International Journal of Innovative Computing, Information and Control
Volume **8**, Number **4**, April **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 2761–2776

VIDEO SURVEILLANCE USING FACIAL FEATURES-BASED TRACKING

Ibrahim Hemdan, Stephen Karungaru and Kenji Terada

Department of Information Science and Intelligent Systems
University of Tokushima
2-1, Minami-Josanjima, Tokushima 770-8506, Japan
ibrahim@is.tokushima-u.ac.jp

Received October 2010; revised March 2011

Abstract. *In this work, we present a real time system for facial features detection and tracking in image sequences. We are interested in developing a human tracking system that can improve human-computer interaction and benefit video surveillance problems by making it invariant to rotation, illumination and subject's movement. However, since human body and face movements are very complicated to detect and track, all available cues that can narrow the search space should be considered. This paper describes a novel strategy for both face tracking and facial feature detection. Face detection is important because it reduces the search space and consequently saves time for further face processing, e.g., recognition or transmission. Moreover, facial feature detection enables face normalization which leads to size invariant face recognition. The proposed face tracker resembles human perception in that, initially, it utilizes motion as the major cue and thereafter, searches for the eyes in the areas likely to contain human faces. The presence of a face is determined using an eye tracker. Eyes are important facial features due to their relatively constant interocular distance. In this work, efficiency improvement focuses on two points: reducing template matching area and speeding up the matching process. Our method initially detects two rough eye candidate regions using a feature based method. All other processes are thereafter performed inside the candidate regions. In addition, we can evaluate the size of eye template according to the size of the regions. Altogether, the proposed method combines the accuracy of template based methods and the efficiency of feature based methods in the visual spectrum. To prove the effectiveness of this approach, we performed comparative experiments using real video images. We achieved a real time detection accuracy of about 96%.*

Keywords: Skin detection, Face tracking, Template matching, Eye detection

1. Introduction.

Machine detection and tracking of human faces from video frames is emerging as an active research area spanning several disciplines such as image processing, pattern recognition and computer vision. Numerous attempts have been conducted in face localization and identification for a variety of applications including intelligent surveillance, law enforcement systems, virtual reality interfaces, etc. The main focus is on the extraction of primary facial features such as eyes, eyebrows and nose and mouth. Facial features are very important for many applications like face expression recognition, rotation invariant face recognition and person's state monitoring.

However, face detection provides interesting challenges due to the stringent requirements for pattern classification and learning techniques. The dimension of the feature space is extremely large. Therefore, training and testing face classifiers is computationally expensive. Moreover, the other factors that affect the system's performance include scale, pose, illumination, facial expression, occlusion and age. The human body, especially the face or hand movements are very complicated to track compared with artificial articulated objects. Therefore, all available cues which can narrow the search space in tracking or detection systems should be considered. Among these features, spatial, temporal and textural features are the most important. Moreover, the simplicity and obviousness of skin color as an effective cue for hand and/or face segmentation and tracking has caused many researchers to develop methods based skin color [1]. Face detection can be viewed as a two-class recognition problem classifying an image region as either a "face" or a "non-face".

D. Lin [2] proposed an algorithm combining the Principal Component Analysis (PCA) with Scale Invariant Feature Transform (SIFT) for face recognition. Initially, feature vectors invariant to image scaling and rotation are extracted by SIFT with a different local descriptor. Then, PCA is applied to projecting the feature vectors to the new feature space as PCA-SIFT local descriptors. Finally, the local descriptors are clustered by K-means algorithm. This local and global information of images are combined to classify face images. In [3], a novel approach to multi-view face tracking by taking advantage of a learning based classifier and template matching is proposed. The classifier trained with a horizontal projection histogram of a rectangular face region robustly locates the upper and lower boundaries of the face in spite of the face viewpoint change. Due to the translation invariant property of the horizontal projection histogram, considerable computational efficiency is achieved by increasing the moving steps of the scanning window. Then, template matching precisely determines the left and right boundaries of the face. To enhance the matching accuracy, a new distance measure incorporating the classifier outcome into the sum of the absolute distances (SAD) is used. In [4] an Adaboost machine is applied to improving face recognition accuracy. The machine consists of several classifiers of radial basis function neural networks (RBFNN). To speed up the training process and increase the generalization of RBFNNs, the principal component analysis (PCA) technique is applied to selecting adequate centers for the RBFNN classifiers and a novel weight updating mechanism is adopted to update the connection weights between the hidden neurons and output neurons. A novel Non-parametric Kernel Discriminant Analysis (NKDA) based facial feature extraction method insensitive to varying illumination is proposed for face recognition. Theoretical analysis on the recently proposed Non-parametric Discriminant Analysis (NDA) shows that NDA has its limitations on extracting the non-linear features owing to the high nonlinear and complex distribution of facial images under varying lighting conditions. In order to improve NDA for face recognition NKDA was proposed by improving the NDA using a kernel [5]. With over 170 reported approaches in face detection [6], the impact of the research has broad implications on face recognition. However, in spite of all the work, challenges still remain especially in real time invariant facial features tracking for special applications, for example, the automatic monitoring of a desktop worker's state using inexpensive USB cameras.

Therefore, this paper describes a novel strategy for both face tracking and facial feature detection. Face detection is important because it restricts the field of view and thus reduces the amount of computation for further face recognition or transmission, while facial feature detection is important because it enables face normalization and leads to size invariant face recognition. Resembling the way of human perceiving face and facial features, the proposed face tracker initially utilizes motion as the major cue and then searches for the human eyes in the areas that contain a human face. Eyes are important facial features due to their relatively constant interocular distance. Efficiency improvement in this work focuses on two points: reducing template matching area and speeding up the matching process. Our method initially detects two rough eye candidate regions using a skin color based method. A good skin classifier must be able to discriminate between skin and non-skin pixels for a wide range of people with different skin types and perform well under different illumination condition indoors or outdoors. Thereafter, all other processes are performed inside the candidate regions. Moreover, we can evaluate the size of eye template based on the size of the blobs. In other words, profiting from the possibility of evaluating the size of eyes, our algorithm performs the template matching just once thereby improving computation cost. Altogether, the proposed method combines the accuracy of template based methods and the efficiency of feature based methods in the visual spectrum for 2D images. In addition, using different skin color models, we achieve lighting invariance for the database used.

The contribution of this paper is in two areas. The use of an eye tracker to verify the location of the face from the results of skin color detection and computation time reduction using single template resizing based on the size of the skin color blobs. This results in a robust real time location and tracking of face features (eyes, mouth, nose and lips). The Facial features detected can be used in other applications like face and gesture recognition.

The rest of this paper is organized as follows. Section 1.1 presents the traditional methods. Section 2 describes the proposed method system outlines. Section 3 investigates the performance of the approach using experiments while Section 4 concludes the paper and suggests possible future works.

1.1. Face detection techniques – traditional approaches.

Face detection can be view- ed as a two-class recognition problem classifying an image region as either a "face" or a "non-face". With over 170 reported approaches in face detection [6], the impact of the research has broad implications on face recognition. The various approaches for face detection can be classified into four categories:

- Knowledge-based methods that use rules encoding the human knowledge of a face.
- Feature invariant approaches find robust structural features, invariant to pose, lighting, etc.
- Template matching methods compute the correlation between a face and an input image for detection.
- Appearance-based methods, in contrast to template matching, use models learned from training sets to represent the variability of facial appearance.

Table 1 summarizes the four categories by indicating representative approaches for each category.

1.2. Automatic eye detection.

There are two purposes of eye detection. One is to detect the existence of eyes, and the other is to accurately locate eye positions. Under most situations, the eye position is measured using the pupil center.

Current eye detection methods can be divided into two categories: active and passive [7]. The active detection methods use special types of illumination. Under IR illumination, pupils show physical properties which can be utilized to localize the eyes [8,9]. The active eye detection methods are very accurate and robust but because they need special lighting sources, they have more false detections outdoors, where the outdoor light impacts the IR illumination.

Passive methods directly detect eyes from images using distinct features including image gradients [22], projection [23] and templates [24,25]. However, in these methods, heuristics and post-processing are usually necessary to remove false detections, and these features are sensitive to image noise. Huang and Wechsler selected optimal Wavelet packets and classified the eye and non-eye regions using Radial Basis Functions (RBFs) [27]. Gabor wavelets are robust to moderate illumination change, and the similarity measurement

Table 1. Categorization of face detection methods

Approach	Representative Approaches
Knowledge-based	Multi-resolution rule-based method [10]
Feature invariant	
Facial features	Grouping of edges [11]
Skin color	Gaussian mixtures [12]
Multiple features	Integration of skin color, size and shape[13]
Template matching	
Predefined Face templates	Shape template [14]
Deformable templates	Active Shape Model (ASM) [15]
Appearance-based	
Eigen face	Eigenvector decomposition and clustering [16]
Distribution-based	Gaussian distribution and multilayer perceptron [17]
Neural Network	Ensemble of neural networks and arbitration schemes [18]
Support Vector Machines	SVM with polynomial kernel [19]
Hidden Markov Model	Higher order statistics with HMM [20]
Adaboost	Attentional cascade with boosted features [21]

based on Gabor wavelets is sensitive to localization change so that they can be used to detect fiducial points [26]. In [28], a two-layer Gabor wavelet network (GWN) is proposed to localize facial points from coarse to fine. The first layer localizes face region while the second layer further refines the facial points. The experiments on FERET show that about 95% eyes are located with a distance error smaller than 3 pixels. Some passive methods consider eye detection as a typical two-class pattern recognition problem. In [29], eye detectors are trained with the rectangle Haar features and AdaBoost algorithm to detect eyes in images. In [30], the critical features are selected from both rectangle and center surrounded Haar feature sets. GentleBoost is applied to constructing a final eye detector. The same algorithm is used to train a frontal face detector. After a frontal face is detected, eyes are located inside the face region.

2. Proposed System.

2.1. **Overview.** There are a lot of techniques in objects detection: Motion, Color, Texture, Shape, etc., each with its pros and cons. Color will be the main focus in this project because of its advantages especially, the ability to differentiate objects. Color provides useful information especially during occlusions and can be robust to non-stationary background. In human tracking applications, skin color is often used to signify the presence of human targets. However, tracking skin color is unreliable due to the small skin area available. In addition, the skin color can easily be blocked by objects or lost when the subject is not facing the camera. In this paper, a method that uses human clothings color is discussed. The system overview is shown in Figure 1. After the video is captured, skin color processing is performed followed by face detection. Facial features are then detected and tracked.

2.2. **Color system.** Color information is commonly represented in the widely used RGB coordinate system. This representation is hardware oriented optimized for image acquisition or display devices but not particularly suitable for describing the perception of

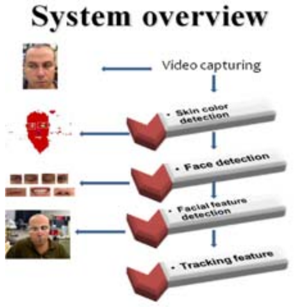


Figure 1. System overview

colors. On the other hand, the HSV (hue, saturation, value) color model corresponds more closely to the human perception of color. HSV domain allows color and brightness isolation. This is useful as brightness information is not useful in color object detection. The extra brightness information could be due to external factors (external light source), and therefore, holds no meaning.

2.2.1. *Advantages of using color.*

- Unique to an object so they can easily be identified even after occlusion.
- Robust representation that is invariant to complex, deformed and changeable shape (human), orientation, size, partial occlusion.
- Resolution independent to low resolution, fast to process, low cost.

2.2.2. *Disadvantages of using color.*

- Object not unique to color: Affected by objects with the same color as targeted object.
- Inaccurate measurement: affected by illumination/brightness and low saturation (Too bright looks white, too dark looks black). Sources of poor lightings can be highlights, shadows, light spectrum emitted by different/numerous lightings, daily variation due to sunlight and temperature. Also, color white balance in every camera is different.
- Not distinctive: Wide range of skin color (reddish, yellow, brown, white, black).

Depending on the application of color detection, some parameters like camera calibration and scene lighting, can be controlled.

2.3. **Skin color.** Human skin is composed of several layers of tissue which consist essentially of blood cells, and a yellow pigment called melanin. The appearance of the skin is affected by a number of factors including the degree of pigmentation (varies amongst individuals and different races), the concentration of blood, and the incident light source. The combination of these entire factors give rise to a variation in skin color which spans over the range of red, yellow and brownish-black. Nevertheless, these variations correspond to a restricted range of hue values.

2.4. **Skin color modelling.** Pixel-based skin color detection methods introduce a tool for measuring the distance of pixel to skin color tones. The color itself can be represented in many different ways. In this work, the skin color is modelled using the following color spaces.

2.4.1. *RGB and normalized RGB.* The RGB color space has been widely used for processing and storing digital image data. This model describes each color as a weighted combination of three base components Red, Green and Blue. However, high correlation between components and mixing luminance with chromaticity makes it very sensitive to changes in imaging conditions such as lighting. Normalized RGB tries to reduce the dependence of each component to the brightness of the pixel by normalizing each component.

The simplicity of this color spaces has been the main reason for its popularity in skin color detection [31].

2.4.2. *Hue saturation value model.* Hue Saturation Value (HSV) model describes color with dominant color (Hue), colorfulness in proportion to the brightness (Saturation), and the amount of luminance (Value). The most important characteristic of the model is its explicit discrimination of luminance from chrominance. This makes the model insensitive to brightness at white color and ambient light; the properties that have attracted many researchers to this model [32]. H, S and V components are computed using Equations (1), (2) and (3). The model, however, has the disadvantage of being discontinuous at points where the brightness is very low.

$$H = \text{arcos}\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-B)+(G-B)}} \tag{1}$$

$$S = 1 - 3\frac{\min(R,G,B)}{R+B+G} \tag{2}$$

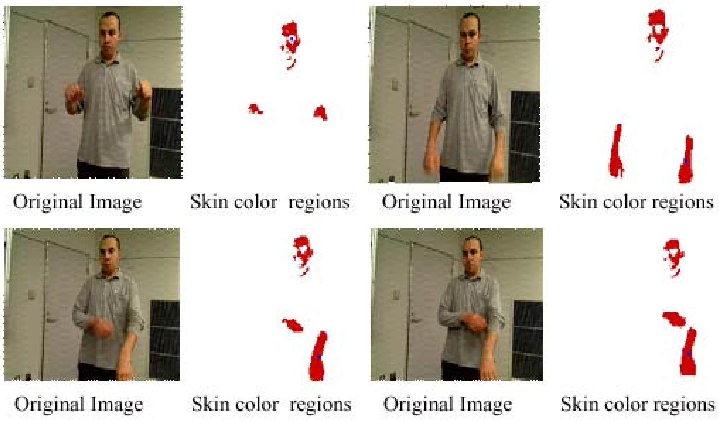
$$V = \frac{1}{3}(R+B+G) \tag{3}$$

2.4.3. *YCbCr.* The YCbCr color space was developed as part of ITR-R BT.601 during the development of a worldwide digital component video standard which is commonly used by European television studios. YCbCr separates luminance from chrominance in RGB values using a linear transform consisting of a weighted some of the three components. The simplicity of the transform and the explicit separation of luminance have made the model very attractive for skin color detection [33].

2.5. **The segmentation process.** The segmentation process consists of identifying skin-color regions in the input image based on the generated model. Initially, all pixels in input image are converted to the chromatic space. Using the skin color model, a gray-level image is generated and the intensity of each pixel in the new image represents its probability to be skin-color. Since skin-color pixels present a stronger response in the gray-level image, we can identify such regions by a thresholding process. The threshold value was chosen empirically.

In general, due to noise and distortions in input image, the result of thresholding may generate non-contiguous skin-color regions. Furthermore, after this operation, spurious pixels may appear in the binary image. In order to solve these problems, first we have applied a morphological closing operator to obtaining skin-color blobs. Subsequently, a median filter was used to eliminate spurious pixels. Regions with size less than 1% of image size are eliminated. Furthermore, structural aspects are considered so that regions that do not present the face structure are removed. An example for skin color detection is shown in Figure 2.

2.6. **Template-based object tracking.** A template describes a specific object to be tracked. Object tracking is done using the highest correlation between template and image sequence under analysis. Two possible matching criteria are:



Original Image Skin color regions Original Image Skin color regions

Original Image Skin color regions Original Image Skin color regions

Figure 2. An example of skin color detection using skin color model: (left) the original image, (right) skin detection

2.6.1. *Image subtractions.* In this technique, the template position is determined from minimizing the distance function between the template and various positions in the image. Image subtraction techniques require less computation time than the correlation techniques, and perform well in restricted environments where imaging conditions, such as image intensity and viewing angles between the template and images containing the template are the same.

2.6.2. *Correlation.* Matching by correlation utilizes the position of the normalized cross correlation peak between a template and an image to locate the best match. This technique is generally immune to noise and illumination effects in the images, but suffers from high computational complexity caused by summations over the entire template. Point correlation can reduce the computational complexity to a small set of carefully chosen points for the summations.

There are two types of object template matching, fixed and updated matching.

2.6.3. *Fixed template matching.* Simple template-based tracking algorithms use a fixed template over the entire image sequence. Fixed templates are useful when object shapes do not change with respect to the viewing angle of the camera [34] and no occlusions. Illumination change is also a problem as the initial template may differ from image sequences in the video. Fixed template matching algorithm implementation time can be improved using a technique called difference decomposition.

Difference decomposition was introduced by [35] to solve registration problem in tracking. The tracking problem is posed as the problem of finding the best set of registration parameters describing the motion and deformation of the target frame through the image sequence. The difference images are generated by adding small shift to the initial template. In this method, parameter variations are written as a linear function of a difference image (the difference between the reference target image and the current image). This approach is very efficient as motion can be easily deduced from the difference images and the target frame.

2.6.4. *Template update matching.* An underlying assumption with fixed template matching is that the object remains unchanged. This assumption is reasonable for a certain period of time, but eventually the template is no longer an accurate model of the appearance of the object [36]. In order to account for appearance changes, illumination changes, occlusion in the video, the template is updated throughout the video sequence [35]. The question is when to update the template. One approach is to update the template every frame (or every n frames) with a new template extracted from the current image at the current location of the template. The problem with this algorithm is that the template drifts. Each time the template is updated, small errors are introduced in the location of the template.

A better approach would be updating the template only when it satisfies certain predetermined conditions. Usually the template update is based on a threshold value [37], when the matching correlation value is greater than a threshold value, the template is updated. Choosing an appropriate value for threshold is very important. The template update algorithm is more robust to appearance and illumination changes compared to fixed template matching.

An alternative to the template update algorithm is to model the appearance of the object in a linear subspace. The subspace is obtained by applying Principal Component Analysis (PCA) to a set of training images. When the object (itself or its class) is available before tracking, appearances under different viewpoints [38] or illumination conditions could be used to create a linear subspace off-line. When there is no prior information about the object to be tracked, a linear subspace could be updated on-line [39]. Incremental singular value decomposition, application of the Gramm-Schmidt orthogonalization to a subset of images in the sequence [40], and active appearance models [34] are recently developed methods of on-line template updating.

2.6.5. *Challenges in template matching.* Appearance Changes: As an object moves through the field of view of a camera, its appearance may change dramatically. In a typical application, appearance may change due to changes in the viewpoint, illumination, shape (such as deformations), and partial occlusion of the object. So it is important to keep track of appearance changes [40,41].

Scale and Rotation: Robustness to scale and rotation variations is another important challenge. If the object of interest is moving towards the camera or away from it, the size of the object changes tremendously in the video. If the camera rotates or the object is rotated then the orientation of the object of interest also changes. The best matching region could be searched for at different scale and rotation parameters for each frame, either exhaustively or using a gradient-descent technique. Due to computational limitations, typically only a small portion of the scale-rotation space is searched. As we check for different scale and rotation parameters for each frame since we do not know beforehand which frame has changed, huge computational overhead for each frame is required.

Occlusion: Occlusion is defined as the covering of the object partly or fully by its surroundings [38]. Since we have a template defined, if the object is occluded, matching the occluded object and template produces a mismatch. If the occlusion persists for a while, there is a problem of permanently loosing track of the object. So occlusion is a serious problem which needs to be solved for efficient object tracking.

Illumination Changes: As an object moves through the light source, the brightness of the images and objects changes. In a typical application, illumination changes due to the movement of object towards or away from light source.

2.6.6. *Template matching method.* Let I be an image of dimension $m \times n$ and T be another image of dimension $p \times q$ such that $p < m$ and $q < n$. Template matching is defined as a search method which finds the portion in I of size $p \times q$ where T has the maximum cross correlation coefficient. The normalized cross correlation coefficient [42] is defined as:

$$\gamma(x, y) = \frac{\sum_s \sum_t \delta_{I(x+s,y+t)} \delta_{T(s,)}}{\sum_s \sum_t \delta^2_{I(x+s,y+t)} \sum_s \sum_t \delta^2_{T(s,)}} \tag{4}$$

where

$$\begin{aligned}
&\delta_{I(x+s,y+t)} = I(x+s, y+t) - \bar{I}(x,y), \quad \delta_{T(s,t)} = T(s,t) - \bar{T},\\
&s \in \{1, 2, \cdots, p\}, \ t \in \{1, 2, \ldots, q\},\\
&x \in \{1, 2, \cdots, m-p+1\}, \ y \in \{1, 2, \ldots, n-q+1\},\\
&\bar{I}(x,y) = \frac{1}{pq} \sum_s \sum_t I(x+s, y+t)\\
&\bar{T} = \frac{1}{pq} \sum_s \sum_t T(s,t)
\end{aligned}$$

The value of cross-correlation coefficient γ ranges between $[-1, +1]$. A value of $+1$ indicates that T is completely matched with $I(x, y)$ and -1 indicates a complete mismatch. For template matching the template, T slides over I and γ is calculated for each coordinate (x, y). After completing this calculation, the point which exhibits maximum γ is referred to as the match point.

Basically, correlation coefficients ρ are computed in different positions within the skin region as follows:

$$\rho = \frac{cov(x, y)}{\sigma(x)\sigma(y)} \tag{5}$$

where x is the eye template and y is the image patch in a specific position within the skin region. The function *cov* in the Equation (5) returns the covariance between x and y.

2.7. **Face detection: verification process.** The main goal in identifying skin-color regions is to reduce the search space for faces. However, it is important to note that not all detected regions contain faces. Some regions correspond to parts of human body, while other correspond to objects with colors similar to that of skin. Thus, it is necessary to verify the presence or absence of a face in each detected region.

To accomplish this task, we have used an eye detector based on an efficient template matching scheme. It is worth saying that there is biological evidence that eyes play the most important role in human face detection [43], because of their relatively constant interocular distance [44]. The eye template used is illustrated in Figure 3 (top half).

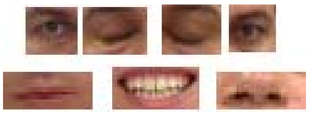


Figure 3. Eye, mouth and nose templates

In general, existing methods use several templates with different scales to accomplish the detection. For instance, the work of Brunnelli and Poggio [45] uses five templates with scales 0.7, 0.85, 1, 1.5 and 1.35 to detect eyes in an image. Although these methods have obtained good results, they are computational expensive and not suitable to real time processing.

So the solution presented in this work is to improve the efficiency of face detection focusing on two points: reducing the area in the face image for template matching and cutting down the times of this type of matching. In fact, our method firstly detects the two rough regions of eyes in the face using a feature based method. Template matching is then performed only in the two regions which are much smaller than the face.

Our method estimates the person eye size and resizes the eye template. Thus, the template matching is performed only once, using the upper portion of the detected skin-color region.

2.8. **Detection and tracking of facial features.** Once a face is detected, the system proceeds with the search for pupils, nostrils and lip corners. After, these facial features are tracked in the video sequence. The approach used is similar to the work of Stiefelhagen and Yang [46].

2.8.1. *Searching the pupils.* Assuming a frontal view of the face initially, we locate the pupils by looking for two dark pixels that satisfy certain geometric constraints [47] and that lie within the eye region detected by template matching.

2.8.2. *Searching the lip corners.* First, the approximate positions of lip corners are predicted, using the position of the face and the pupils. A small region around those points is then extracted and a Sobel horizontal edge detector is applied. The approximate horizontal boundaries of the lips are determined by first computing the vertical integral projection P_v of this horizontal edge image:

$$P_v(y) = \sum_{x=1}^{H} E_h(x, y), \quad 0 \leq y \leq W \tag{6}$$

where E_h(x,y) is the intensity function of the horizontal edge image, and W and H are the width and height of the search region, respectively.

The approximate left and right boundaries of the lips are located where P_v exceeds or falls below a set threshold, t. We chose t to be the average of the projection P_v. The vertical position of the left and right lip corners are simply located by searching for the darkest pixels along the columns at the left and right estimated boundaries of the lips in the search region.

2.8.3. *Searching the nostrils.* Similar to searching for the pupils, the nostrils are located by searching for two dark regions, which satisfy certain geometric constraints [47]. The search region is restricted to an area below the pupils and above the lips.

2.8.4. *Tracking.* For tracking pupils and nostrils in the video sequence, simple darkest pixel search in the search windows around the last positions is used. In order to track lip corners, the same detection process is performed in a small region around the last lip corners position. Figure 4 shows the the detected facial features.

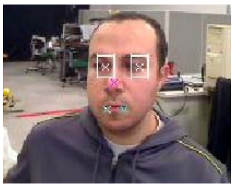


Figure 4. An example of facial feature detection

3. **Verification Experiments.** The system was tested using a computer with the following specifications: Intel (R) Core Due CPU, 2.10 GHZ and 3G RAM.

The system was tested using six different video sequences captured inside our laboratory to track human under some difficult situations such as orientation, illumination changes and movement. In this work, we used two cameras to provide different views of the human face. This considerably improves the accuracy of the system.

The experiments consist of three stages. The first stage tracks the human face using the skin color model. The second shows the ability of the system to tracking humans using the proposed eye tracker by following the eye template through the video frames using template matching. Finally, the third stage which is the core of the work, tracks humans using facial feature in video frames [48].

3.1. **Face tracking.** Figure 5 illustrates an example for human face tracking for two persons. As shown in the figure, the proposed system tracked the human successfully. The ability to track faces in different orientations is also shown on the right.

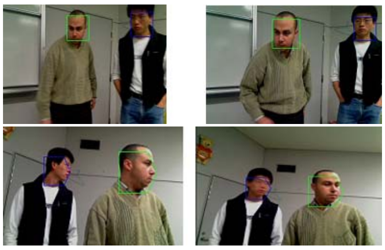


FIGURE 5. Top 2: An example for human face tracking for two persons: (left) Frame 306, (right) Frame 400, Bottom 2: human face tracking in different orientations of the face: (left) Frame 323, (right) Frame 390

3.2. **Eye tracking.** Figure 6 show examples of eye tracking using the eye tracker. After choosing a template for the eye in the very first frame, the system successfully tracked the person on all the following frames.

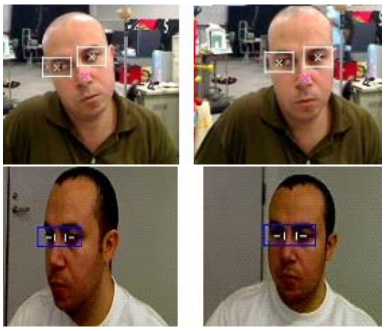


FIGURE 6. Top 2: An example of human eye tracking using eye tracker: (left) Frame 101, (right) Frame 119, Bottom 2: Localized eyes from face images: (left) Frame 132, (right) Frame 207

3.3. **Facial features tracking.** Initially, in face tracking, color-based frontal view face detection is performed in the case of rotation variations. Once a face is detected, its facial features are located and tracked in the image sequence. If there is more than one face in the first frame, only the bigger face is considered. The face detection procedure has proved to be robust in identifying skin color of people in different rotation as seen in Figure 7.

In addition, color-based frontal view face detection is performed and the frames are captured at different distances between the person and the camera and considering varying illumination. The system produced good results especially in some difficult situation such as a hand or other objects occluding the face, as well as under strong varying illumination as seen in Figure 7.

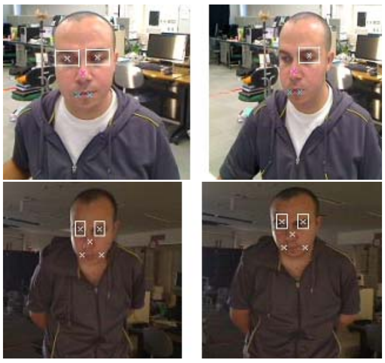


Figure 7. Top: Video sequence one, shows a person tracked with different rotation: (left) Frame 193, (right) Frame 198, Bottom 2: Tracking under different light conditions movement: (left) Frame 59, (right) Frame 243

Moreover, the system was challenged by a person making large movements. During the tracking process, we verified that the system successfully tracked the person, Figure 8. Finally, the system's ability to track a person in the presence of others was also confirmed, Figure 8.

Moreover, to detect sleeping persons, we alter the algorithm to the mark such situations using straight lines. For smiling persons, the eyes are marked using a white star and the mouth using a yellow star, Figure 9.

3.4. **System performance evaluation and comparison.** In this section, we report on the performance of the proposed method and compare it with three other traditional methods: the template matching (TM) [49], the feature point tracking method (FPT) [50] and RU method [3].

The TM method uses a 2-dimensional face region as its template. An initial template is generated from the sample images and the template is updated every frame by adding the detected face region to the template. The FPT tracking method selects features within a face region which are optimal for tracking, and keeps track of these features. For

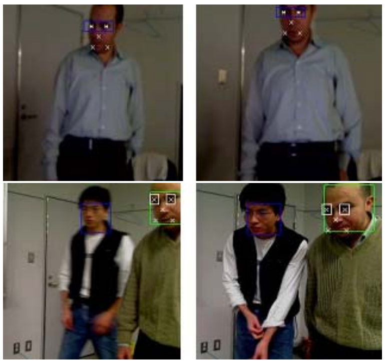


Figure 8. Top: Video sequence one, shows a person tracked with different rotation: (left) Frame 193, (right) Frame 198, Bottom 2: Tracking under different light conditions movement: (left) Frame 59, (right) Frame 243

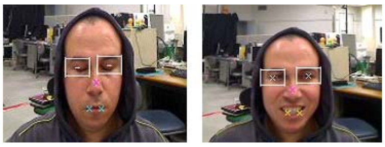


Figure 9. Sleep and laughter detection: (left) Frame 51, (right) Frame 59

our experiments, the tracking process of both methods has been initialized by using the proposed face detector tracking in each frame. In the FPT tracking method, as soon as any of the features within the face region disappear, the face detector is used to reinitialize the tracking. The RU method for the face template is represented by two projection histograms of the face region and matching methods to determine the candidate face region.

In order to determined the feature detection rate, Equation (7) is used:

$$\text{Detection Rate} = \frac{\text{No. of Features detected}}{\text{Total No. of Features}} \tag{7}$$

Six video sequences captured under different conditions were used to evaluate the system. The video sequences are labelled $T1$ to $T6$. Table 2 shows the results of the experiments.

Table 2. Detection rates and processing times

Try	Detection Rates (%)			Process Time
	Eyes	Nose	Mouth	(sec/frame)
T1	99	96	97	0.040
T2	96	93	95	0.035
T3	98	94	96	0.037
T4	96	93	95	0.036
T5	98	94	96	0.037
T6	99	96	97	0.040
Average	97.4	94.3	96.3	0.0375

Table 3 shows the comparison of our method with the three traditional methods.

Table 3. Comparison with other traditional methods

Method	Detection Rate Average (%)	Process time (sec/frame)
Proposed	96.3	0.0375
RU	84.3	0.036
TM	74.8	0.053
FTP	67.7	0.033

Tables 2 and 3 show the comparison results performed on the TV broadcast sequences. In terms of the detection rate, the proposed method gives improvements of up to 12%, 21.5% and 28.6%, compared with RU, TM and FPT, respectively.

3.5. **Discussions.** The RU method overcomes the drift problem due to the face verification and refinement process based on information about the facial features. However, the results are not perfect. On the other hand, the proposed method overcomes this problem using the separated information of face: Eyes, Nose and Mouth. Compared with TM method, the proposed method effectively tracks faces with pose variations. In particular, the TM method fails to track the face correctly due to the template drifting problems. Compared with FPT, the proposed method can reliably track faces not only for frontal faces, but also for faces viewed from other orientations, such as left, right and front. The results obtained are satisfactory and better than those obtained by the FPT method.

In terms of tracking speed, the proposed method takes on average 0.0375 seconds per frame which is faster than TM but a little slower than FPT and RU. As already mentioned, the uses of multiple templates and frames usually need more computation and time. The time needed for our tracking system to process a frame is a function of the type of movement occurring in the scene. The fastest processing time is obtained when no motion is detected. In these situations, no template shifting occurs. The slowest processing time is obtained when large head translation and rotation are detected.

4. **Conclusions.** The paper has described a novel strategy for both face and facial features detection. Face detection is important because it restricts the field of view and thus reduces the amount of computation for further face recognition or transmission, while facial feature detection is important because it enables face normalization and leads to size invariant face recognition.

The work has presented a real time system for detection and tracking of facial features in video sequences by developing a computer vision approach using correlation-based template matching. A skin-color model to segment face-candidate regions in the image is used. The presence or absence of a face in each region is verified by means of an eye detector, based on an efficient template. Once a face is detected, the pupils, nostrils and lip corners are located and tracked in the image sequence. System Features

1. Invariant to different lighting
2. Tracking even with multiple subjects
3. Detection rate of 96.3%.

The proposed system is efficient in tracking a moving human in real time under varying illumination, different movement and rotation. This research area has many applications in face identification systems, model-based coding, gaze detection, human-computer interaction, teleconferencing, etc. Furthermore, we intend to improve the system robustness by developing a recovery module and considering motion to increase the detection accuracy and tracking.

REFERENCES

[1] A. A. Argyros and M. I. A. Lourakis, Real-time tracking of multiple skin-colored objects with a possibly moving camera, *Proc. of European Conference on Computer Vision*, pp.368-379, 2004.

[2] S. D. Lin, J.-H. Lin and C.-C. Chiang, Combining scale invariant feature transform with principal component analysis in face recognition, *ICIC Express Letters*, vol.3, no.4(A), pp.927-932, 2009.

[3] H. Ryu, M. Kim, V. Dinh, S. Chun and S. Sull, Robust face tracking based on region correspondence and its application for person based indexing system, *International Journal of Innovative Computing, Information and Control*, vol.4, no.11, pp.2861-2874, 2008.

[4] C.-Y. Chang and H.-R. Hsu, Application of principal component analysis to a radial-basis function committee machine for face recognition, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4145-4154, 2009.

[5] J.-B. Li, Nonparametric kernel discriminant analysis for face recognition under varying lighting conditions, *ICIC Express Letters*, vol.4, no.3(B), pp.999-1004, 2010.

[6] M.-H. Yang, D. Kriegman and N. Ahuja, Detecting faces in images: A survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.1, pp.34-58, 2002.

[7] Q. Ji, H. Wechsler, A. Duchowski and M. Flickner, Special issue: Eye detection and tracking, *Computer Vision and Image Understanding*, pp.1-3, 2005.

[8] A. Haro, M. Flickner and I. Essa, Detecting and tracking eyes by using their physiological properties, dynamics, and appearance, *IEEE International Conference on Computer Vision and Pattern Recognition*, vol.1, pp.163-168, 2000.

[9] Z. Zhu, Q. Ji and K. Fujimura, Combining kalman filtering and mean shift for real time eye tracking under active ir illumination, *International Conference on Pattern Recognition*, pp.318-321, 2002.

[10] G. Yang and T. S. S. Huang, Human face detection in complex background, *Pattern Recognition*, vol.27, no.1, pp.53-63, 1994.

[11] T. K. Leung, M. C. Burl and P. Perona, Finding faces in cluttered scenes using random labeled graph matching, *Proc. of IEEE Int. Conf. Computer Vision*, pp.637-644, 1995.

[12] S. McKenna, Y. Raja and S. Gong, Tracking color objects using adaptive mixture models, *Image and Vision Computing*, vol.17, no.3, pp.223-229, 1998.

[13] K. Sobottka and I. Pittas, Face localization and feature extraction based on shape and color information, *Proc. of IEEE Int. Conf. Image Processing*, 1996.

[14] V. Govindaraju, Locating human faces in photographs, *Int. J. Computer Vision*, vol.19, no.2, pp.129-146, 1996.

[15] A. Lanitis, C. J. Taylor and T. F. Cootes, An automatic face identification system using flexible appearance models, *Image and Vision Computing*, vol.13, no.5, pp.393-401, 1995.

[16] M. A. Turk and A. P. Pentland, Face recognition using eigenfaces, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.586-591, 1991.

[17] K.-K. Sung and T. Poggio, Example-based learning for view-based human face detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.20, no.1, pp.39-51, 1998.

[18] H. Rowley, S. Baluja and T. Kanade, Neural network-based face detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.20, no.1, pp.23-38, 1998.

[19] E. Osuna, R. Freund and F. Girosi, Training support vector machines: An application to face detection, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.130-136, 1997.

[20] A. Rajagopalan, K. Kumar, J. Karlekar, R. Manivasakan, M. Patil, U. Desai, P. Poonacha and S. Chaudhuri, Finding faces in photographs, *Proc. of IEEE Int. Conf. Computer Vision*, pp.640-645, 1998.

[21] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. of IEEE Int. Conf. Comp. Vision and Pattern. Rec.*, 2001.

[22] R. Kothari and J. L. Mitchell, Detection of eye locations in unconstrained visual images, *ICIP*, vol.3, pp.519-522, 1996.

[23] Z. H. Zhou and X. Geng, Projection functions for eye detection, *Pattern Recognition*, vol.37, no.5, pp.1049-1056, 2004.

[24] T. D'Orazio, M. Leo, G. Cicirelli and A. Distante, An algorithm for real time eye detection in face images, *ICPR*, pp.278-281, 2004.

[25] T. Kawaguchi, D. Hidaka and M. Rizon, Detection of eyes from human faces by hough transform and separability filter, *ICIP*, vol.1, pp.49-52, 2000.

[26] L. Wiskott, J.-M. Fellous, N. Kruger and C. von der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp.775-779, 1997.

[27] J. Huang and H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions (RBFs), *International Journal of Pattern Recognition and Artificial Intelligence*, vol.13, no.7, pp.1009-1026, 1999.

[28] R. S. Feris, J. Gemmell, K. Toyama and V. Kruger, Hierarchical wavelet networks for facial feature localization, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.118-123, 2002.

[29] Y. Ma, X. Ding, Z. Wang and N. Wang, Robust precise eye location under probabilistic framework, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.339-344, 2004.

[30] I. Fasel, B. Fortenberry and J. Movellan, A generative framework for real time object detection and classification, *Computer Vision and Image Understanding*, vol.98, no.1, pp.182-210, 2005.

[31] J. Brand and J. Mason, A comparative assessment of three approaches to pixel level human skin detection, *Proc. of IEEE Int. Conf. on Pattern Recognition*, pp.1056-1059, 2000.

[32] L. Sigal, S. Sclaroff and V. Athitsos. Estimation and prediction of evolving color distributions for skin segmentation under varying illumination, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1883-1892, 2000.

[33] C. Liu, A bayesian discriminating features method for face detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.25, no.6, pp.725-740, 2003.

[34] I. Matthews, T. Ishikawa and S. Baker, The template update problem, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.6, pp.810-815, 2004.

[35] M. Gleicher, Projective registration with different decomposition, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.331-337, 1997.

[36] G. Hager and P. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.20, pp.1025-1039, 1998.

[37] I. Matthews, T. Ishikawa and S. Baker, The template update problem, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.6, pp.810-815, 2004.

[38] M. Black and A. Jepson, Eigen tracking: Robust matching and tracking of articulated objects using a view-based representation, *Proc. of the European Conference on Computer Vision*, pp.329-342, 1996.

[39] N. Gupta, P. Mittal, S. Roy, S. Chaudhury and S. Banerjee, Condensation-based predictive eigen tracking, *Proc. of the Indian Conference on Computer Vision, Graphics and Image Processing*, 2002.

[40] J. Ho, K.-C. Lee, M.-H. Yang and D. Kriegman, Visual tracking using learned linear subspaces, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol.1, pp.782-789, 2004.

[41] C. Grable, T. Zimber and H. Niemann, *Model Based Tracking*.

[42] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd Edition, Prentice Hall, India, 2002.

[43] B. Moghaddam and A. Pentland, Face recognition using view-based and modular eigenspaces, *Proc. of Automatic Systems for the Identification and Inspection of Humans, SPIE*, vol.2277, 1994.

[44] R. T. Kumar, S. K. Raja and A. G. Ramakrishnan, Eye detection using color cues and projection functions, *IEEE ICIP*, 2002.

[45] R. Brunelli and T. Poggio, Face recognition: Features versus templates, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.15, no.10, pp.1042-1052, 1993.

[46] R. Stiefelhagen and J. Yang, *Gaze Tracking for Multimodal Human Computer Interaction*, University of Karlsruhe, http://werner.ira.uka.de/ ISL.multimodal.publications.html, 1996.

[47] R. Joan-Arinyo and A. Soto-Riera, Combining constructive and equational geometric constraint-solving techniques, *Trans. on Graphics*, vol.18, no.1, 1999.

[48] I. Hemdan and K. Terada, Video surveillance using facial features, *Proc. of the 16th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Hiroshima, Japan, 2010.

[49] Z. Liu and Y. Wang, Major cast detection in video using both speaker and face information, *IEEE Trans. on Multimedia*, pp.53-56, 2000.

[50] T. Burghardt and J. Calic, Analysing animal behaviour in wildlife videos using face detection and tracking, *IEE Proc. of Vision Image and Signal Processing*, vol.153, pp.305-312, 2006.