World Applied Sciences Journal 32 (2): 289-301, 2014
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2014.32.02.343

Video-Surveillance System for Tracking Moving People Using Color Interest Points

[1]Khadija Laaroussi, [1,2]Abderrahim Saaidi,
[1]Mohamed Masrar and [1]Khalid Satori

[1]LIIAN, Department of Mathematics and Computer Science, Faculty of sciences,
Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco
[2]LIMAO, Department of Mathematics, Physics and Computer Science Faculty Polydisciplinary of Taza,
Sidi Mohamed Ben Abdellah University, B.P 1223, Taza, Morocco

Abstract: Local image features or interest points provide compact and abstract representations of patterns in an image. In this paper, we address the problem of detecting and tracking multiple moving people based on color interest points. The proposed method uses the statistical Gaussian Mixture Model (GMM) for the segmentation, extraction of moving people and background area. After that, from the detected foreground we determine the rules that define skin regions for good people detection. Color Interest Points are identified in the detected regions of skin using Harris algorithm. The use of an interest points set allows us to track people by matching these ones from image to image based on ZNCC correlation approach (*Zero mean Normalized Cross Correlation*). Finally, by calculating Euclidean distance between the best matches and other interest points detected on each consecutive images of video sequence, we can observe the motion of people tracked in the scene. Proposed results are obtained from two different types of videos, namely sport video and class video. The simulations and the experimental results show the robustness of our method to achieve the track with a good precision. The results are very encouraging, as well as, our proposed method fits well with noise conditions and contrast changes.

Key words: Video Surveillance · Gaussian Mixture Model · Moving People Detection · Color Interest Points · Skin color Segmentation · Correlation-based Matching · Feature Tracking · People Tracking · Euclidean Distance

INTRODUCTION

Automatic visual surveillance in dynamic scenes (both in indoor and outdoor environment) has recently got a considerable interest to researchers [1]. Technology has reached a stage where mounting video camera is cheap causing a widespread deployment of cameras in public and private areas [2]. Finding available human resources to sit and watch the imagery is too expensive for most organizations to afford the cost of human operators [3]. Moreover, surveillance by operators is error prone due to fatigue, negligence and lack of ubiquitous surveillance. Therefore, it's important to develop an accurate and efficient automatic video analysis system for monitoring human activity that will create enormous business opportunities. It will allow us to detect unusual events in the scene and warrant the attention of security officers to take preventive actions [2]. The purpose of visual surveillance is not to replace human eyes with camera, but to accomplish the entire surveillance task as automatic as possible [1]. Other applications of automatic video surveillance include preventing theft at parking and shopping areas [2], detecting robbery in bank and secured places [3], video conferencing or analysis of human movement.

Corresponding Author: Khadija Laaroussi, LIIAN, Department of Mathematics and Computer Science, Faculty of Sciences, Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco.

Detection and tracking of moving objects are important task for computer vision, particularly for visual-based surveillance system [4] and the literature on tracking objects is abundant. Several reference papers are available [5, 6]. Difficulties in tracking objects can arise due to abrupt object motion, changing appearance patterns of both the object and the scene, nonrigid object structures, object-to-object and object-to-scene occlusions, less of information caused by projection of the 3D word on a 2D word and camera motion. In general, the automatic video surveillance system has two major components, they are detecting moving objects and tracking them in a sequence of video images. The accuracy of these components largely affects the accuracy of overall surveillance system. Detecting moving regions in the scene and separating them from background image is a challenging problem. In the real world, some of the challenges associated with foreground object segmentation are illumination changes, shadows, camouflage in color, dynamic background and foreground aperture. Foreground object segmentation can be done by three basic approaches: frame differencing, background subtraction and optical flow. Frame differencing technique does not require any knowledge about background and is very adaptive to dynamic environments [3], but suffers from the problem of foreground aperture due to homogeneous color of moving object. Background subtraction can extract all moving pixels, but it requires perfect background modeling. It's extremely sensitive to scene changes due to lighting and movement of background object. Optical flow is the most robust technique to detect all moving objects, even in the presence of camera motion, but it's computationally expensive and cannot be used for real-time systems.

In general, tracking methods are firstly based on the choice of people representation and on a method of tracking. The possible representations are numerous [7], we can cite for example the centroid of person, set of interest points, primitives geometric shapes for example a rectangle or ellipse, silhouette and contour, articulated shape models which composed of body parts and skeletal models. In our case, we have chosen color interest points to represent people tracked, these later are particularly interesting because they concentrate the information, originally contained in the entire image, in a few specific points and they are considered more reliable sources of information than the contours because there is more stress on the intensity function, they are less sensitive to occultation since only a part of the descriptors is assigned, are also present in a large majority of images relative to the contours. Also, the skin color is a primitive often used as a first estimate of location and segmentation to reduce the search area. The color information is undoubtedly relevant to the presence of persons in video because they are very specific and allow for fast algorithms invariants to the changes in orientation and scale.

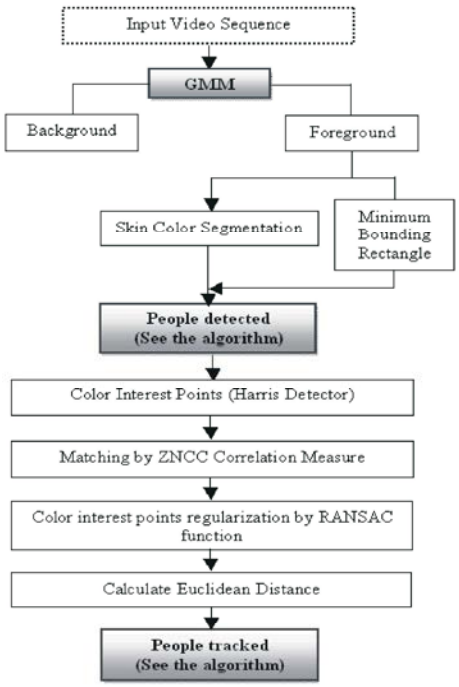

Fig. 1: Flowchart for the whole system.

A new approach will be discussed in this paper; it's an approach for detecting and tracking moving people in a sequence of images under the assumption that our camera is fixed. Our system is briefly as follows: foreground and background segmentation are performed by the statistical Gaussian Mixture Model (GMM) [8, 9] to extract moving people. After skin regions segmentation we detect color interest points on these ones using Harris algorithm [10, 11]. The use of an interest points set allows us to track people, by calling ZNCC correlation measure [12, 13]. After this step, several points are wrong pairings, so a regularization step is very important, in our experiments we used RANSAC function [14]. Best matches are identified by considering maximum correlation coefficient. Finally, by calculating Euclidean distance between the best matches and other interest points in each consecutive frames of video sequence, we can observe the motion of people tracked in the scene.

The main steps of our proposed tracking people system are summarized in the flowchart in Figure 1.

This paper is organized as follows. In section 2, we present some previous works in this area. Section 3 gives all mathematical formulation especially for skin color segmentation, detection and matching of color interest points and Gaussian Mixture Model. We illustrate our proposed method for tracking in section 4. The results of simulations and experiments are presented in section 5 and the conclusions are presented in section 6.

Related Works: The main task of tracking algorithms is to track moving objects from one frame to another in an image sequence. Tracking over time typically involves matching objects in consecutive frames using features such as points, lines or blobs. Tracking methods are divided into four major categories [1], region-based tracking, active-contour-based tracking, feature-based tracking and model-based tracking. It should be pointed out that this classification is not absolute in these algorithms different categories can be integrated together.

Region-based tracking algorithms [3, 5] track objects according to variations of the image regions corresponding to the moving objects. For these algorithms, the background image is maintained dynamically and motion regions are usually detected by subtracting the background from the current image. These regions are often modeled using a density distribution of the probability of their color. This distribution can be described with the help of a color histogram or a mixture of Gaussians. Other versions of these algorithms use Template Matching to increase the robustness of the tracking algorithm. These methods have a strong robustness for camera motion and background mutation; it also has a strong capacity to adapt to noise. Although they work well in scenes containing only a few objects. However, these methods also have some limitations: the sudden deformation of objects; the method cannot be followed quite true; it suffers from computational complexity, as it matches a window with all candidate windows in the next frames; and it cannot reliably handle occlusion between objects [1]. Other approaches take into account the spatial information thanks to many small regions and using time average of color by pixels [15]. Moving objects can also be modeled with a fixed form or input, as in the approach Mean-Shift [16] and for the particulate matter filters [17]. The disadvantages of these techniques are unable to describe an arbitrary form, which changes between consecutive images; have a low complexity which offers a general and reliable solution regardless of the characteristics representing the target object. Also, the mean shift fails at the monitoring of small and fast objects and at recovering a track after a total occlusion.

Active contour-based tracking algorithms track objects by representing their outlines as bounding contours and updating these contours dynamically in successive frames [18, 19]. These algorithms aim at directly extracting shapes of subjects and provide more effective descriptions of objects than region-based algorithms. Paragios *et al.* [20] detect and track multiple moving objects in image sequences using a geodesic active contour objective function and a level set formulation scheme. Peterfreund [21] explores a new active contour model based on a Kalman filter for tracking nonrigid moving targets such as people in spatio-velocity space. To overcome the limitations of association, the active contour algorithm can be combined with an algorithm describing color of object [22, 23], or with an interest points set, for example, in [24, 25] Gouet and Lameyre presented a method based on a model of appearance that is a feature vector for each interest point detected. The use of contours reduces the points to follow in a limited area for each image and the spatial description of this point is exploited for tracking contour. This approach is valid only for grayscale images and a single object in the scene. In contrast to region-based tracking algorithms, active contour-based algorithms describe objects more simply, more effectively and reduce computational complexity. Even under disturbance or partial occlusion, these algorithms may track objects continuously. However, precision of tracking is limited at the contour level. The recovery of the 3D pose of an object from its contour on the image plane is a demanding problem. A further difficulty is that the active contour-based algorithms are highly sensitive to tracking initialization, making it difficult to start tracking automatically.

Feature-based tracking algorithms [26-29] perform recognition and tracking of objects by extracting elements, clustering them into higher level features and then matching the features between images. Feature-based tracking algorithms can further be classified into three subcategories according to the nature of selected features: global feature-based algorithms, local feature-based algorithms and dependence-graph-based algorithms [1]. In all these methods, several features of blobs are used for matching. Such features could include

centroids, perimeters, areas, size, position, velocity, ratio of major axis of best-fit ellipse [2], line segments, curve segments, corner vertices, orientation, coordinates of bounding box, geometric relations between features etc. However, there are several serious deficiencies in feature-based tracking algorithms [1]:

- The recognition rate of objects based on 2D image features is low, because of the nonlinear distortion during perspective projection and the image variations with the viewpoint's movement.
- These algorithms are generally unable to recover 3D pose of objects.
- The stability of dealing effectively with occlusion, overlapping and interference of unrelated structures is generally poor.

Model-based tracking algorithms track objects by matching projected object models, produced with prior knowledge, to image data. The models are usually constructed off-line with manual measurement, or computer vision techniques. As model-based rigid object tracking and model-based nonrigid object tracking are quite different, we review separately model-based human body tracking (nonrigid object tracking) and model-based vehicle tracking (rigid object tracking). This provides position, movement of tracked object and at the same time gives a pose estimation of object. Authors of articles [6, 30-33] present methods of tracking parts of body like face, head and hands or body by constructing a model. However, tracking is carried out mainly in structured scenes where only one or two objects in motion are seen by the camera. Generally, there is a constraint on pixels number for moving objects, as well as requirements on composition of scene and object pixels number is more stringent, the computational complexity increases rapidly with the amount of details in the model, as well as correspondence to a model with an image requires expensive calculations.

Mathematical Formulas

Skin Color Segmentation: Skin color segmentation can be defined as the process of discrimination between skin and non-skin pixels. However, there are some difficulties in robustly detecting the skin color. The ambient of the light and shadows can affect the appearance of the skin-tone color. Moreover different camera produce different color values even from the same person and moving object can cause blurring of colors. Finally, people have varied skin color-tones individually [34] such as Asians skin gives big different with Caucasians skin type. A method for establishing a fast classifier of skin color is to explicitly define (by a certain number of rules) the terminals of skin regions in a color space. Nevertheless, the main part in skin color segmentation is to choose the suitable color space, as well as, the rules and thresholds of empirical decision. Several color spaces are used to label the pixels as pixels of skin color such as: RGB, RGB normalized, HSV (or HSI), YcbCr, YIQ. In addition, methods to construct a model of skin color are proposed: a method using the tone of pixel, method based on the histogram and a method using a Gaussian function.

There are several approaches for segmenting skin regions. Therefore, a simple, effective, robust and real-time method is the right choice for our work. Among the approaches used, we found the approach presented in [35], uses a combination of three color spaces RGB, HSV and YCbCr. The latter should be well suited for our work, with respect to its high rate of accuracy.

Detection and Matching of Color Interest Points: An interest points is a point in an image where significant changes occur, several detectors have been developed over the last two decades. Schmid and Mohr [36] compare the performance of many of them. The most popular is the Harris detector [10, 11] with its adaptations [10, 37-39]. While this detector only applies to grayscale images, Montesinos *et al*. [10] generalized it to color images. The interest points produced by their detector are defined as the positive local extrema of the intermediate grayscale image:

$$F = \det(G_{color}) - \alpha \cdot trace^2(G_{color}) \quad (1)$$

where α=0.04 (Harris parameter response) and G_{color} is the 2×2 matrix given by:

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

With R, G and B are three components of color such as Red, Green and blue. According to the comparisons made by Gouet and Boujemaa [40], the above detector appears to be the most stable among the popular color interest points detectors with regard to illumination changes, noise, rotation and viewpoint changes.

After the detection of color interest points by Harris algorithm [10, 11], appeal is made to one of matching methods which are looking for the pixels that are similar to

the ones using correlation measure. The search of what is locally corresponding is carried out in an area of research. Very many measures have been proposed to take into account the various difficulties encountered during this step such as noise or occultation. In this paper, we chose the extent of correlation ZNCC (*Zero mean Normalized Cross Correlation*) [12, 13]. The robustness of these functions is classified according to their capabilities to match in the case of large displacement and change of brightness.

Gaussian Mixture Model: A Gaussian Mixture Model *(GMM)* is a parametric probability density function represented as a weighted sum of Gaussian component densities [8, 9]. This method is suggested by Stauffer and Grimson, which models each pixel as a mixture of Gaussian distributions and uses an online approximation to update the model. The model assumes that each pixel in the frame is modeled by a mixture of K Gaussian distributions where different Gaussian distributions represent different colors. We consider the values of a particular pixel over time as a 'pixel process', i.e., a time series of scalars for grayvalues or vectors for color pixel values. At any time, t, what is known about a particular pixel, $\{x_0, y_0\}$, is its history:

$$\{X_1,\ldots,X_t\} = \{\mathrm{I}(x_0, y_0, i) : 1 \le i \le t\} \quad (3)$$

where I is the image sequence. We chose to model the recent history of the color features of each pixel, $\{\times_1,\ldots, \times_t\}$, as a mixture of K Gaussian distributions. So, each pixel is characterized by its intensity in the RGB color space. Then, the probability of observing the current pixel value is considered given by the following formula in the multidimensional case:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * f(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (4)$$

where K is the number of the distributions, $\omega_{i,t}$ is an estimate of the weight (what portion of the data is accounted for by this Gaussian) of the i^{th} Gaussian in the mixture at time t, $\mu_{i,t}$ is the mean value of the i^{th} Gaussian in the mixture at time t, $\Sigma_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time t, X_t is a random variable vector and where f is a Gaussian probability density function:

$$f(X_t, \mu_t, \Sigma) = \frac{1}{\sqrt{(2\Pi)^K |\Sigma|}} \exp\left[-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)\right] \quad (5)$$

K is determined by background multimodality, available memory and computational power, currently, from 3 to 5 are used. Also, for computational reasons, Stauffer and Grimson [8] assumed that the RGB color components are independent and have the same variances. So, the covariance matrix is of the form:

$$\sum_{K,t} = \sigma_k^2 I \quad (6)$$

So, each pixel is characterized by a mixture of K Gaussians. Once the background model is defined, the different parameters of the mixture of Gaussians must be initialized. The parameters of the GMM are the number of Gaussians K, the weight $\omega_{i,t}$ associated to the i^{th} Gaussian at time t, the mean $\mu_{i,t}$ and the covariance matrix $\Sigma_{i,t}$. The initialization of the weight, the mean and the covariance matrix is made using an EM algorithm *(expectation maximization)* [41]. However, this model present some disadvantages: the number of Gaussians must be predetermined, the need for good initializations, the dependence of the results on the true distribution law which can be non-Gaussian and slow recovery from failures. Others limitations are the needs for a series of training frames absent of moving objects and the amount of memory required in this step.

Proposed Method

Background Modeling and Foreground Object Segmentation: Moving objects detection is the basic step for further video analysis; it's a very important step because the following steps will be based on its results. Most of the works on foreground object segmentation is based on three basic methods, namely frame differencing, background subtraction and optical flow [28]. Only background subtraction requires modeling of background. It's faster than other methods and can extract maximum features pixels. In [3], Collins *et al*. used a hybrid of frame differencing and background subtraction for effective foreground segmentation. The basic operation of these methods is the separation of moving objects (foreground) and the static objects (background). In the literature, a lot of works has been done on modeling dynamic background, researchers usually use a single Gaussian [5, 6], a mixture of Gaussians [8, 9], kernel density function [42] or temporal median filtering techniques [43]. In our case, we adopt the statistical Gaussian Mixture Model (GMM) for the segmentation, extraction of moving people and background area [8, 9]; this method is most common in the case of background dynamic and presents several advantages. Indeed, it can work without having to store

an important set of input data in the running process. The multimodality of the model allows tackling multimodal backgrounds and gradual illumination changes. Also, it returns a foreground with minimal errors in terms of misclassified pixels.

Tracking People: We developed our tracking system based on the basic tracking algorithm proposed by Collins *et al*. [3], which is as follows:

- Predict positions of known objects
- Associate predicted objects with current objects
- If tracks split, create new tracking hypothesis
- If tracks merge, merge tracking hypotheses
- Update object tracking models
- Reject false alarms

Most of the tracking system is built on the basis of this algorithm and therefore use prediction of features in the next frame. It reduces the search space, but predicting features requires use of a predictor like Kalman filter. It requires significant computation time to built and update the model [28]. In our system, we skipped the prediction of features to save computation time; rather we compared features obtained in the previous frame with features obtained in the current frame.

Our algorithm proposed in this paper is based on the following steps: moving people detection, skin color segmentation, location of color interest points detected on the moving skin regions, tracking people by matching the color interest points detected and calculates Euclidean distance between the best matches and other interest points in each consecutive frames of video sequence. We firstly use background subtraction to separate moving objects (foreground) and the static objects (background) based on Gaussian Mixture Model. After that, we detect all skin regions using the threshold technique, it helps to assume that the objects are all humans and not requires any type of classification, in addition to, it's to simplify the subsequent treatments by locating interest regions in the image to optimize the computation time. Then, the number of wrong detections is also reduced because the image regions are limited and the probability of finding a person is very low. Note that the HSV and YCbCr color spaces have been used only for skin region segmentation and shadow removal. For background modeling, interest points detection and correlation matching, we used RGB color space. Although HSV color space can separate color component from intensity, RGB color space can represent the color distribution of an interest point more accurately.

After detecting all skin regions in each frame of a sequence, smaller ones are discarded. In our case, a minimum size of regions is 400 pixels worked well. In particular, noise appearing in skin regions found should be removed before tracking phase. We have removed noise by two morphological operations, erosion and dilatation. After, we will limit the search areas of our moving people and we bounded by minimum bounding rectangle manually in order to simplify and reduce execution algorithm. Then a Harris detector is applied [10, 11] on these regions whose aim is to reduce the maximum information of an image in a few points and for positioning the interest points. The use of an interest points set allows us to track people frame by frame.

Tracking is performed by matching interest points features in the current frame with the ones in the previous frame. They are several kinds of matching techniques, namely normalized Euclidean Distance [2], correlation-based approach [3, 28, 29] and histogram-based matching [16]. The problem of Euclidean distance is that, the feature which has a higher value dominates others. To solve this problem, Xu *et al*. [2, 26] suggested using Mahalanobis distance, but it's computationally expensive. To avoid this problem and to give importance to significant features, we propose in this paper to use ZNCC correlation approach [12, 13] for feature-based tracking. The formula for ZNCC Correlation coefficient is showed in equation 7:

$$ZNCC(P,Q) = \frac{\sum_{i=1}^{n}\left(p_i - \overline{P}\right)\left(q_i - \overline{Q}\right)}{\sqrt{\sum_{i=1}^{n}\left(p_i - \overline{P}\right)^2 \sum_{i=1}^{n}\left(q_i - \overline{Q}\right)^2}} \quad (7)$$

With respectively $P=(p_i)_{i=1...n}$ and $Q=(q_i)_{i=1...n}$ the coordinates of points P and Q. And $\overline{P}$ and $\overline{Q}$ are the averages of P and Q respectively such as:

$$\overline{P} = \frac{\sum_{i=1}^{n} p_i}{n} \quad , \quad \overline{Q} = \frac{\sum_{i=1}^{n} q_i}{n}$$

It provides a measure of the interval [-1, 1], the similarity between two points P and Q is high when ZNCC (P, Q) approaches to 1. In practice, we set a threshold s (in general s=0.5) such as the pair (P, Q) is considered the true correspondent if the following constraint is satisfied:

$$ZNCC(P,Q) > s \quad (8)$$

During this step, several points are wrong pairings, this has for effect to alter completely the process of monitoring. Several robust regression techniques exist in the literature; in our experiments we have used RANSAC function (*Random Sample Consensus*) [14]. The main idea of this method is to perform draws of eight corresponding points among the set of two detected corresponding successive images. In our case, we have considered the step of interest points regularization as very important because the last step in our tracking system depends on it, so we have calculated Euclidean distance between the interest points detected in each frame in order to demonstrate the motion in these regions. Knowing that, Euclidean distance between two points *X* and *Y* is defined as length of line segment connecting them. In our digital image processing between two pixels M(x, y) and N (u, v), the Euclidean distance function [44] is:

$$d_e(M,N) = \sqrt{(x-u)^2 + (y-v)^2} \quad (9)$$

The main steps of our system for detecting and tracking moving people in video sequences proposed in this paper are described as follows:

Algorithm: Tracking Moving People Using Color Interest Points

Step1: Sequence of successive images.

Step 2: Detect moving people by separating background and foreground using Gaussian Mixture Model.

Step 3: Apply a skin detector in the combination of three color spaces RGB-YCbCr-HSV to segment skin regions.

Step 4: Apply morphological operations dilation and erosion on the skin regions founded in step 3 and surrounded them by minimum bounding rectangle.

Step 5: Apply Harris color detector on the skin regions founded in step 4.

Step 6: Pair color interest points detected using ZNCC correlation measure.

Step 7: Regularization of all interest points by RANSAC function to eliminate false matching detected in step 6.

Step 8: Calculate Euclidean distance between the best matches and other interest points in each consecutive frames of video sequence in order to motion detection regions.

Step 9: Detection and tracking results: the people tracked are numbered and surrounded by minimum bounding rectangle.

This algorithm has been 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 simulation results show that our method is capable to track multiple persons and therefore, to detect and manage the interactions between them with excellent precision, also it's robust in the case of noise existence.

Experiment Results: This section is composed of two distinct parts. In the first, we lay interest in some specific properties, such as the robustness of color interest points against Gaussian noise and contrast modification, after we compare our method to other existing tracking methods. In the second part, we present our experiment results on different video sequences used in this study.

Simulations: In order to test the performance of our application, two situations were discussed: contrast modifications and Gaussian noise addition. The Gaussian noise with varying deviation ó is added to the image coordinates point. We simulate an image with 320×240 resolution taken from a video sport. Color interest points are detected by the Harris algorithm [10. 11]. Table 1 shows the number of color interest points obtained for different contrast and noise conditions respectively:

The evaluation is performed by observing the variations of the number of interest points compared with the initial situation (no contrast modification and no

Table 1: Influence of contrast and noise.

a) Contrast influence

Contrast [LOW; HIGH]	Color interest points
[0.99 1]	1
[0.8 1]	7
[0.4 1]	37
[0 1] reference	52
[0 0.5]	55
[0 0.2]	64

b) Noise Influence

Gaussian noise	Color interest points
σ = 0 reference	52
σ = 0.001	62
σ = 0.01	954
σ = 0.1	1006
σ = 1	1071
σ = 2	1096

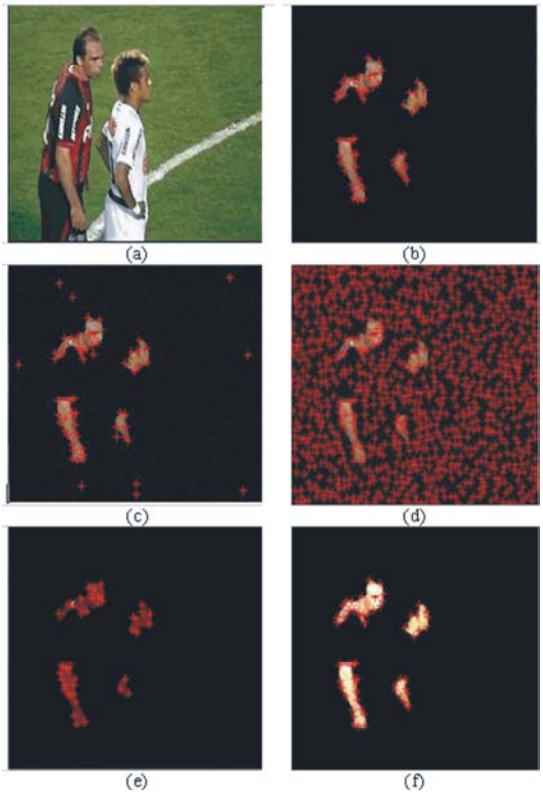

Fig. 2: Influence of contrast and Gaussian noise on the color interest points detection, (a) original image, (b) result of interest points detection with no contrast modification and no noise, (c) Gaussian noise with ó=0.001, (d) Gaussian noise with ó=0.01, (e) result detection with low contrast and (f) result detection with high contrast.

Table 2: Quality of tracking moving people in percentage for each method.

Quality of tracking moving people in percentage (%)		
Our proposed method	Method 1 [15]	Method 2 [12, 13]
98.86	97.17	93.27

noise: 52 color interest points by frame). It can be noticed that the number of interest points is relatively stable with respect to contrast modification as indicated in the images (e) and (f) in Figure 2. On the contrary, the color interest points detection is very sensitive to Gaussian noise as indicated in the images (c) and (d) in Figure 2. In general, the impact of noise in people tracking algorithms cannot be ignored and some of them can be completely lost and trace the wrong people.

In order to fully validate our method, we compare it to other existing tracking methods. The first method chosen presents object tracking system that uses a Variable Search Window (VSW) algorithm based on color and multiple points features [27], i.e. it's an effective improvement of meanshift with scale invariant feature transform (SIFT) algorithm. It's a method that extracts the feature points of the detected object and generates a variable search window (VSW) using the given information. This information is the positions of extracted feature points. An advantage of this method is that it's robust to specific color objects and can solve the problem of a similar color distribution; however, a disadvantage it's sensitive to non specific color objects due to illumination and noise. The second method used for comparison based on interest points and contours [24, 25] but with images in gray level and for a single object in the scene.

For estimating tracking accuracy of our proposed method we used the following formula:

$$\Pr ecesion(\%) = \frac{NT - NF}{NT} \times 100 \quad (10)$$

With NT is the total number of images in the sequence and NF is the number of images for false followed.

Table 2 shows the quality of tracking moving people for each method considered:

According to these results, we can point out that our proposed method has a higher precision than the other methods; the average accuracy is 98.86%. So, we have drawn various conclusions from this comparative study. First of all, it may be pointed out that the methods that are working on videos in gray levels are less efficient than those which work with color video as the case of the method [24, 25]. Then, only the proposed method allows to correctly tracking people with excellent precision.

Real Data: To characterize the performance of our method, we tested it on two different types of video data with different duration; these videos can be types of dynamic background (changing illumination) from static cameras:

Class Video: A sequence of 245 images with resolution 450×360 pixels taken from a class video that lasts 25 minutes and 40 seconds (a person who moves front of a static camera).

Sport Video: A sequence of 207 images with resolution 420×356 pixels taken from a football video that lasts one minute and five seconds (two players football who meet and walk side by side).

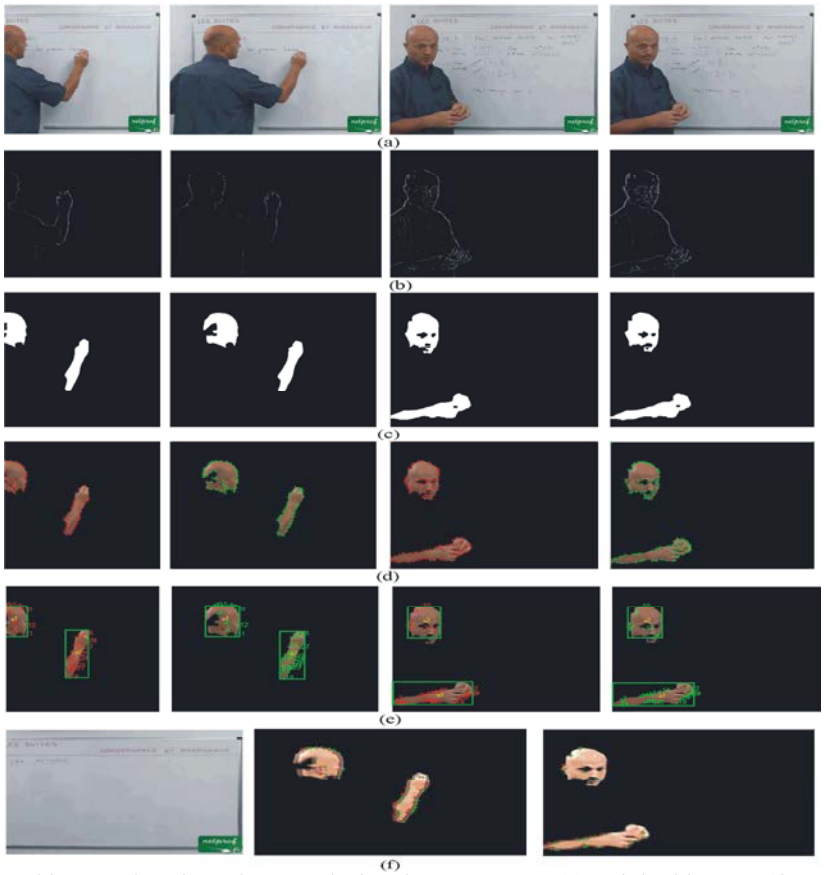

Fig. 3: Detection and tracking results of moving people for class sequence. (a) Original images (frames 31, 35, 196 and 199), (b) Moving pixels detected, (c) Results of skin regions segmentation, (d) Color Interest Points detected are shown in red and green for each pair of images, (e) Results of detection and tracking, (f) The first frame in the left represents the background identified for this sequence, the second frame represents fusion results with the matching of Harris's points by RANSAC method between the frames 31 and 35, the last frame shows the same results between the frames 196 and 199.

For all these types of videos we firstly use background modeling to separate moving objects (Foreground) and static objects (background) using the statistical Gaussian Mixture Model. After that, we selected all skin regions in RGB-YCbCr-HSV color spaces for its advantages of speed and simplicity. Once the classification of the skin pixels is done, we apply erosion followed by dilation to remove scattered pixels representing misclassifications and we strengthen the detected regions. After the detection of color interest points by the Harris algorithm [10, 11], the matches between each pair of images are determined by ZNCC correlation function [12, 13] and are regularized by RANSAC algorithm [14].

We present now several experiments for our method, the scenarios are varied from an only simple walking person to two people moving who meet and walk side by side. The first analyzed sequence is about one person who moves independently in a classroom under the assumption that the camera is fixed. The background in this sequence is relatively stable, snuff images from different views: front, profile and behind. The results are shown in Figure 3, where the first line (a) represents the original images considered in this video; the second line (b) shows moving pixels detected (foreground); and the third line (c) shows the results of skin regions segmentation. The line (d) shows color interest points detected only on skin regions, the interest points are shown in red and green in each pair of images; the line (e) shows the results of detection and tracking people for this sequence; the matches are shown and numbered in red and green in each pair of images; all moving regions are correctly detected, followed and automatically surrounded by bounding box rectangle. The last line (f) shows the

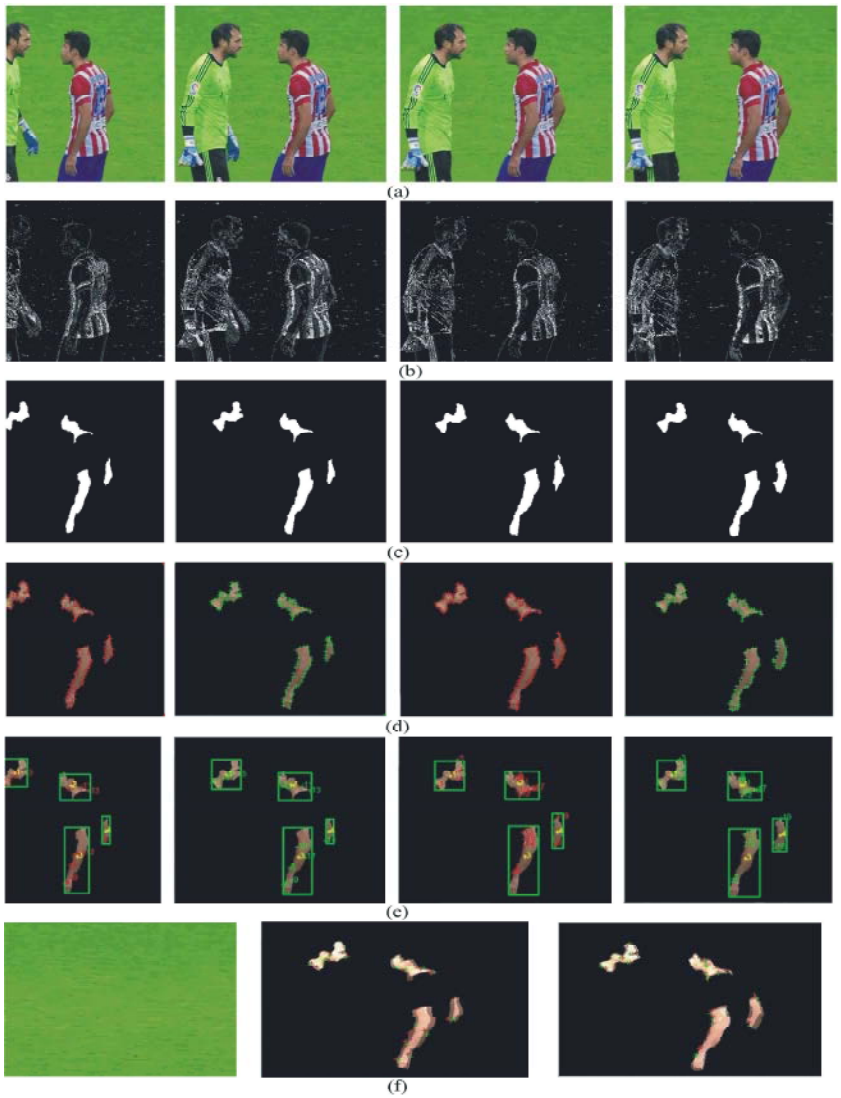

Fig. 4: Detection and tracking results of moving people for sport sequence. (a) Original images (frames 99, 102, 105 and 107), (b) Moving pixels detected, (c) Results of skin regions segmentation of, (d) Color Interest points detected shown in red and green for each pair of images, (e) Results of detection and tracking, (f) The first frame in the left represents the background identified for this sequence, the second frame represents fusion results with the matching of Harris's points by RANSAC method between the frames 99 and 102, the last frame shows the same results between the frames 105 and 107.

background identified from this sequence and the results of fusion for each pair of consecutive images. The corresponding displacement vector for this sequence is given by $V_1=(v_1,\ldots v_n)$ where n is the number of matches detected in each pair of images. The different traversed Euclidean distance for frames 31, 35, 196 and 199 respectively are d_1=3.2463e+003, d_2=2.7077e+003, d_3=2.9711e+003, d_4=4.8297e+003. We can notice that the face and hands are correctly detected and tracked in all images considered for this sequence; also we can show the difficulties of this video are that the background is changed in frames 196 and 199 and even in the presence of illumination changes but the results of detection moving objects is excellent.

The second selected sequence is a sport video of two players football moving toward each other. The results of detection and tracking are presented in Figure 4, the first line (a) shows the original images filmed mostly in profile. The second line (b) shows moving pixels detected (foreground). The third line (c) presents the results of skin

color segmentation. The fourth line (d) shows color interest points detected on discovered skin areas, these later are shown in red and green in each pair of images; the line (e) shows detection and tracking results of these two players football; the matches are shown and numbered in red and green in each pair of images; all moving regions detected are correctly followed and automatically surrounded by bounding rectangle. The last line (f) shows the background identified for this sequence and fusion results for each pair of consecutive images. The corresponding displacement vector for this sequence is given by $V_2=(v_1,\ldots v_m)$ where m is the number of matches detected in each pair of images. The different traversed Euclidean distance for frames 99, 102, 105 and 107 respectively are: d_1=2.1019e+003, d_2=1.9172e+003, d_3=2.5233e+003, d_4=2.5925e+003. Hence, people are correctly detected and tracked in all frames considered; all skin regions are correctly detected and surrounded by rectangles; the matching results are better after the removal of wrong matches; all this is thanks to RANSAC function. Hands and faces are detected, segmented and followed in the majority of images considered even in the presence of illumination changes. The results obtained by our method if they are not perfect, are very encouraging given complexity of sequence. As well as, our method fits well with noise conditions and contrast changes. In summary, the combination of color interest points and skin color will develop a tracking method that implements the time parameter and quality of tracking.

CONCLUSION

Intelligent Video Surveillance (IVS) has become a major research area in computer vision. In particular, tracking and identifying humans in the scene is a central problem in IVS. We proposed solutions to several problems related to tracking people in a dynamic scene. We have proposed an algorithm based on skin segmentation technique and color interest points for feature-based tracking of multiple people in indoor and outdoor environment. We used feature-based tracking, as it's faster than other methods. In particular, we have used ZNCC correlation approach for the matching of feature points detected by Harris algorithm, as it gives better results than histogram-based approach and Euclidean Distance-based approach. Finally, by calculating the Euclidean distance between the best matches and other interest points in each consecutive frames of video sequence we can observe the motion of people tracked in the scene. Simulations and experimental results show the robustness of our method. The results are very encouraging and our method fits well with noise conditions and contrast changes.

REFERENCES

1. Hu, W., T.N. Tan, Fellow, L. Wang and S. Maybank, 2004. A survey on visual surveillance of object motion and behaviors. IEEE Trans on systems, Man and Cybernetics-Part C: Applications And Reviews, 34(3): 334-352.
2. Xu, L., J.L. Landabaso and B. Lei, 2004. Segmentation and tracking of multiple moving objects for intelligent video analysis. BT Technology Journal, 22(3): 140-150.
3. Collins, R.T., A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver N. Enomoto and O. Hasegawa, 2000. A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, CMU.
4. Foresti, G.L., P. Mahonen and C.S. Regazzoni, 2000. Multimedia Video-Based Surveillance Systems: from User Requirements to Research Solutions, Kluwer Academic Publishers.
5. McKenna, S.J., S. Jabri, Z. Duric, A. Rosenfeld and H. Wechsler, 2000. Tracking groups of people. Computer Vision and Image Understanding, 80(1): 42-56.
6. Wren, C.R., A. Azarbayejani, T. Darrell and A.P. Pentland, 1997. Pfinder: Real-Time Tracking of the Human Body. IEEE Trans on Pattern Analysis and Machine Intelligence, 19(7): 780-785.
7. Yilmaz, A., O. Javed and M. Shah, 2006. Object Tracking: A Survey. ACM Computing Surveys, 38(4).
8. Stauffer, C. and W.E.L. Grimson, 2000. Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intell., 22(8): 747-757.
9. Bouwmans, T., FE. Baf and B. Vachon, 2008. Background modeling using mixture of gaussians for foreground detection–a survey. Recent Patents Comput Sci., 1(3): 219-237.
10. Montesinos, P., V. Gouet and R. Deriche, 1998. Differential invariants for color images. International Conference on Pattern Recognition, pp: 838-840.
11. Harris, C. and M. Stephens, 1988. A combined corner and edge detector. In: Proc of the Fourth Alvey Vision Conference, pp: 147-151.
12. Chambon, S. and A.Crouzil, 2011. Similarity measures for image matching despite occlusions in stereo vision. Pattern Recognit., 44(9): 2063-2075.

13. Di Stefano, L., S. Mattoccia and F. Tombari, 2005. ZNCC-based template matching using bounded partial correlation. Pattern Recognit. Letters, 26: 2129-2134.
14. Fischler, M.A. and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing.
15. McKenna, S.J., S. Jabri, Z. Duric and H. Wechsler, 2000. Tracking Interacting People. Proc of International Conference on Automatic Face and Gesture Recognition, pp: 348-353.
16. Comaniciu, D., V. Ramesh and P. Meer, 2000. Real-time tracking of non-rigid objects using mean shift. Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2: 142-149.
17. Nummiaro, K., E. Koller-Meier and L.V. Gool, 2003. Color Features for Tracking Non-Rigid Objects. Special Issue on Visual Surveillance, Chinese Journal of Automation, 29: 345-355.
18. Mohan, A., C. Papageorgiou and T. Poggio, 2001. Example-based object detection in images by components. IEEE Trans. Pattern Recognit. Machine Intell., 23(4): 349-361.
19. Nguyen, H.T., M. Worring, R. Van den Boomgaard and A.W.M. Smeulders, 2002. Tracking Nonparameterized Object Contours in Video. IEEE Trans on Image Processing, 11(9): 1081-1091.
20. Paragios, N. and R. Deriche, 2000. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Pattern Anal. Machine Intell., 22(3) 266-280.
21. Peterfreund. N., 1999. Robust Tracking of Position and Velocity With Kalman Snakes. IEEE Trans On Pattern Analysis and Machine Intell., 21(6).
22. Isard, M. and A. Blake, 1998. Condensation-Conditional Density Propagation for Visual Tracking. International Journal of Computer Vision, 29(1): 5-28.
23. Rasmussen, C. and G.D. Hager, 2001. Probabilistic Data Association Methods for Tracking Complex Visual Objects. IEEE Trans On Pattern Analysis and Machine Intelligence, 23(6).
24. Gouet, V. and B. Lameyre, 2004. SAP: A robust approach to track objects in video streams with snakes and points. British Machine Vision Conference, pp: 737-746.
25. Lameyre, B. and V. Gouet, 2005. Object Tracking and Identification in Video Streams with Snakes and Points. Proc of the 5th Pacific Rim conference on Advances in Multimedia Information Processing, 3333: 61-68.
26. Gabriel, P., J.B. Hayet, J. Piater and J. Verly, 2005. Object tracking using color interest points. Advanced Video and Signal Based Surveillance, pp: 159-164.
27. Lim, H. and D. Kang, 2011. Object tracking system using a VSW algorithm based on color and point features. EURASIP Journal on Advances in Signal Processing.
28. Boufama, B. and M.A. Ali, 2007. Tracking Multiple People in the Context of Video Surveillance. Proc of the 4th international conference on Image Analysis and Recognition, pp: 581-592, Springer.
29. Ali, M.A., S. Indupalli and B. Boufama, 2006. Tracking multiple people for video surveillance. International Journal of Computer Vision-IJCV.
30. Decarlo, D. and D. Metaxas, 2000. Optical Flow Constraints on Deformable Models with Applications to Face Tracking. International Journal on Computer vision, 8(2): 99-127.
31. Zhang, Y. and C. Kambhamettu, 2002. 3D head tracking under partial occlusion. Pattern Recognition, 35(7): 176-182.
32. Paterson, J. and A. Fitzgibbon, 2003. 3D Head Tracking using Non-Linear Optimization. Proc Of the British Machine Vision Conference, pp: 609-618.
33. Stenger, B., A. Thayananthan, P.H.S. Torr and R. Cipolla, 2006. Model-Based Hand Tracking Using a Hierarchical Bayesian Filter. IEEE Trans on Pattern Analysis and Machine Intelligence, 28(9): 1372-1384.
34. Bouirouga, H., S. Elfkih, A. Jilbab and D. Aboutajdine, 2012. Skin detection in pornographic videos using threshold technique. Journal of Theoretical and Applied Information Technology, 35(1).
35. Singh, S.K., D.S. Chauhan, M. Vatsa and R. Singh, 2003. A Robust Skin Color Based Face Detection Algorithm. Tamkang Journal of Science and Engineering, 6(4): 227-234.
36. Schmid, C., R. Mohr and C. Bauckhage, 2000. Evaluation of interest point detectors. International Journal of Computer Vision, pp: 151-172.
37. Schmid, C. and R. Mohr, 1997. Local grayvalue invariants for image retrieval. IEEE Trans on Pattern Analysis and Machine Intelligence, pp: 530-534.
38. Mikolajczyk, K. and C. Schmid, 2001. Indexing based on scale invariant interest points. In International Conference on Computer Vision, Vancouver, Canada,.
39. David G. Lowe, 2004. Distinctive image features from scale-invariant keypoints. Accepted for publication in the International Journal of Computer Vision.

40. Gouet, V. and N. Boujemaa, 2002. About optimal use of color points of interest for content-based image retrieval. Internal Report, INRIA Rocquencourt.
41. Dempster, A., N. Laird and D. Rubin, 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. Royal Statistical Soc., 39(Series B): 1-38.
42. Elgamal, A., R. Duraiswami, D. Harwood and L. Davis, 2002. Background and foreground modeling using nonparametric kernel density estimation for visual Surveillance. Proc of the IEEE, 90(7).
43. Zhou, Q. and J.K. Aggarwal, 2001. Tracking and classifying moving objects from video. Proc of 2nd IEEE Intl Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2001), Kauai, Hawaii, USA.