Asian Journal of Multidisciplinary Studies

ISSN: 2321-8819 (Online)
2348-7186 (Print)
Impact Factor: 0.92
Vol. 3, Issue 12, December 2015

A Video Surveillance system at Security zone

Abouzar Ghasemi[1] and C. N. Ravikumar[1]

Dept of Computer Science & Engineering
SJ College of Engineering, Mysore-570006, India

Abstract- *In this paper, we introduce an real time video based face recognition framework at security zone. The created framework recognizes subjects(e.g., travellers/customer) while they are entering a security entryway. Uncontrolled varieties of facial appearance because of occlusion, changing brightening, posture or expression of non cooperative subjects should be handled to correct recognition. To achieve this goal, the framework first detects and tracks the eyes for registration. Then registered subject faces are classified by a local appearance based face recognition algorithm individually. In next step by outcome of confidence scores from each classification are progressively combined to get the identity estimate of the all sequences. We propose three unique measures to weight the contribution of all frames individually to determine the general classification. These are distance to model, distance to second nearest representative, and their composition. We have closed set identification experiments on a database of 125 subjects and the result show that the our framework is capable to get more than 92 percent of correct recognition rate.*

Keywords: Video surveillance system, Face detection, Face recognition, Eye tracking, DCT.

Introduction :

Building a strong face recognition framework is one of the greatest difficulties in computer vision research. An extensive variety of conceivable application zones, for example, access control, surveillance systems, video teleconferencing ,smart shop and so on., have filled huge measure of exploration endeavours on this issue. On the other hand, the majority of the studies on face recognition have been directed on information that was gathered under controlled conditions [1]. This kind of information contains changes in facial appearance that are produced by improving a single or a mix of two variety sources in a controlled manner. The fundamental variety sources that have been for the most part centred around are expression, pose, time gap between training and testing data, illumination and occlusion. In spite of the fact that the studies that have been led on this kind of information demonstrate the tried calculations' execution against a particular sort of facial appearance variety and give experiences about face recognition under these particular conditions, they are not adequate to copy real life conditions because of two fundamental reasons. In real life, the varieties of facial appearance are brought about by blends of various sources in a nonstop way, that is, for instance, one needs to manage face images from any view angle, while in the databases gathered under controlled conditions, there are just discrete head pose classes. Therewith, the majority of the face recognition algorithms are tried on cropped and adjusted face images that are enlisted by marked fiducial points on the faces. In any case, it is realized that fault in the record of the face deteriorate performance of face recognition [2] and [3] . In this manner, a real life framework must be sturdy against record errors that may happen because of imperfect fiducial point localization.

The need to focus or confirm the identity of a object in an extensive variety of utilization ranges has additionally prompted numerous business face recognition frameworks. The vast majority of these business frameworks are essentially centred around security related applications, for example, access control or surveillance systems [4], [5], [6], [7] and [8]. What's more, frameworks with multimedia centre are additionally accessible, for example, looking superstars in video [9] or automatic photo labelling [10].

In this paper, we display a strong genuine face recognition framework for shrewd situations that distinguishes the people while they are entering a security door. The principle inspiration to assemble a face recognition framework to screen the individuals entering to air port the extensive variety of utilizations in which it can be utilized. Both for monitoring for security purpose of open regions, e.g., air ports, and for individuals checking in smart regions, e.g., smart homes, one of the best moments to recognize the persons is the minute they are going into the room. This gives face images resolutions going from 45×45 pixels to 100×100 pixels.

Our Method:

In this research work, we building a real time video based face recognition framework for the mentioned real life setting. Sample application region can be airport hall ,hotel's lobby or

store(shop) that can recognize specific passenger or regular guest in hotel or regular customer . The framework comprises of a strong eye tracking algorithm, that gives predictable eye areas to permit face registration, and a video based face classification algorithm, that uses registered face region images to derive identify objects.

The improved face classification framework advantages from local appearance-based face representation [12] and [13] and uses the video data in order to strong handle robust varieties in the information. Two fundamental observation are exploited to determine two distinct approaches to weight the contribution of every individual frame to the general classification result. Distance to model (DTM) is the first in which, considers how alike a test sample is to the representatives of the training set and the next one is distance to second nearest (DTSN), decreases the effect of frames which convey obscure classification results and the third measure is a mixture of the two schemes is utilized.

Video segmentation:

The face recognition framework needs to first detect the moments at which passenger/customer is entering lobby/security door. This subsystem, that we named as face recorder, comprises of three main parts: colour based skin segmentation utilizing ratio histogramming in order to choose face traveller/costomer, feature based face detection to accept or reject them and an fundamental tracking system to guarantee the complete entering sequence is recorded. Fig. 2 gives a diagram of the framework. An example of a recorded objects faces can be found in Fig. 3.

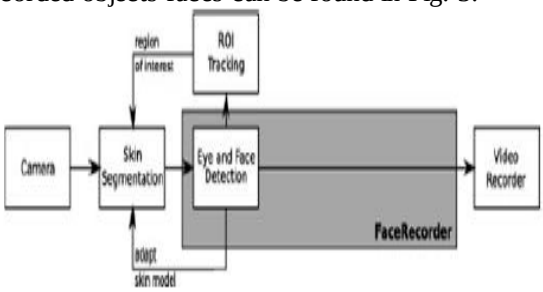

Fig. 2. Overview of the data collection system.

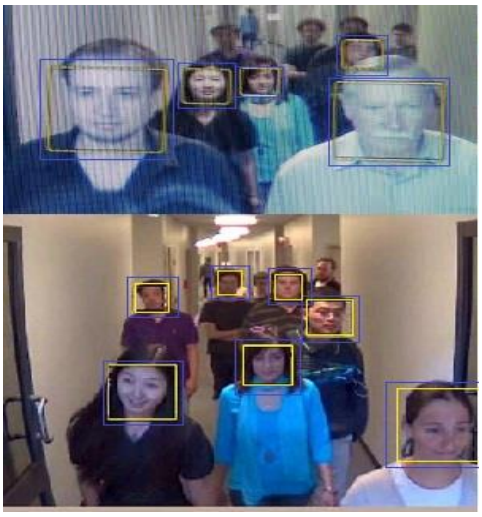

Fig. 3. Example of a recorded passenger face. Blue box shows the search region and the yellow box shows the detected face.

Skin colour segmentation:

In the given situation, the traveller/customer face to be recognized is comparatively small with Considering the image dimensions of 640*480 pixels. To avoid dispensable processing of the background, it is crucial to focus on important areas of the image. To distinguish these regions, the image is analyzed for skin-like colours.

Skin colour representation:

In this paper, a histogram based model of 128*128 is utilized to represent the skin colour distribution and it is found out from a representative training set of skin samples which are physically trimmed from images by selecting vast skin region in faces in a set of input images of passengers. It is non-parametric and makes no prior presumption about the actual distribution of skin colours. The model used in this paper is situated in the normalized rg colour space. The merit of selecting a chrominance-based colour space is a decreased sensitivity to illumination. In the meantime, diverse skin tones, because of variety ethnic background, get more like one another in this representation, forming a compact cluster in colour space. According to a physical model for skin-reflectance, Storring et al [13] demonstrate that the skin colour of objects with distinctive background under illumination of differing colour temperature commonly forms an eye brow like shaped area in the chromaticity plane. This area is generally referred to as skin-reflectance locus or skin locus. Fig. 4a shows the skin model that we got from 354 training samples or, to be more exact, 799,785 training pixels, taken with a Canon VC-C1 camera. The state of our model is more elliptic, due to that the real shape of the skin locus is camera dependent [14]. Since in the given situation, the collision face sizes range from approximately 65 * 65 to 100 *100 pixels, the model is scaled regarding a mean

face size of 80 * 80 pixels. We use M_0 as primary model.

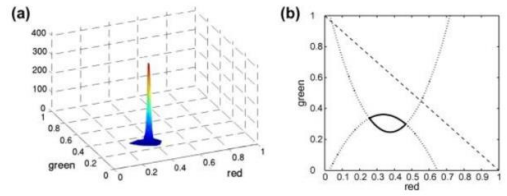

Fig. 4. (a) The skin colour distribution as determined from a training set. (b) The skin locus in normalized-rg colour space, described by two functions of quadratic order

During skin segmentation, we will utilize the skin locus to extent model adaptation. To do this, we depict its outline with two quadratic function $f_{\min}$ and $f_{\max}$ that we fit to the frontier point of the skin distribution, i.e., to the outer most histogram bins with non-zero count. The outcome can be found in Fig. 4b.

A certain shading (r,g) is a piece of the locus if

$$g > f_{\min}(\mathrm{r}) \wedge f_{\max}(\mathrm{r})$$

whith

$$f_{\min}(\mathrm{r}) = 6.38r^2 - 4.79r + 1.15$$

$$f_{\max}(\mathrm{r}) = -3.51r^2 + 2.53r - 0.06$$

Image segmentation:

The segmentation procedure is based on histogram backprojection, a strategy that highlights hues in the image which are a part of a histogram-based colour model [15]. For a solitary image, the probability of a pixel being skin given a colour vector (r,g) can be effortlessly inferred by use of Bayes' rule as described in point of interest in [16]. The outcome is a ratio histogram R, figured from the skin model histogram S and the image histogram *I*

$$p(\mathrm{skin} \mid \mathrm{r},\mathrm{g}) = \mathrm{R}(\mathrm{r},\mathrm{g}) = \frac{\mathrm{S}(\mathrm{r},\mathrm{g})}{I(\mathrm{r},\mathrm{g})}$$

where r and g mean the histogram bins. Next, R is backprojected onto the first(original) image, which denote, that every pixel i(x,y) is replaced by $R(\mathrm{r}_{x,y},\mathrm{g}_{x,y})$, where $r_{x,y}$ and $g_{x,y}$ mean the normalized colour estimations of i(x,y). In the other words, R is utilized as lookup table between pixel colour and skin probability. This outcomes in a gray scale image which can be interpreted as a probability guide of skin presence. As explained in [16] , utilization of Bayes' principle is just right, if applied to the same image from which the histograms were initially processed. But, practically speaking, this research works reasonably well for different(other) images taken in a similar situation. Backprojecting the ratio histogram instead of the model histogram itself emphasizes hues that are characteristic for the model. Thus, hues which are a part of the model however which are additionally in the background are weakened.

In [17], it is presented that the background stays noisy in cluttered situations however this issue is successfully addressed with a two-stage thresholding algorithm based on region growing. The main stage is a fundamental binary threshold at level T_{high}, which is set to 100. The next one is a hysteresis threshold like the one presented by Canny [18] for edge detection. It utilizes a lower threshold value T_{low} than the beginning one but it just adds those pixels to the beforehand made binary image which are 8-neighborhood connected to chosen pixels. The thresholded image is less cluttered, if the backprojection is smoothed utilizing a Gaussian kernel because this mitigates interlacing impacts and noise. Morphological administrators have been crossed off for rate reasons. possible face candidates are extracted from the thresholded image utilizing an associated components algorithm [19].

The lower threshold T_{low} is determined adaptively. It is picked as the mean gray level of the non-black pixels of the back projection, i.e., as the mean probability of all skin-like hue pixels. This method has a main point over a constant value of T_{low}. If the skin of an entering passenger is just poorly represented by the present model, because of hue, size or both, just a little rate of the skin pixels will be bigger than T_{high} while the greater part will have comparatively little values. If a constant T_{low} is selected too large, these pixels won't be segmented. selecting T_{low} correctly to successfully segment the badly modelled skin pixels, issues emerge when a well modelled face is encontered. The skin pixels of such a face will, to a large scope, get high probabilities of being skin. As a result, utilization of H_{high} already leading to rational segmentation. The little value of T_{low} from before will then add dispensable clutter to the segmented image.

Model adaptation:

The model produced from the skin samples, M_0, is utilized for first detection and is then adjusted to the present illumination and the object's specific skin colour. While a face of object(passenger) is successfully detected in a skin coloured region, the histogram H_{face} of this region is utilized to update the present model M_t.

$$M_{t+1}(\mathrm{r},\mathrm{g}) = (1-\alpha)\,\mathrm{M}_t(\mathrm{r},\mathrm{g}) + \alpha H_{face}(\mathrm{r},\mathrm{g})$$

with recalculate parameter α and bin indexes r and g. With α=0.4, this guarantees quick adjustment to each specific case. Because of the Gaussian

smoothing, the thresholding procedure depicted above leads to segmentation of non-skin pixels near to skin-coloured ones, e.g., eyes, lips and hair. Because of adjustment to these hues, just hues inside the skin locus are utilized to compute H_{face} .

Feature-based face and eye detection:

In order to detect the passenger faces and the eyes we have utilized the method proposed by Viola and Jones 20]. We utilize the execution of their algorithm from the Open Computer Vision Library (OpenCV) 21]. We designed our own face and eye detection cascades. we rotate training face images up to 45°.

Region of interest tracking:

A passenger's face is not need detected in every frame because passengers may turn sideways or look down so that the face detector, which has been trained for quasi frontal faces fails. Because of to be able to record the all sequence until the passenger leaves the camera's zone of view, a simple yet effective tracking algorithm has been utilized. Based on the fact that a profile view of a face still produces a object face during the skin segmentation step. This leads to the basic assumption that a face applicant at or near to a position where a face was successfully detected in past frames is similar to be this face. Essentially, this passenger face is selected as a face if its centre lies within the bounding box of the past detected face. To account for movement, the search area is enlarged by a certain amount. The processing of the following frame will then be confined to this region which leads to an enormous speedup as image information is diminished to a fraction.

Face registration:

Because of stable eye detections are crucial, eye areas are tracked over consecutive frames utilizing Kalman filter. Both eyes are tracked independently. The state of each of the two Kalman filter covers the x-and y-position of one of the eyes, together with its velocity of motion, v_x and v_y . The state estimates are supported by estimations of the (x,y) region of the eyes as determined by eye detectors.

The issue that emerges with eye detection is, that an eye tracer with a reasonable detection rate produces quite a couple of false positives. This is because of the fact that the intensity distribution of an eye, as caught by the classifier, is sort of simple. It can be seen in different parts of the processed range as well, e.g., on wavy hair. This is particularly valid since the tracer is trained with input information which is rotated up to 45°. Because of initialize the Kalman filters, it is important to decide on the "true" detection among every accessible ones. It is observed that the majority of false positives just appear in single frames or sets of frames. nonetheless, some of them are recognized more consistently.

To solve this issue, the method depicted in Fig. 5 is executed [13]. The detections of every eye cascade are utilized to create track hypotheses over consecutive frames. Close detections in consecutive frames are related to each other to form a track. Tracks that don't get redesigned with another estimation are extrapolated based on past observations. If few detections are associated with one track, it gets separate into two. If two tracks overlap for a few frames, one of them is rejected.

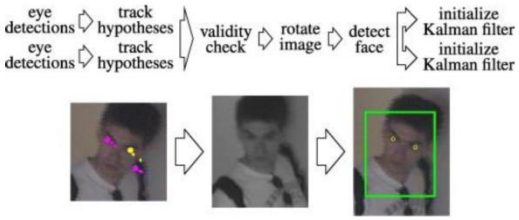

Fig. 5. Initialization of the Kalman filters for eye tracking.

From the arrangement of tracks, eye pairs are created with the down constraints

- Left eye is left of right eye.
- Eye distance is greater than a min.
- Left and right eye move into a similar direction.
- Left and right eye move at similar velocity.

The number of possible eye passengers is already greatly decreased at this point. To check the eye pair hypotheses, the image is initially rotated, so that the eye positions are on a horizontal line. Next, a face tracer is utilized to last confirm or reject the hypothesis. The rotation is important because of the face tracer is limited to upright faces. Without that limitation, the false positive rate would robustly increase as in the eye tracer case. If the face tracer is successful, the Kalman filters are initialized accordingly. Because of a fallback solution, eye candidates trigger the Kalman filter initialization if they show up consistently over a long time. From one viewpoint, this is important because of that the face tracer may still fail on an upright face. On the other point, it is possible because normally just the genuine eye areas are consistently detected over a more extended duration of time. The face tracer method has the capacity succeed within three frames while the fallback solution is triggered after successful detection of a valid eye pair more than 25 frames. Regardless of the fact that the eye locator is trained to account for some amount of rotation, it works best on horizontal eyes, i.e., upright faces. The detection results can be enormously improved if subsequent face applicants are rotated based on Kalman filter prediction prior to any detection or confirmation. If eye detection unsuccessful nevertheless, the prediction can be utilized as substitute.

Face will registered such that, the face image is rotated to bring the detected or predicted eye areas into horizontal position. After that, the image is scaled and cropped to a size of 75×75 pixels, so that the eyes are situated at certain coordinates in the resulting image. Fig. 6 demonstrates a few examples acquired with this strategy.

Fig. 6. Sample registered face images with the proposed system

Face recognition:

A local appearance based face recognition algorithm is utilized [21] and [22] for face recognition. Generally it is a face recognition method that has been designed to be robust against occlusion variations as well as real life situation ,expression and illumination. This algorithm has been evaluated on a few benchmark face databases, for example, AR [23], CMU PIE [24], FRGC [25], Yale B [26] and Extended Yale B [27] face databases, and found to be significantly better than other previous face recognition algorithms, for example, eigenfaces [28], Fisherfaces [29], embedded hidden Markov models [30] and Bayesian face recognition [31].

The method uses representation of local facial area and combines them at the element level, which gives conservation of the spatial relationships. This algorithm utilizes discrete cosine transform (DCT) for local appearance representation. There are a few merit of utilizing the DCT. Its information independent bases make it exceptionally useful to utilize. There is no need to set up a representative set of training information to process a subspace. Moreover, it gives frequency data, which is extremely helpful for handling changes in facial appearance. For example, it is realized that some frequency groups are useful for combating against illumination variation. In addition, we have found that the DCT-based local appearance representation is more useful than representations based on the Karhunen-Loève, Fourier, Wavelet and Walsh–Hadamard transforms in terms of face recognition execution [14]. In the proposed method, a registered and detected face image is divided into non overlapping windows of 8×8 pixels size. The explanation behind picking a window size of 8×8 pixels is to have sufficiently little windows in which stationary is given and transform complexity is kept simple on one hand, and to have sufficient windows to give adequate compression on the other hand. Moreover, the experiments conducted with different window sizes also demonstrated that utilizing a window size of 8×8 pixels is also useful for face recognition execution. After that, on each 8×8 pixels window, the DCT is performed. The got DCT coefficients are ordered utilizing zig–zag scanning. From the ordered coefficients, the initial five AC coefficients are chosen in order to create compact local feature vectors. The DC coefficient is rejected for illumination normalization as recommended in [32]. Moreover, robustness against illumination varieties is expanded by normalizing the local feature vectors to unit standard[33]. This reduces illumination impacts, particularly illumination variety with an gradient pattern, while keeping the fundamental frequency data. At the end, the local feature vectors extracted from every window are concatenated to develop the general feature vector. Both a discriminative and a generative methods are followed to classify the so-achieved feature vectors. With both methods, individual models are determined for each passenger. The granularity of these models related on the respective amount of accessible training information. The window diagram of the face recognition framework is given in fig 7.

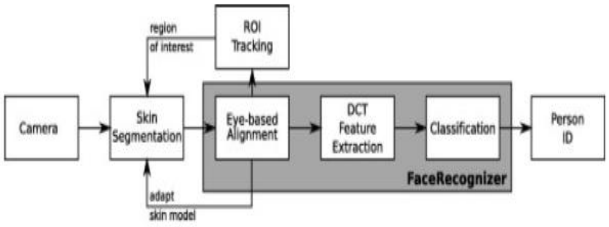

Fig. 7. Overview of the face recognition system.

K-nearest neighbours model:

The mean merit of discriminative methods like K-nearest neighbours (KNN) is that they don't make a assumption about the distribution of the fundamental information. This permits to design meaningful models with less information than would be important to training high-dimensional generative models like Gaussian mixtures. To select the nearest neighbours, the L_1 norm is utilized as distance measure $d(.,.)$, as it was demonstrated to perform best among a few popular distance measurements in [12]. The k nearest neighbours S_i, i=1,2,… ,k of a test vector x are chosen with score $s_i = d(\mathrm{x}, \mathrm{S}_i)$. Since the distance and, thus, the resulting scores can differ largely between frames, they should be normalized. It is got with straight min-max standardization [33].

$$S_i' = 1 - \frac{S_i - S_{\min}}{S_{\max} - S_{\min}}$$
$$i = 1, 2, ..., k$$

which maps the scores to [0, 1]. To have equal contribution of every frame, these scores are re-

normalized to $\sum_{i=1}^{k} s_i' = 1$. Among the k nearest representatives, there can be a few ones from the same class. Since a few passengers have far fewer representatives than others, care must be taken that their scores are not dominated by those. Individual scores are chosen by a simple max-rule [34], which just chooses the max score for every class. A sum-rule [25] decision fusion scheme is utilized to take merit of all frames in a video sequence to decide the identity of a object. Two baseline performances are selected. The first one is, each and every frame is evaluated independently to have the capacity to evaluate the change contributed by video based classification. Second one is, the standard video based recognition execution is determined by simply adding the scores of all frames.

Gaussian mixture model:

Although generative models, for our situation Gaussian mixture models (GMM), generally require more training information(data) than discriminative ones, they permit to model the information with probability density function (pdf), and, as a consequence, the computation of conditional pdfs. The Gaussian mixture model method trains one GMM per class utilizing a expectation maximization algorithm [35].

In like manner the KNN model, the number of components per mixture related on the number of training samples available for a object. At runtime, object x is classified as one of the N registered individuals in a max log-likelihood manner utilizing:

$$\text{argmax}_{i\in N} \log P(\mathrm{x} \mid \mathrm{i}) = \text{argmax}_{i\in N} \log \sum_{i=1}^{k_i} \alpha_{ij} . N(\mathrm{x}, \mu_{ij} . \sum_{ij})$$

where K_i denotes the number of modes per object, α_{ij} the mixing parameters, and μ_{ij} and $\sum_{ij}$ the mean and the variance of the j^{th} component of object i's model, respectively. To keep the computational effort within reasonable limits, just a diagonal rather than the full covariance matrix is utilized. Three methods are utilized to assess the classification performance of the GMM setup on video input. Like the KNN model, frame based evaluation determines the baseline performance of the model. Each frame is evaluated on its own based on a min-max normalization.

Bayesian inference:

Utilizing Bayes' rule, posterior probabilities are calculate for every class. These posteriors are utilized as in the next frame. The posterior probability $p(\mathrm{i}_t \mid \mathrm{x}_{0:t})$ of object i at casing t given the all the past observations $x_{0:t}$ is formulate as:

$$p(\mathrm{i}_t \mid \mathrm{x}_{0:t}) = \frac{p(\mathrm{x}_t \mid \mathrm{i}_t).\mathrm{p}(\mathrm{i}_t \mid \mathrm{x}_{0:t-1})}{p(\mathrm{x}_t)}$$

The conditional observation likelihood $p(\mathrm{x}_t \mid \mathrm{i}_t)$ is calculated by the GMM for object i, the unconditional one by

$$p(\mathrm{x}_t) = \sum_{i=1}^{N} P(\mathrm{x}_t \mid \mathrm{i}_t).\mathrm{p}(\mathrm{i}_t \mid \mathrm{x}_{0:t-1})$$

with N being the number of objects. The priors are initialize uniformly:

$$p(\mathrm{i} \mid \mathrm{x}_0) = \frac{1}{N}$$

This method takes into account the temporal dependency by calculating probability to observe a given sequence of input frames.

Bayesian inference with smoothing:

In view of the past approach, the thought of a consistent identity is presented as recommended in [14]. The identity of an entering passenger(object) does not change but depending on frame and model quality the classification of single frames can vary from past ones. As an outcome, the influence of frames which are not consistent with the present sequence hypothesis, i.e., the present classification for a given sequence, is reduced. Equation (6), the smoothed posteriors are computed as :

$$p(\mathrm{i}_t \mid \mathrm{x}_{0t}) = \frac{\mathrm{p}(\mathrm{x}_t \mid \mathrm{i}).\mathrm{p}(\mathrm{i}_t \mid \mathrm{i}_{t-1}) p(\mathrm{i}_t \mid \mathrm{x}_{0t-1})}{p(\mathrm{x}_t)}$$

with

$$p(\mathrm{i}_t \mid \mathrm{i}_{t-1}) = \begin{cases} 1-\varepsilon & \text{if } i_t = i_{t-1} \\ \frac{\varepsilon}{N} & \text{otherwise} \end{cases}$$

The measure of smoothing is dictated by the smoothing parameter ϵ, where smaller values denote stronger smoothing. With a value of 0, the sequence is fundamentally classified solely in view of the first frame. The values near to 0 lead to an fixing of the sequence hypothesis while as yet permitting a change to a distinct identity as the experiments in the following section

Frame weighting :

Due to the real world quality of the information, not all frames are suitable to classify the object. Low resolution, large occlusion and faulty alignment are samples of negative impacts on frame quality. Also, certain perspectives of a object might just not be captured by the model because of small training information or because of preparing information that contains too little variety. Two mean observations have been produced from the experiments conducted on a parameter estimation group, which are used in order to reduce the impact of ambiguous frames. To start with, for wrong classifications, the distance to the nearest representative is on average, more than for right ones and additionally, badly aligned frames result in larger distances as well. To this issue, we present the weighting technique distance to model (DTM). The frames f_i ,i=1,2,… , are weighted as for the nearest representative c with

$$W_{DTM}(\mathrm{f}_i) = \begin{cases} 1 & if \quad d(\mathrm{f}_i,\mathrm{c}) < \mu \\ e^{-\frac{d(fi,\mathrm{c})}{2\sigma^2}} & otherwise \end{cases}$$

This weighting function is selected by observed distribution of all frames distances $d(\mathrm{f}_i, \mathrm{c}_{f_{i,corect}})$, the distances of all frames f_i to the nearest representative $f_{i,correct}$ of the corresponding correct class. The distribution, decided on a parameter estimation set, resembles a normal distribution $N(.,\mu,\sigma^2)$. To get more robustness against outlier, μ is selected as sample median and σ^2 as median absolute deviation (MAD) [35]. A example distribution and weight function is demonstrate in Fig. 8. Utilizing the weight function W_{DTM}, the influence of frames which are not adequately near to the model is reduced.

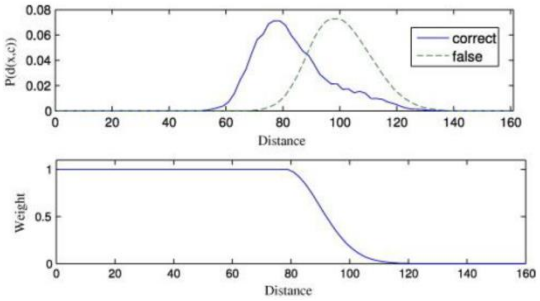

Fig. 8. Distance to model weight function. (top) Distribution of the distances to the closest representative of the correct class (blue, solid) and to all other classes (green, dashed) and (bottom) the actual weight function.

The next observation is that, in case of misclassification of frame f_i, the differences of the distances Δ(f_i) to the nearest and the second nearest representation is generally smaller than in the right case. The distribution of these distances follows approximately an exponential distribution:

$$\varepsilon(\mathrm{x};\lambda) = 0.1\lambda e^{-\lambda x}$$

with

$$\lambda = 0.5$$

The weights are then processed as the cumulative distribution function of ε(·)

$$W_{DT2ND}(\mathrm{f}_i) = E(\Delta(\mathrm{f}_i)) = 1 - e^{-i\Delta(\mathrm{f}_i)}$$

An example distribution and weight function is demonstrated in Fig. 9. This weighting scheme will be alluded to as distance-to-second-nearest (DTSN).

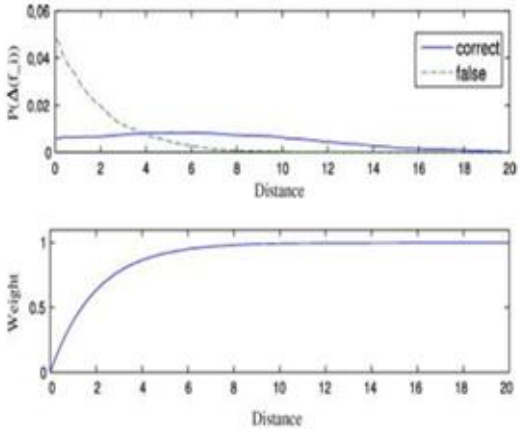

Fig. 9. DTSN weight function. (top) Distribution of the distances between the closest and second closest representatives for correct (blue, solid) and false classifications (green, dashed) and (bottom) the actual weight function.

Distance To Model and Distance To Second nearest use diverse kind of data. DTM considers how comparable a test sample is to the representative of the training set, though DTSN considers how well the nearest and second nearest representatives are separated. Foe example, a badly aligned face image causes an large distance to the model. On the other hand, the best matches can be still well separated. It is imaginable to have both conditions satisfied. That is, having a small distance to the nearest representative, and a well separation between the nearest and second nearest representatives. By virtue of this reason, in addition to individual weighting schemes, a joint weighting scheme is utilized that utilizes the product of W_{DTM} and W_{DT2ND} to weight the frames.

Experiments: In this section the assessment after effects of the video segmentation also, face recognition frameworks are exhibited and examined.

Assessment of the video segmentation: To evaluate the execution of the video segmentation calculation, four persistent video streams were recorded on three distinct days and physically named for ground truth. They cover a time period of 24 h and comprise of pretty nearly 2.2 million frames. Table 1 gives a point by point review. According to the share of under one percent of relevant information inside of the recorded video, it is evident that continuous recording isn't an option for sensible information collection, not just concerning memory prerequisites however especially in terms of effort and time utilization of repetitive manual segmentation. The outcomes in Fig. 10 are given as Correct Detection Rate and False Detection Rate. A correct detection is given if a detected sequence overlaps no less than half of a labelled one. The aggregate correct detection rate for diverse overlaps values demonstrate in Fig. 11.

Sequence	Duration	Tolal no. Of frames	No. Of Sequences	No. Of relevant frames
A	02:53:16	259,910	42	2929
B	04:04:13	366,318	12	3233
C	03:25:25	308,124	12	989
D	06:07:38	551,443	63	6220
Total	16:30:32	1,484,795	129	13371

Table 1:The number of sequences refers to situations in which subject(traveller/costumer) is actually entering the security gate.

Sequence	CDR	FDR
A	92.9	9.3
B	83.3	0.0
C	100.0	0.0
D	95.2	9.1
Total	93.8	7.6

Fig.10:Performance of face detection

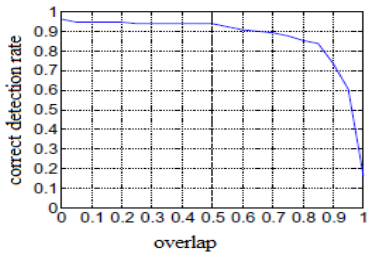

Fig.11:Detection performance of face recorder with respect to overlap ratio

Assessment of the face recognition system: To assessment the face recognition framework, we utilized a database that consists of 6960 video sequences (626463 frames) of 125 objects recorded during one mount. This database is chronologically divided into three sets for training, parameter estimation and testing as listed in Table 2. Face images are automatically extracted from training sequences utilizing the registration process specified in Section 3. Training data is completed with virtual samples. Those are produced during the extraction process by artificial perturbations of the main detected eye regions by ±2 pixels in x- and y-direction. The face is then adjusted by new coordinates and saved in the training data set. Since nine areas per eye are assessment, this increases the training set size by factor 81. The range number of raw training samples is vast which would slow down the KNN method. Since numerous samples from consecutive frames are fundamentally the same, k-means is applied to choose representative models. The clustering is performed for each passenger individually.

Size of the three subset	Number of sequences
Training set	2795
Parameter set	1105
Test set	3060
Total	6960

Table 2:Sizes of three subsets

Closed set identification:

For close set identification, the framework is just confronted with subjects that are recorded in the database. The framework needs to classified every subject as one of the possible classes. The efficiency is measured as Correct Classification Rate (CCR), the rate of aright classified sequence in the test set. Each single frame is assessed separately, that is, correct classification rate is computed as the rate of aright classified frames between all frames. The outcomes are given in Table 3. Uniform indicates no weighting and hence equal share of every frames, combined the combination of Distance To Model and Distance To Second Closest by weight multiplication. Video based assessment better than frame based assessment since the increased amount of accessible data helps to resolve a few ambiguities. Clearly, both weighting scheme enhance the classification resolute over uniform weighting. The combination takes merit of both and fulfill better. The increase is somewhat bigger for Distance To Model, as it allocate smaller weights to frames that are not similar enough to the representative of the training set. Distance To Second, conversely, diminishes the effect of ambiguous frames, i.e., frames which yield same scores for the top two objects, independent of how well the face is modelled. Actually badly aligned image can lead to a distinct score, however it is maybe to have a large distance to the model. Distance To Model has the capacity handle this case, Distance To second is not. Nonetheless, decline of obscurity lead to better result over uniform weighting too.

Since the various models are concerned, the discriminative methods perform better than the generative ones. Since parametric models like GMMs need more training data with expanding dimensionality, this is maybe caused by insufficient training data for some objects which can avoid derivation of meaningful models. In addition, the number of mixture components may not be adequate to calculate the underlying probability distribution dissemination. The discriminative models are less influenced by little training data, as they classify new information just based on existing information, without making any suppositions about its distribution. To investigate the more better results, it valuable looking at the outcomes including rank-2 and rank-3 classification i.e., cases in which the right identity is among the best a few hypotheses. As clearly demonstrated in Fig. 12 and Table 4, the frame based method regularly get near to the right decision. Although, it has to decide on the identity even in the case that the single element vector is of suspicious quality. The method does not have an open door to support or discard the hypothesis utilizing extra data as done by the sequence based approach. These have the capacity to exploit the temporal dependency of consecutive frames and to advance the rank-2 and rank-3 classifications of the

frame based approaches to first place. Since numerous frames contribute to the choice, the overall betterment change is bigger than the difference between the right and rank-3 classification in the frame based method. The more frames can be assessed, the more probable it is to acquire a right result. This gets affirmed by the observation that the mean length of correctly classified sequences is bigger – 39 frames – than that of misclassified ones with 28 frames as demonstrate in Fig.13.

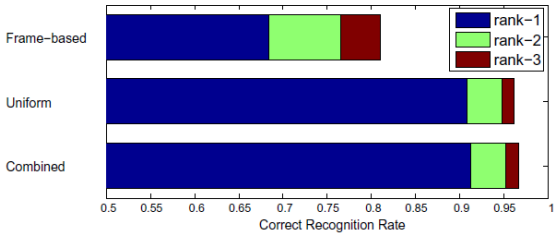

Fig 12:correct recognition rate by rank for the KNN models.

To justify the growth training efforts caused by the bigger training set size, an experiment was led to compare the recognition performance utilizing completed and uncompleted training data. The comparison can just cover the KNN approach as it is impossible to training suitable GMMs because of the fact that numerous objects have fewer images in the training set than the component vector's dimensionality. As recorded in Table 5, recognition efficiency increments significantly in each of the three KNN cases. This demonstrates that the information growth is certainly justified regardless of the expanded memory and time resources. Adding noise to recognized eye areas leads to samples of varying scale and rotation which increment the variety bans width and reduces the influence of conceivable record error. For as much as the data set size is expanded by factor 81, even persons with couple of real training image can be modelled properly.

KNN	CCR	GMM	CCR
Frame based	68.4	Frame based	62.7
Uniform	90.0	Uniform	86.7
DTM	92.0	Smooth	87.8
DTSN	91.3	DTM	90.6
Combined	92.5	DTSN	89.1
		Combined	91.8

Table 3:closed set correct classification results with different schemes. smooth GMM used $\varepsilon = 10^{-5}$

	Frame base	Uniform	Combined
Rank-1	68.7	90.9	92.5
Rank-2	76.5	94.8	95.6
Rank-3	81.1	96.2	96.7

Table 4:correct recognition rate by rank for KNN models

	Frame base	Uniform	Combined
Unaugmented	56.6	87.6	88.2
Augmented	68.4	90.9	92.5
p-value	0.00	0.02	0.00
Significantly better	√	√	√

Table 5:Influence of data set augmentation.

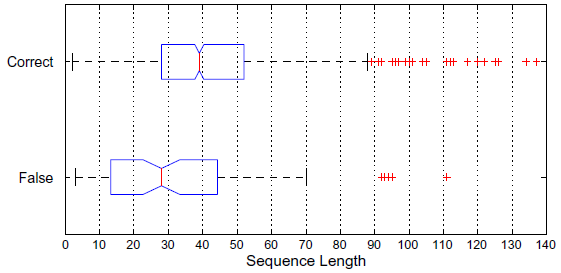

Fig.13:Box plot showing the distribution of sequence lengths for foe correct and false classification.

Conclusion:

In this paper, we presented a real time video based face recognition framework. This system is developed for security zones like airports ,hotels and smart stores. This system able to process data under every time conditions and to accomplish this in real time. Violations of this requirement necessarily interrupt the subject(passenger/customer) to be identified or restrict them in their actions. It has been introduced three weighting schemes to weight the contribution of individual frames in order to get better classification performance. The introduced face recognition framework required on average 27ms per frame on Pentium 4 with 3 GHz and 2 GB ram.

In the future we are planning to use kinect camera technology to design system to utilize in very crowded area like stadiums to recognize human action, people tracking and region of interest processing in occluded zone with high accuracy.

References:

1. Gross, R., et al., *Face recognition across pose and illumination*, in *Handbook of Face Recognition*. 2005, Springer. p. 193-216.
2. Zana, Y., et al., *Local approach for face verification in polar frequency domain.* Image and Vision Computing, 2006. **24**(8): p. 904-913.
3. Ekenel, H.K. and R. Stiefelhagen. *Face alignment by minimizing the closest classification distance*. in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. 2009. IEEE.
4. Park, U., Y. Tong, and A.K. Jain, *Age-invariant face recognition.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(5): p. 947-954.
5. Bazakos, M.E., Y. Ma, and A.H. Johnson. *Fast access control technology solutions (FACTS)*. in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. 2005. IEEE.
6. Akhloufi, M., A. Bendada, and J.-C. Batsale, *State of the art in infrared face recognition.* Quantitative InfraRed Thermography Journal, 2008. **5**(1): p. 3-26.
7. Wiehl, T., *Human and Computerized Facial Recognition: Comparison and Constitutional Analysis.* Nw. Interdisc. L. Rev., 2013. **6**: p. 95.
8. Lim, F.-L., W. Leoputra, and T. Tan, *Non-overlapping distributed tracking system utilizing particle filter.* The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, 2007. **49**(3): p. 343-362.
9. Srirama, S.N., C. Paniagua, and H. Flores, *Croudstag: Social group formation with facial recognition and mobile cloud services.* Procedia Computer Science, 2011. **5**: p. 633-640.
10. Akgül, C.B., et al., *Content-based image retrieval in radiology: current status and future directions.* Journal of Digital Imaging, 2011. **24**(2): p. 208-222.
11. Turk, M. and A.P. Pentland. *Face recognition using eigenfaces*. in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. 1991. IEEE.

selection and feature normalization. in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*. 2006. IEEE.

12. Gao, H., H.K. Ekenel, and R. Stiefelhagen, *Pose normalization for local appearance-based face recognition*, in *Advances in Biometrics*. 2009, Springer. p. 32-41.
13. Störring, M., H.J. Andersen, and E. Granum, *Physics-based modelling of human skin colour under mixed illuminants.* Robotics and Autonomous Systems, 2001. **35**(3): p. 131-142.
14. Martinkauppi, J.B., M.N. Soriano, and M.V. Laaksonen. *Behavior of skin colour under varying illumination seen by different cameras at different colour spaces*. in *Photonics West 2001-Electronic Imaging*. 2001. International Society for Optics and Photonics.
15. Swain, M.J. and D.H. Ballard, *Colour indexing.* International journal of computer vision, 1991. **7**(1): p. 11-32.
16. Adam, A., E. Rivlin, and I. Shimshoni. *Robust fragments-based tracking using the integral histogram*. in *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*. 2006. IEEE.
17. Demirel, H. and G. Anbarjafari, *Pose invariant face recognition using probability distribution functions in different colour channels.* Signal Processing Letters, IEEE, 2008. **15**: p. 537-540.
18. Perona, P. and J. Malik, *Scale-space and edge detection using anisotropic diffusion.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1990. **12**(7): p. 629-639.
19. Rosenfeld, A. and J.L. Pfaltz, *Sequential operations in digital image processing.* Journal of the ACM (JACM), 1966. **13**(4): p. 471-494.
20. Viola, P. and M.J. Jones, *Robust real-time face detection.* International journal of computer vision, 2004. **57**(2): p. 137-154.
21. Bradski, G. and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. 2008: " O'Reilly Media, Inc.".
22. Fortmann, T.E., Y. Bar-Shalom, and M. Scheffe, *Sonar tracking of multiple targets using joint probabilistic data association.* Oceanic Engineering, IEEE Journal of, 1983. **8**(3): p. 173-184.
23. Martinez, A. and R. Benavente, *Cvc technical report# 24.* The AR Face Database, 1998.
24. Sim, T., S. Baker, and M. Bsat. *The CMU pose, illumination, and expression (PIE) database*. in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. 2002. IEEE.
25. Phillips, P.J., et al. *Overview of the face recognition grand challenge*. in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*. 2005. IEEE.
26. Georghiades, A.S., P.N. Belhumeur, and D.J. Kriegman, *From few to many: Illumination cone models for face recognition under variable lighting and pose.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001. **23**(6): p. 643-660.
27. Lee, K.-C., J. Ho, and D. Kriegman. *Nine points of light: Acquiring subspaces for face recognition under variable lighting*. in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001. IEEE.

28. Turk, M. and A. Pentland, *Eigenfaces for recognition.* Journal of cognitive neuroscience, 1991. **3**(1): p. 71-86.
29. Belhumeur, P.N., J.P. Hespanha, and D.J. Kriegman, *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(7): p. 711-720.
30. Nefian, A. and M.H. Hayes III, *A hidden Markov model-based approach for face detection and recognition*, 1999, School of Electrical and Computer Engineering, Georgia Institute of Technology.
31. Moghaddam, B., T. Jebara, and A. Pentland, *Bayesian face recognition.* Pattern Recognition, 2000. **33**(11): p. 1771-1782.
32. Ekenel, H.K. and R. Stiefelhagen. *Local appearance based face recognition using discrete cosine transform*. in *13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey*. 2005.
33. Snelick, R., et al., *Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005. **27**(3): p. 450-455.
34. Kittler, J., et al., *On combining classifiers.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1998. **20**(3): p. 226-239.
35. Huber, P.J., *Robust statistics*. 2011: Springer.