracking Algorithm Based on Particle Filter and Mean-Shift. International Conference on Information Technology and Software Engineering, pages 409-418, 2013.

[20] S.Lefèvre, N.Vincent. Real Time Multiple Object Tracking Based on Active Contours. International Conference, ICIAR, Proceedings, Part II, pages 606-613, 2004.

[21] Y.ZHAO, H.PEI. Object tracking based on particle filter with discriminative features. Journal of Control Theory and Applications, Volume 11, Issue 1, pages 42-53, February 2013.

[22] W.Jian; Y.Heng-jun; Yan-yan Cao ; Zhi-ming Cui. Video Object Tracking Method Based on Normalized Cross-correlation Matching. Distributed Computing and Applications to Business Engineering and Science (DCABES), pages 523 – 527, 2010.

[23] S.Patrick, V.Yap Vooi. Tracking Using Normalized Cross Correlation and Color Space. International Conference on Intelligent and Advanced Systems, pages 770 – 774, 2007.

[24] G. Castellano, J. Boyce, and M. Sandler. Regularized cdwt optical flow applied to movingtarget detection in IR imagery. Machine Vision and Applications, pages 277–288, 2000.

[25] V. Gouet and B. Lameyre. Object Tracking and Identification in Video Streams with Snakes and Points. Advances in Multimedia Information Processing - PCM 2004 Lecture Notes in Computer Science Volume 3333, , pages 61-68, 2005.

[26] P. Gabriel, J.-B. Hayet, J. Piater, and J. Verly. Object tracking using color interest points. Advanced Video and Signal Based Surveillance, AVSS. IEEE Conference on, pages 159–164, 2005.

[27] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 530–534, May 1997.

[28] H.Lim, D.Kang, Object tracking system using a VSW algorithm based on color and point features, EURASIP Journal on Advances in Signal Processing, September 2011.

[29] K.Mikolajczyk, C.Schmid. Indexing based on scale invariant interest points. In International Conference on Computer Vision, Vancouver, Canada, July 2001.
[30] David G. Lowe. Distinctive image features from scale-invariant keypoints. Accepted for publication int the International Journal of Computer Vision, 2004.
[31] P. Gabriel, J.-B. Hayet, J. Piater, and J. Verly. Utilisation des Points d'Intérêts Couleurs pour le Suivi d'Objets. Actes du Congrès ORASIS, pages 159–164, 2005.
[32] V. Gouet and N. Boujemaa. About optimal use of color points of interest for content-based image retrieval. Internal Report, INRIA Rocquencourt, 2002.
[33] J. Harris and M. Stephens. A combined corner and edge detector. 4th ALVEY Vision Conference, pages 147–151, 1988.
[34] M.Lhuillier and L.Quan. Quasi-dense reconstruction from image sequence. ECCV, 2002.
[35] S.Chambon, A.Crouzil, M.El miziani, G.Lemarie, P. Le callet. Complémentarité de mesures de corrélation pour la mise en correspondance: 4TReconnaissance des Formes et. Intelligence Artificielle4T”4T, RFIA, 2010.
[36] M.A.Fischler, R.C.Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing, Juin 1981.