CHAPTER 4

TRACKING AND CONSISTENT LABELING OF HUMAN IN A HOMOGENEOUS ENVIRONMENT

4.1 INTRODUCTION

Video indexing and retrieval of evidence regarding the occurrence of an event are necessary for many security and vigilance applications. The ultimate goal in all these applications is to provide surveillance, detection and alerts. In order to detect the possible threats or suspicious activities the informations like ‘the number of people presented in the scene’ and ‘the identity of each person’ are needed. In recent years, tracking algorithms on video surveillance have gained a lot of attention. The idea of tracking during video analysis of human motion has two components: (1) motion segmentation and (2) temporal correspondences. Motion segmentation is the process of separating the objects of interest (humans) from the rest of the image (the background). The temporal correspondences, is the process of associating the detected humans in the current frame with those in the previous frames and hence providing temporal trajectories through the state space.

The tracking methods are generally categorized based on the image measurements such as region, active-contour, feature and model. The feature-based analysis is one of the tracking methods to identify a person and to analyze the action. They aim at tracking local features, like corners or vertices inside the given region (Tomasi and Kanade 1991) or more global features of the object like perimeter, area of the object region (Polana and Nelson 1994).
Once the tracking is commenced, the problem is to find the temporal correspondences between predicted and measured states. Furthermore, interactions between humans, during occlusions, are difficult to analyze, when they are represented as one foreground object. The problem of finding the correspondences and identity establishment during tracking, namely labeling in normal sequence and also under occlusion, have been investigated recently (Moselund et al 2006).

The objective of any feature-based tracking algorithms is that it should provide the identity of the object by exploiting information on object motion, local features (corners, vertices), global features (centroid, boundary) and dependence graphs (distances, geometry). But, when the objects are having similar color distributions, labeling the objects consistently during tracking under occlusion is a challenging one. The identities provided by the features like centroid and bounding box would fail when the objects are occluded due to the similarity. This inconsistency in labeling leads to confusions and wrong identifications during tracking and also, for later processes such as person identification and action analysis. A novel feature based algorithm is developed in this chapter to circumvent the problem of inconsistency in labeling while people wear similar color dress.

4.2 TRACKING AND CONSISTENT LABELING

One of the main issues in tracking and consistent labeling is, the difficulty in distinguishing a moving object from its shadow. A lot of work has been reported in the literature on shadow removal which is discussed at length in Chapter 2. Shan et al (2007) studied the features of moving object and shadow in different color spaces, and a model for moving object detection is derived using grey and gradient of shadow. Liu et al (2007) presented a method using information from pixel-level, local-region level and global level
to remove shadow. At the pixel-level, Gaussian Mixture Model (GMM) is combined with homomorphic filter to model surface reflectance value which fits Gaussian distribution because the GMM deals with slow lighting changes, periodic motions. Yet, there are several practical issues concerning the use of the existing algorithms for the shadow modeling as each of them involve different parameter tuning for indoor and outdoor shadow removal. Therefore, an effective shadow modeling algorithm which is suitable for both significant and insignificant shadows is to be developed (Martel and Zaccarin 2007).

Despite the merits of feature-based tracking techniques that rely uniquely on color, there are several serious deficiencies in tracking algorithms. In the scenario where the color distribution of the object is similar to the background, the tracking algorithm is unable to retrieve sufficient information to determine the position of the object (Huang et al 1999). If a sub-region of the object has the same color distribution as the whole object, pure color-based tracking techniques might be unable to properly determine the real size of the object (Krumm et al 2000). This leads to the tracking of a sub-region of the object instead of the whole object. These scenarios promote the issues of determination of the position and size of the object in surveillance applications. Also, when several objects with similar color distributions are tracked simultaneously, without depth information, the evaluation of the object’s identities makes confusion. Nakajima et al (2000) used color histograms based appearance modeling as they are relatively invariant to changes in object orientation, scale, partial occlusion, and viewing position and object deformation. This method captures only the color distribution in an image and do not include any spatial correlation information. Therefore it has limited discriminative power.
The simplest scheme reported in the literature to establish consistent labeling may be to match color or other features of objects being tracked in each camera, to generate correspondence constraints. This matching may be performed statistically in a Kalman Filter framework (Utsumi and Ohya 2000) or using a Bayesian Network approach (Chang and Gong 2001). In both cases, the authors do not restrict themselves to a single type of features using single camera but applicable for multi camera environment. For a single camera system, McKenna et al (2000) proposed an approach using the appearance features such as color, shape and texture to re-establish identity. The system ‘W4’ proposed by Haritaoglu et al (2000) uses grayscale texture, silhouette shape information as well as a dynamic template. However, they do not address the issue of splitting and of re-establishing identity. Piater and Crowley (2001) detect occlusion by considering the interaction of adaptive Gaussian regions of interest. Their technique allows them to avoid merging blobs that barely occlude each other. Senior et al (2006) used an appearance based model to track humans through occlusions. These models are used to localize objects during partial occlusions, detect complete occlusions and resolve depth ordering of objects.

Li et al (2002) followed a different approach and model the appearance of humans and objects using a color correlogram. Dan and Yuan (2007) proposed a multi-object motion-tracking method based both on region and feature. Munoz-Salinas et al (2008) proposed a multiple object tracking approach that combines color and depth information using a confidence measure to address the problems such as occlusion and also, the case when both the background and objects have similar color. Still constraints are added to disambiguate cases in which objects are in close proximity to each other. Such constraints exploit some property of the objects before and after the ambiguous event assuming that this property will be invariant; for example,
constant velocity assumptions, motion uniformity constraints, object shape models, and color features.

The tracking algorithm is generally developed, to act as an intermediate level of processing, getting input from low level background subtraction, and feeding higher level analysis such as person identification, or activity analysis. In this chapter, to describe the approach ‘consistent labeling’ rather than ‘tracking’, as the term ‘tracking’ generally implies reliance on a motion model or spatial constraints, alone or in combination with visual appearance constraints, while ‘consistent labeling’ relies entirely on visual appearance constraints. The specific assumptions needed are the information about the camera angle or position when video be taken with a static camera. Also, the image environment such as education institutions, offices, shopping malls and airport, where several people wearing similar color dress is considered for analysis.

4.3 MOTIVATION

The primary objective is to maintain a consistent labeling of humans through a video sequence, maintaining continuous labeling even if a person leaves the scene for a period of time and returns. Many tracking and labeling algorithms using multiple cameras are reported in the literature. But with a single camera and assuming that the face region is visible in the camera view, the problems identified are, effective shadow elimination to improve the classification accuracy, feature selection for providing consistent labeling even under occlusion. Also, keeping the identity of people during tracking even when they merge into groups or separate from groups is another issue. The algorithm should be robust to variations in view point (Stable labeling performance even during changes in view of camera). Hence, the objective of the proposed algorithm is to provide tracking and labeling group of persons
consistently using face skin-color model in the optimal color space to address the issue of consistent labeling in a homogeneous environment. In order to segment skin regions from non-skin regions and to extract face regions, a reliable skin color model is needed. Also, the algorithm which addresses the issue of adaptability of the skin color model to people of different skin colors is required.

4.4 PROPOSED METHODOLOGY

As a first step, the background subtraction is done to remove the shadow using a homomorphic filter and GMM. To identify the person in the group in the homogeneous environment, the statistical skin-color model is generated by means of a supervised training. The flowchart of the proposed algorithm for labeling is shown in Figure 4.1. The proposed statistical skin color model distinguishes skin regions and non-skin regions based on the color intensity variation. Although skin colors of different people are nearly the same, they differ in intensities. The skin color distribution is represented by a Gaussian model and the likelihood of skin for this pixel is then computed using statistical formulae (Yang and Ahuja 1998).

On a color normalized images, the ‘likelihood’ estimation of a pixel being skin is mostly high. So, an optimal global thresholding on ‘likelihood’ can be considered as the indicator of skin regions. Such skin regions are labeled by connected component analysis and each label is identified with distinct colors. The face region is extracted using region descriptors like filled areas and number of holes. Subsequently, the histogram of each face region is plotted, and the peak value of the histogram (peak intensity/color) of the individual face region identifies the person depending on the peak amplitude. At the same time, the tracking is realized by extracting the features such as centroid and area, average luminance of moving objects and matching by
similarity functions such as weighting measuring function and distance measuring function. The similarity functions are used to match object’s characteristic so as to make sure whether they are similar which is applicable for the environment where people in different color attires. Along with these features, matching functions and face skin color based identity; the people have been tracked with consistent labeling.

![Diagram](image)

**Figure 4.1 An overview of the Proposed Algorithm**

4.4.1 Shadow Classification and Optimal Color Space Selection

The rule for the classification of shadows (Shan et al 2007) was discussed in section 2.4.3 of Chapter 2. Based on the pixel being classified whether as significant or insignificant the appropriate color space is chosen for the background subtraction. Analysis of the performance of various color spaces over the shadow removal on a variety of datasets, Shan et al (2007) suggested that the scale invariant property of HSV color space is suitable for representing significant shadow, since the significant shadow is less affected
in the luminance space, whereas the shift invariant property of YCbCr color space is suitable for representing insignificant shadows. In this chapter, the experiments are focused on indoor sequences since it is assumed that the image of the human is relatively large with respect to the artifacts introduced by the background subtraction stage. This is a reasonable assumption for indoor but might not be always valid for outdoor sequences, where scenes often have a wide range of depths and the sizes of humans and objects can be very small (Liu et al 2007).

4.4.2 Background subtraction and shadow removal

After the selection of color space using shadow classifier, the background is modeled and the shadow is removed by Gaussian Mixture Model (GMM) and the homomorphic filter respectively. The intensity of a pixel in an image, $S_i(x, y)$, at time instant $t$, can be modeled as shown in Equation (4.1).

$$S_i(x, y) = E_i(x, y)\rho_i(x, y)$$

where, $x$ and $y$ are the coordinates of pixel, $E_i(x, y)$ is the irradiance and $\rho_i(x, y)$ is the reflectance of the object surface. Liu et al (2006) found that the scene illumination component $E_i(x, y)$ changes slowly, except the transition from the illuminated area to umbra area (penumbra), whereas the reflectance component contains medium high-frequency details (object information). Two assumptions made were: penumbra changes are slower than the reflectance value of the surface and the shadow region is minimum fifty pixels area. Since, the $\rho_i(x, y)$ component only is connected with the object surface, if $S_i(x, y)$ has been separated as $\rho_i(x, y)$ and $E_i(x, y)$ component, the shadow can be easily removed. As described above, the $\rho_i(x, y)$ component has more medium high-frequency component and homomorphic filtering is used to extract the reflectance component and after
the filtering, values of each pixel represent a measurement of the high
frequency component of object surface. The frequency response of
homomorphic filtering also fits Gaussian distribution well as shown in
Figure 4.2. Hence, Gaussian Mixture Model (GMM) which can deal with
slow illumination changes and clutter is chosen here to model the
background.

![Frequency Response of the Homomorphic Filter](image.png)

**Figure 4.2 Frequency Response of the Homomorphic Filter**

GMM models the background, with ‘$K$’ Gaussian distributions of
pixels as discussed in section 2.4.1. Once the background has been subtracted,
the residue of the chosen image region represents objects and possibly the
shadows.

4.4.3 Skin-Color Model

To identify the person in a group in the homogeneous environment,
the statistical skin-color model is generated by means of a supervised training,
using a set of skin-color region samples. The samples are obtained from the
color face databases from YALE face database\(^1\), CMU face database\(^2\), AT and
T (Olivetti) Database\(^3\) and from college campus database with 140 face

1. [http://cvc.yale.edu](http://cvc.yale.edu)
2. [http://www.cs.cmu.edu/~har](http://www.cs.cmu.edu/~har)
3. [http://www.uk.research.att.com](http://www.uk.research.att.com)
images. Such images were obtained from people of different races, ages and gender, with varying illumination conditions. For the identification of face regions using skin color, to separate skin objects from non-skin objects like wood, which can appear to be skin colored is important. The cluster for human skin generally form a continuous homologous series because of characterization caused by absorption of melanin and hemoglobin to separate skin objects from those which have skin colored appearance in chrominance space. Yet, the color space selection to develop the skin color model is another issue because the human skin-color is distributed in RGB color space. But, in the chromatic color space, the color distribution of skin colors of different people is found to be clustered in a small area as shown in Figure 4.3.

![Figure 4.3 Skin cluster in Chromatic Space](image)

Although skin colors of different people are very close, but they differ mainly in intensities. With this finding, to develop a skin-color model in the chromatic color space is preceded. Here, the YCbCr color space is adopted since it is perceptually uniform, similar to the TSL (Tint, Saturation and Luma) space in terms of the separation of luminance and chrominance as well as the compactness of the skin cluster.
First, the input color image is converted from RGB to YCbCr color space. The conversion from RGB color space to YCbCr color space can be done by using Equations (4.2) through (4.4). Therefore, if a pixel, in the chromatic color space, has a value of \((Cb, Cr)\).

\[
Y = (0.299*R) - (0.587*G) - (0.114*B) \tag{4.2}
\]
\[
Cb = (0.5*R) - (0.169*R) - (0.331*G) + 128 \tag{4.3}
\]
\[
Cr = (0.5*R) - (0.419*G) - (0.081*B) + 128 \tag{4.4}
\]

A statistical skin-color model is generated by means of a supervised training, using a set of 40 skin-color regions of size 45 x 45, obtained from a color face database. Thus, a skin color distribution can be represented by a Gaussian model \(N(m, C_r)\), where, the mean \((m)\) is obtained for the chrominance component \((x_c)\) by Equation (4.5),

\[m = E[x_c], \text{ where } x_c = (Cb \quad Cr)^T\]  \tag{4.5}

and the covariance is obtained using Equation (4.6),

\[C_c = E[(x_c - m)(x_c - m)^T]\] \tag{4.6}

where, \(T\) stands for transpose. Figure 4.4 shows the Gaussian distribution \(N(m, C_r)\) fitted using the sample skin color data base.

Figure 4.4 Skin Color model
With this Gaussian fitted skin color model, the likelihood of skin for any pixel of an image can be obtained. The likelihood of skin for this pixel $L(i, j)$ can then be computed using Equation (4.7),

$$L(i, j) = \exp(-0.5(x - m^T)C_e^{-1}(x - m))$$

(4.7)

where, the matrices $x$ (chrominance component), the mean $m$ and $C_e$ (covariance of $Cb, Cr$) are obtained from the Equations (4.5) and (4.6). This skin-likelihood image is obtained as a gray-scale image whose gray values represent the likelihood of the pixel belonging to skin.

The next step is to segment skin and non skin regions. The important point here is that this process can reliably point out the non skin regions are need not to be considered for the further process of finding face region. Since the skin regions are brighter than the other parts of the images, these regions can be segmented from the rest of the image through a thresholding process. To process skin color of different people, a fixed threshold value is not sufficient. An optimal global thresholding process is used and to obtain the threshold ‘$T_r$’ automatically, the algorithm is as follows.

1. Selection of an initial estimate for $T_r$.
2. Segmentation of the image using $T_r$ which produces two groups of pixels, $G_1$ consisting of pixels with gray level values greater than $T_r$ and $G_2$ consisting of pixels with values less than or equal to $T_r$.
3. Computing the average gray level value, $m_1$ and $m_2$ for the pixels in regions $G_1$ and $G_2$.
4. Computing a new threshold value as shown in Equation (4.8).

$$T_r = 0.5(m_1 + m_2)$$

(4.8)

5. Repeating steps 2 through 4 until the difference in $T_r$ in successive iterations is smaller than a predefined small value. After thresholding, a
binary image is obtained with white pixels representing the skin regions.

4.4.4 Face Detection and Identification

A set of skin color samples database is given as input and the skin-likelihood image is obtained from the chrominance components using the skin model based on the Equation (4.10). After, the skin-segmented image is obtained by using an optimal global thresholding, the skin regions are labeled with distinct colors. From the skin regions, the exact face region boundary is extracted.

4.4.4.1 Connected Component Analysis

The foreground region, representing skin region is marked by a ‘1’ (white pixel) after the previous segmentation process. The aim is to find the regions that are exclusive to face. This could be realized by analyzing the connected components. The face is represented by largest connected component.

Therefore, the effort is to label the foreground components based on the connectivity. A label is an integer value. An 8-connected neighborhood (considering all the neighbors of a pixel), shown in Figure 4.5, is used to determine the labeling of a pixel. If any of the neighbors had a label, the current pixel is assigned with that label. Otherwise, a new label will be assigned. At the end, the number of labels is counted and this will be the number of regions in the segmented image.
Then, the number of holes (Black pixel) in each connected component is determined by computing the Euler number of the region, which is defined in Equation (4.9),

\[ E = C_{co} - H_0 \]  \hspace{2cm} (4.9)

where \( E \) is the Euler number, \( C_{co} \) is the number of connected components and \( H_0 \) is the number of holes in a region. For the proposed algorithm, the number of connected components (i.e. the skin region) is set to ‘1’ since one skin region is considered at a time. The number of holes is defined by Equation (4.10).

\[ H_0 = 1 - E \]  \hspace{2cm} (4.10)

Next, based on the fact that a skin region corresponding to the face has at least two holes and the assumption that face occupies a significant portion in the input image (\( E = C_{co} - H_0 \), here, \( C_{co} = 1 \) and \( H_0 = 2 \)). Hence, the largest filled area which satisfies the Equation (4.10) is retained in connected component and corresponds to skin region of face. The total number of face regions within the group is identified based on the above criteria. For the identified face regions \( I_f \), the histogram \( h_i \) is obtained as shown in Equation (4.11),

\[ h_i(r_i) = \frac{n_i}{N_i} \]  \hspace{2cm} (4.11)

where, \( i \in [0,1,...,L-1] \) such that , \( n_i \) is the number of occurrence of \( i^{th} \) intensity, \( N_i \) is the total number of pixels , \( r_i \) is the probability of occurrence of gray level \( i \).
and $L$ is the number of Gray levels. The histogram is plotted and according to the value of peak from the plot the person’s ID in the group is assigned since the peak value for each face region is varying. From these values persons attire similar or in a homogeneous environment will be identified correctly.

4.4.5 Proposed Consistent Labeling Algorithm

After detecting all the moving objects with corresponding labels, the objects are tracked. The motion direction, spatial position, shape and luminance are the most widely used features for feature based tracking. But in fact the new entry or overlapping often happens unexpectedly will affect tracking efficiency. So, an associated method is applied to achieve steady tracking. Three features such as centroid, area, and average luminance of each blob are extracted for tracking. The three features need to be extracted are centroid, area and mean luminance. Centroid extracting is to show each object’s spatial position in the monitoring field. It is a clue to tracking and matching the distance between moving objects. In the present frame $(F_n)$, centroid of labeled block $(BO^i_n)$ is given in Equations (4.12) and (4.13).

$$x_c = \frac{1}{A(BO)} \sum_{x \in BO} x_i$$  \hspace{1cm} (4.12)

$$y_c = \frac{1}{A(BO)} \sum_{y \in BO} y_i$$  \hspace{1cm} (4.13)

where $x_c, y_c$ are the $x, y$ co-ordinates of the centroid, $(BO^i_n)$ represents the labeled block, $i$ and $n$ are corresponding to number of binary blocks and frame number, $A(BO)$ is the labeled block’s area, which is adopted as follows in Equation (4.14).

$$A(BO) = \sum_{(x,y) \in BO} 1$$  \hspace{1cm} (4.14)

Each moving object’s average luminance $(G^i_n)$ is defined as in Equation (4.15).
Figure 4.6 presents the means of borderline extraction which is defined as the minimal rectangle that surrounds the borderline. The rectangle’s height is \( y_{\text{max}} - y_{\text{min}} \) and width is \( x_{\text{max}} - x_{\text{min}} \).

**Figure 4.6 Extraction of borderline**

Also, two similarity functions proposed by Dan and Yuan (2007) are used to match objects characteristic so as to make sure whether they are similar. One is the weighting measuring function that can resist the change of luminance and object’s distortion in some degree and the other function is the moving object’s distance measuring function between the successive two frames. If the old object can move independently and the distance between two frames is smaller than the distance measuring function value, substitute the old for the new. It is assumed that \( F_n \) and \( F_{n-1} \) are the two sequential frames, there are ‘i’ binary blocks in \( F_n \) and ‘j’ in \( F_{n-1} \). Then, the new object in \( F_n \) is matched with the old one in \( F_{n-1} \) and it is considered to be matched if they are similar using the similarity functions. The average luminance difference \( (\gamma_n^i) \) and area difference \( (\beta_n^j) \) are used to obtain the similarity functions from the following Equations (4.16) and (4.17) with the use of \( G_n^i \) and \( A(BO_n^i) \):

\[
G_n^i = \sum \frac{F_n(BO_n^i)}{A(BO_n^i)}
\]
Average Luminance Difference,
\[
\gamma_i^{ij} = C1 \left[1 - \frac{(G_i^j - G_n^i)}{G_n^i}\right]
\]  \hspace{1cm} (4.16)

Area Difference,
\[
\beta_i^{ij} = C2 \left[1 - \frac{(A(BO_n^i) - A(BO_n^i))}{A(BO_n^i)}\right]
\]  \hspace{1cm} (4.17)

where \(ij\) is the index of the object, and \(C1\) and \(C2\) are weighting coefficient of the average luminance difference and area difference. The weighting function \((\phi_i^{ij})\) can be adopted using Equation (4.18):
\[
\phi_i^{ij} = \sqrt{(\gamma_i^{ij})^2 + (\beta_i^{ij})^2}
\]  \hspace{1cm} (4.18)

The distance similarity function \((\theta_i^{ij})\), is obtained using Equation (4.19):
\[
\theta_i^{ij} = \sqrt{(\bar{X}_n^i - \bar{X}_n^{ij})^2 + (\bar{Y}_n^i - \bar{Y}_n^{ij})^2}
\]  \hspace{1cm} (4.19)

Whether the two objects are matching or not can be judged by the Equation (4.20):
\[
\left[(\phi_i^{ij} < Th_1) \cap (\theta_i^{ij} < Th_2)\right]
\]  \hspace{1cm} (4.20)

where, \(Th_1, Th_2\) are the thresholds of weighting and distance similarity functions. The algorithm for multi-object motion tracking method is as follows. It is assumed that in the image frames, human motion are previously regulated that people come into the eyeshot one by one. Once, a new one is detected, an identification label (ID) in terms of number, is assigned to the person. Each person is described as a vector HB \(((ID, x, y, A, G_n^i))\) where HB is the ID for single person, ID is a unique label assigned to each tracked person \((x, y)\) are the centroid coordinates of the person’s projection and \(A\) is the object’s area, \(G_n^i\) is object’s average luminance. When they move separately, their matching can be found through the two similarity functions as defined above; When two or more persons are close to each other or overlapped, they
are considered to be a group described as GP(ID list, \(x, y\)), where ID list take
down all the person’s ID and \((x, y)\) record the group’s centroid coordinate.
When they turn into a group, track the group using the group vector; If one of
the group members gets separated from the group the algorithm identify the
person’s ID again by human feature model and at the time of the person is
leaving out of scope, logout the person’s ID. This tracking algorithm works
well in situations where people wear dissimilar color attires. But, there comes
confusion in labeling due to the same similarity functions when people wear
similar color dress. This is overcome by the proposed face skin color based
identification along with this feature based tracking.

In order to evaluate the performance of the proposed method, the
performance metric namely Object Tracking Error (OTE) proposed by Bashir
and Porikli (2006) is used. OTE is defined as the average discrepancy between
the ground truth bounding box centroid and the centroid of the proposed
tracking result as given in Equation (4.21),

\[
OTE = \frac{1}{N_{og}} \sum \sqrt{(x_{og}^i - x_{gt}^i)^2 + (y_{og}^i - y_{gt}^i)^2}
\]  

(4.21)

where, \(N_{og}\) represents the total number of overlapping frames between ground
truth and proposed algorithm’s result, \(x_{og}^i\) and \(y_{og}^i\) represent the \(x, y\) coordinates
of the centroid of the object in \(i^{th}\) frame of ground truth, \(x_{gt}^i\) and \(y_{gt}^i\) represent
the \(x, y\) coordinates of the centroid of the object in \(i_{th}\) frame of proposed
algorithm’s result.

4.5 EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the simulation results and the quantitative
comparison of background subtraction methods for the shadow elimination
such as the proposed Homomorphic filter with GMM in RGB and in optimal
color space are presented. First, image sequences in different illumination conditions from benchmark datasets (Hall monitor, Lobby) and from local environments (College Ground and Image Processing Laboratory) were chosen to test the algorithms. The selected datasets taken in different indoor and outdoor environments are Image Processing (IP) Lab, hall monitor, lobby and Ground with frame size of 320 x 240 are used to demonstrate the performance of the proposed method, where shadow is significant and insignificant and the subjects are human. In the proposed work, the Gaussian Mixture Model’s sensitivity has been analyzed to each parameter by observing the variation of the false negative (FN) which is foreground pixel that were missed, and the false positive (FP) which is background pixels that were marked as foreground caused by the change of one parameter while keeping the others unchanged. The measurements were performed on different video sequences as illustrated in Figure 2.6 of section 2.5 in chapter 2. The FN and FP for different parameter values and clearly shows that 3 is suitable value for \( K \) (number of Gaussian mixtures) and 0.8 is an appropriate value for \( T \). The proper value for \( \alpha \) and \( \sigma_k^2 \) are 0.02 and 26 respectively.

Using the optimal color space rule, the \( \xi \) is obtained as 0.63 for IP Lab dataset and the shadow is classified as insignificant shadow. According to that, YCbCr color space is chosen as discussed in section 2.4.3 in chapter 2. The simulation results were shown in Figure 4.7. The visuals showed that Homomorphic filter with GMM in optimal color space produces better results comparatively with Homomorphic filter with GMM in RGB.
Figure 4.7 Background Subtraction and Shadow Removal

In order to evaluate the proposed method, the shadow detection rate $\eta_s$ in % and Recall rate in % are calculated using Equations (4.22) and (4.23).

$$\eta_s = \frac{TP_s}{TP_s + FN_s} \times 100 \quad (4.22)$$

$$R_s = \frac{TP_s}{TP_s + MO_s} \times 100 \quad (4.23)$$

where, $TP$ is the true positives, $FN$ is the false negatives, $MO$ is missing objects and also the subscript ‘s’ stands for shadow. The proposed shadow elimination method using GMM with Homomorphic filter in optimal color space outperforms than in RGB color space and is shown in Figure 4.8 (a) and Figure 4.8 (b).
a) Shadow Detection Rate \( (\bar{\eta}_s) \) in Percentage

**Figure 4.8 Performance Measures**

The scenario, where the proposed skin color model has not been applied for labeling is shown in Figure 4.10. Figure 4.9 (a) shows the Frame number 7, where two students are there in the scene. The Group vector is built and ID is assigned as Group1. It has been tracked continuously up to Frame 20 and for each student the ID is assigned as HB1 and HB2 by the features centroid, bounding box, area and average luminance. The corresponding results are shown in Figures 4.9(b) and (c).

![Frame 7 (Group)](image)

![Feature based Tracking with ID](image)

![Average Luminance Results](image)

**Figure 4.9 Features based Tracking Results**

In Frame 25 one student is leaving as shown in Figure 4.10 (b). Here, the distance measuring function value \( (\bar{\theta}_d^i) \), obtained using Equation (4.19) is twenty five which is larger; hence, individual IDs are given to both.
Figure 4.10 (c) shows the Frame 47 which has kept the same ID for that student. Frame 83 in Figure 4.10 (d) shows the ‘Group’ and in Figure 4.10 (e), Frame 115 shows the misidentification of second label person as the third label person and vice versa due to same similarity functions for both students obtained by Equations (4.18), (4.20) and (4.21). The threshold values \( Th_1, Th_2 \) are taken as twenty and five. The similarity in dress color provides the values of similarity functions as twelve and three and the IDs for both persons are interchanged, hence this misidentification has occurred.

Figure 4.10 Results of Consistent Labeling before Face Skincolor based Person identification (IP Lab) – (Frames 7, 25, 47, 83, 115)

To avoid this inconsistency in labeling (identification), the skin-likelihood image is obtained from the chrominance components using the skin model the Equation (4.7) which is shown in Figure 4.11.
Figure 4.11 Skin Likelihood from the Chrominance Components

The original image and the skin likelihood images are shown in Figures 4.12 (a) and (b). The skin-segmented image is obtained by using an optimal global thresholding. The resulting skin-segmented image shown in Figure 4.12 (c) that clearly indicates the skin regions present in the original image. Using the connected component analysis the face regions are labeled. Here, the number of face regions is three. The extracted face regions using the connected component analysis is shown in Figure 4.12 (d).

Figure 4.12 Face Regions Extraction
Figure 4.13 Histogram for Face Regions

The histogram is plotted for the extracted face regions and is shown in Figure 4.13 to show the different intensity distributions of the different face regions. The peak values for the three face regions of the students are 247, 234, 255 respectively and according to that the identifications (ID’s) are assigned as one, two and three. Since, the histogram peak value is robust to the view invariant, the identities are correctly maintained in misclassified Frame 115 consecutively and are shown in Figures 4.14.

Figure 4.14 Results of the Proposed Algorithm for Consistent Labeling

In order to evaluate the tracking performance, the metric Object Tracking Error (OTE) given in Equation (4.21) is used. The OTE of the proposed method for the IP Lab dataset is 4.6% for fifty six overlapping frames ($N_{sg}$).
4.6 CONCLUSION

This chapter focused on tracking people with consistent labeling in a homogeneous environment like people wear similar attires. In this work, shadow is removed by GMM with homomorphic filtering in optimal color space and tracking is done by extracting the features such as centroid, bounding box, average luminance difference and area difference of moving objects along with the similarity functions of them in a scene. The skin-color model is proposed to obtain the skin likelihood image and segmentation is done using the statistical likelihood measure among the pixels. The connected component analysis is used to extract the face- regions using the histogram. The peak values of the histogram plots confirmed the person’s ID and the assigned ID helped to identify the person correctly throughout the tracking process. It is proved by simulations and validations with real videos that the proposed algorithm provided an OTE of 4.6%. 