The Effect of Training Data Selection on Face Recognition in Surveillance Application

Jamal Ahmad DARGHAM[a]; Ali CHEKIMA[a]; Ervin Gubin MOUNG[a]; Segiru OMATU[b]

jamalad@ums.edu.my,chekima@ums.edu.my,menirva.com@gmail.com,omatu@rsh.oit.ac.jp

[a] FakultiKejuruteraan, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.

[b]Faculty of Engineering, Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, 5-16-1, Omiya, Asahi-ku, Osaka, 535-8585, Japan.

KEYWORD

ABSTRACT

Face recognition is an important biometric method because of its potential applications in many fields, such as access control and surveillance. In surveillance applications, the distance between the subject and the camera is changing. Thus, in this paper, the effect of the distance between the subject and the camera, distance class, the effect of the number of images per class, and also the effect of session used to acquire the images have been investigated. Three sessions are used to acquire the images in the database. The images in each session were equally divided into three distance classes: CLOSE, MEDIUM, and FAR, according to the distance of the subject from the camera. It was found that using images from the MEDIUM class for training gives better performance than using either the FAR or the CLOSE class. In addition, it was also found that using one image from each class for training gives the same recognition performance as using three images from the MEDIUM class for training. It was also found that as the number of images per class increases, the recognition performance also increases. Lastly, it was found that by using one image per class from all the available database sessions gives the best recognition performance.

1. Introduction

Face detection and face recognition are growing biometric research fields because of their potential applications as important tools for many applications such as security surveillance and human-computer interaction. Many face detection and face recognition techniques have been proposed with satisfactory success (Zhaoet al, 2003).However, face recognition for surveillance is difficult to tackle because in closed circuit television (CCTV) the cameras photograph people at tilted angles or in low light, thus resulting in poor quality video images being captured(Wiliemet al, 2007). The low quality video coupled with the large variations in the subjects face orientations in the acquired video decreases the recognition accuracy. Various techniques have been proposed for face recognition for surveillance application (Jillela and Ross, 2009, Kamgar, 2011, Davis et al, 2010). These can be classified into two groups: software-based and hardware-based techniques. Vera-Rodriguez et al. (2010) investigated the effect of the distance between the camera and the subject on recognition rate. They categorized the images in the database according to the distance between the camera and the subject into three classes, namely; close, medium, and far class. They used mug-shot frontal face for their training library and surveillance videos for their testing. Their work shows that Equal Error Rate

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

58

increases (recognition performance dropped) when the person is further away from the camera regardless of the face recognition technique used. Similarly, YuxiPeng (2011) used mug-shot frontal face for training data and video camera images for testing, the author's work also shows that as the distance between a person and the camera increases, the recognition rate decreases regardless of the face recognition technique used. Yoo et al. (2007) reported that the YC_BC_R and $YC_G'C_R'$ colour spaces improve the performance of face recognition. In the previous works, the training and testing data are different. Thus, we propose to use the same type of images, surveillance video, for both training and testing. However, this raises the question, given that the distance between the subject and the camera is changing, the images from which distance class (FAR, MEDIUM, and CLOSE) should be used for training to produce the lowest error rate? Thus, in this work, an investigation of the selection of training data on the performance of face recognition in surveillance application will carried out. In particular, the effect of images classes on face recognition, the effect of the number of image frames per class, and the effect of database session used on face recognition will be investigated.

2. Surveillance Dataset

The ChokePoint Dataset (National ICT Australia Limited, 2014) is a collection of surveillance videos of 25 persons in portal 1 and 29 persons in portal 2. Portal 2 dataset was used for database preparation since it has more subjects available compared to portal 1. Portal 2 dataset consists of two sub datasets; Portal 2 Entering scene dataset (P2E) and Portal 2 Leaving scene dataset (P2L). P2L dataset contain better quality frontal faces. Thus, P2L dataset were selected for database preparation for all experiments. Although the video images were captured in four sessions, only session 1 (S1), session 2 (S2), and session 3 (S3) were used. Session 4 video images were not used because the captured image of the face is not frontal as seen in Figure 1. All the frontal face images are aligned so that their eyes are in the same horizontal line.

Figure 1 Samples of frontal face images in the P2L dataset, Source ChokePoint Dataset

3. Database Preparation

The ChokePoint database is a series of video frames of a person walking towards the camera. Each person has different number of video images captured since some people walk faster than others thus having lesser video images captured by the camera. The smallest number of video frames per person is 44 frames. The video frames were then divided into three classes, namely; FAR, MEDIUM, and CLOSE class. Since the lowest number of frames available is 44, 42 frames are selected so that each class has the same number of frames, which is 14 frames per class. From the 1st frame to the last 42nd frame, each video image is showing the person getting closer to the camera location. The definitions of the classes are:

- FAR class - The first 14 frames from the 42 frames per person. The distance between the person and the camera is longer than MEDIUM and CLOSE class.
- MEDIUM class - The next 14 frames after the FAR class frames from the 42 frames per person. The distance between the person and the camera is shorter than FAR class and longer than CLOSE class.
- CLOSE class - The last 14 frames from the 42 frames per person. The distance between the person and the camera is shorter than FAR and MEDIUM class.

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

59

A total of 3528 video images with frontal face of persons were selected from the ChokePoint database. Three databases representing three different session containing 1176 video images each were created. They represent 28 different persons. 14 persons are used for training and testing, and the other 14 different persons are used for testing only. All the video images cover from chin to crown of the person and both eyes of the person are aligned in the same horizontal line. The image size is 96 by 96 pixels and all the images are in grey scale format.

3.1. Training Database

Eight cases of selecting training images were prepared as shown in Table 1. As shown in Table 1, these caseswill be used to evaluate the effect of training data selection on the performance of face recognition in surveillance applications.

Table 1 The proposed training database criteria

Case	Selection criteria
Case 1	Three images per person randomly selected from the same class and from the same database session will be used
Case 2	Three images per person. One image randomly selected from each of the three class of the same database session will be used
Case 3	Nine images per person.Three images randomly selected from each of the three class of the same database session will be used
Case 4	Nine images per person.One image randomly selected from each of the three class of the three database sessions will be used
Case 5	Three images per person; three mean images from each class of the same database session will be used.
Case 6	Nine images per person from all the database session with 1 mean image per class will be used.
Case 7	Six images per person; consist of three randomly selected images and three mean images from each class of the same database session will be used. (Case 7 is Case 2 and Case 5 images combined together in the same folder)
Case 8	18 images per person; consist of nine images and nine mean images from all the database session with 1 randomly selected image and 1 mean image per class will be used. (Case 8 is Case 4 and Case 6 images combined together in the same folder)

Case 1 and 2 are used to evaluate the effect of class on face recognition. While case 2 and 3 are used to study the effect of the number of images per class on face recognition. The effect of database used will be evaluated using case 3 and 4. Case 2 and 5 are used to study the effect of using one mean image from each class from the same database while case 4 and 6 are used to study the effect of using one mean image from each class from all the database sessions on face recognition. Lastly, case 5 and case 7 are used to study the effect of using one image and one mean image from each class from the same database session while case 6 and case 8 are used to study the effect of using one image and one mean image from each class from all the database sessions on face recognition.

Table 2 Training database details for case 1, 2, 3, 4, 5, 6, 7, and 8.

Case	Training database size	Number of persons in training database	Number of images per person	Number of class per person	Number of images per class
1	42 images	14 persons	3 images	1 class	3 images
2	42 images	14 persons	3 images	3 classes	1 image
3	126 images	14 persons	9 images	3 classes	3 images
4	126 images	14 persons	9 images	3 classes	3 images
5	42 images	14 persons	3 images	3 classes	1 image
6	126 images	14 persons	9 images	3 classes	1 image
7	84 images	14 persons	6 images	3 classes	2 images

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

60

| 8 | 252 images | 14 persons | 18 images | 3 classes | 2 images |

3.2. Testing Database

Two testing databases were created. The first database, Client test database, has 588 images of the 14 persons for each session. This database will be used to test the recall capability of the face recognition system. The second database, Imposter test database, also has 588 images of 14 different persons for each session. This database will be used to test the rejection capability of the system.

4. The Face Recognition System

The face recognition system used in this work is the Principal Component Analysis (PCA) as shown inFigure 2.Itconsists of an offline training phase and an online testing phase. During the training phase, the set of training images are projected, using PCA, to a lower dimensional feature space. This projection will produce sets of feature vectorshaving much smaller dimensions compared to the original image dimensions. The feature vectors are then stored in the training database. During the online recognition process, a test image is projected to a lower dimensional feature space using PCA. The feature vectors representing the test image are then compared with the feature vectors from the training database in the matching process.
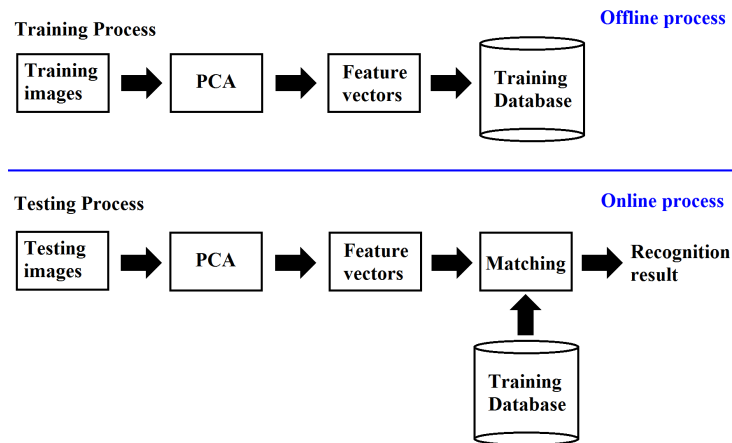

Figure 2 Block diagram of the PCA based face recognition system.

5. The Training Phase

For the training and testing phase, the euclidean distance is used to calculate the distance between two vector. In order to do the matching task during the testing phase, a formula generating a threshold value using the training database, denoted as, t, is propose. Let f be the feature vector from the training database, then, the threshold value, t, is the maximum euclidean distance between any two feature vectors from the training database, divided by a tunable parameter denoted as *Tcpara* as shown in equation (1).

$$t = \frac{max\{\| f_M - f_N \|\}}{Tcpara} \tag{1}$$

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

61

where $M, N = 1$ to Z. M, N is the feature vector index in the training database and Z is the size of the training database. The threshold value, t, will be use during the testing phase where it will be compare with a euclidean distance value generated by the test person during the matching task.

6. The Testing Phase

Let $i = 1, 2, ... 42$ and $M = 1, 2, ... Z$ and Z is the size of the training database. Let $T_R = \{f_{R1}, f_{R2}, ..., f_{R42}\}$ be the test person that have 42 test frames, f_{Ri} is a feature vector that represents the test person frame, and f_M is a feature vector from the training database, the minimum euclidean distance between every test feature vector, f_{Ri}, with all the training database feature vector f_M, is calculated, denoted as $E = \{e_1, e_2, ... e_{42}\}$ where E is a set that contain all the 42 minimum euclidean distance and e_i is the minimum euclidean distance calculated between f_{Ri} and f_M.

The person P from the training database that has the minimum euclidean distance e_i with the test feature vector f_{Ri}, is denoted as l_i and the collection of l_i is $L = \{l_1, l_2, ... l_{42}\}$.

Let $P_{mode} = \text{mode}\{L\}$, be the highest occurrence from the set L, then the euclidean distance average, denoted as E_{av}, of all the e_i from E with l_i equal to P_{mode}, were calculated.

If the average Euclidean distance between test person T_R and P person in the training database, $E_{av}(T_R,P)$ is smaller than a given threshold t, then test person T_R and P person in the training database are assumed to be of the same person. To measure the performance of both systems, several performance metrics are used. These are:

For Recall
- **Correct Classification**. If a test person Ti is correctly matched to the same person Pi in the training database.
- **False Acceptance.** If a test person Ti is incorrectly matched with person Pj, where i and j are not the same person
- **False Rejection.** If a test person Ti is of a person Pi in the training database but rejected by the system.

For Reject
- **Correct Classification**. If a test person Ti, from the Imposter test database is rejected by the face recognition system.
- **False Acceptance**. If a test person Ti from the Imposter test database is accepted by the program.

For this work, the threshold tuning parameter, *Tcpara*, was set so that the system has equal correct classification rates for both recall and reject. This classification rate is defined as the **Equal Correct Rate (ECR)**.

7. Results and Discussion

In case 1 there are a total of 27 separate experiments. Table 3 shows the nine experiments setup for case 1 training database using the FAR class images. The experiment were then repeated using MEDIUM and CLOSE class images, thus resulting in a total of 27 experiments using case 1 training database criteria.

Table 3The nine experiments setup for case 1 using FAR class images.

CLASS	Training Database Session	Testing Database Session
	S1	S1
	S1	S2
	S1	S3
FAR class images were used for training database	S2	S1
	S2	S2
	S2	S3
	S3	S1

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

62

S3	S2
S3	S3

In order to do comparisons between case 1 and the other cases performances, an equation was proposed to represent the performance of case 1 criteria. As can be seen from Table 3, using a FAR class S1 training database tested on S1, S2, and S3 testing database will produce three equal correct rate respectivelyand the average of these three ECR was then calculated as shown in equation (2).

$$avg\_ECR\_FAR\_STNj = \sum_{i=1}^{3} \frac{ECR\_FAR\_STT_i}{3} \qquad (2)$$

Where:

$ECR\_FAR\_STT_i$ = The equal correct rate produced using FAR class Sj training database tested onSi testing database. i, j = 1, 2, 3

$avg\_ECR\_FAR\_STN_j$ = The average equal correct rate produced using FAR class Sj training database tested on S1, S2, and S3 testing databases

The same equation (2) was applied on MEDIUM and CLOSE class training database. In order to determine the best class that give the highest average equal correct rate equation (3) is used.

$$[\text{Best Class, Best Score}] = max \begin{bmatrix} max\ (avg\_ECR\_FAR\_STN_1 & avg\_ECR\_FAR\_STN_2 & avg\_ECR\_FAR\_STN_3 &) \\ max\ (avg\_ECR\_MEDIUM\_STN_1 & avg\_ECR\_MEDIUM\_STN_2 & avg\_ECR\_MEDIUM\_STN_3) \\ max\ (avg\_ECR\_CLOSE\_STN_1 & avg\_ECR\_CLOSE\_STN_2 & avg\_ECR\_CLOSE\_STN_3 &) \end{bmatrix} \qquad (3)$$

Equation (3) returns the best average ECR (Best Score) and the class that give this best average. Thus, the Best Score from case 1 experiments represents the case 1 performance and will be used for comparison with the other cases performances.

In case 2, case 3, case 5, and case 7, there are a total of nine separate experiments in each case, using the same experiment setup shown in Table 3.The same formula shown in equation (2) wasmodifiedbysubstitutingECR_FAR_STT$_i$withECR_STT$_i$ andavg_ECR_FAR_STN$_j$withavg_ECC_STN$_i$ as shown below. The best score for each case is given by equation (4).

ECR_STT_i = The equal correct rate produced using Sj training database tested on Si testing database. i, j = 1, 2, 3

$avg\_ECC\_STN_i$ = The average equal correct rate produced using Sj training database tested on S1, S2, and S3 testing database

$$\text{Best Score} = max\ (avg\_ECR\_STN_1 \quad avg\_ECR\_STN_2 \quad avg\_ECR\_STN_3\ ) \qquad (4)$$

Applying equation (4) on case 2, case 3, case 5, and case 7, return the best average ECR (Best Score) for each case respectively.

In case 4, case 6, and case 8, there are a total of three separate experiments in each case. Equation (2) was modified by substituting $ECR\_FAR\_STT_i$ in the equation (2) withECR_ST_i, and$ECR\_FAR\_STN_j$withavg_ECR as shown below.

ECR_ST_i = The equal correct rate produced using acase 4, case 6, or case 8 training database, tested on Si testing database. i, j = 1, 2, 3

avg_ECR = The average equal correct rate produced using a case 4, case 6, or case 8 training database training database tested on S1, S2, and S3 testing database

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

63

The *avg_ECR* for each case represents the Best Score for that case. The experiments for all the cases were carried out 20 times and all the Best Score for every caseare shown in Table 4.

Table 4 Comparison of the best average equal correct rates from case 1, case 2, case 3, case 4, case 5, case 6, case 7, and case 8.

Experiment number	The best class from Case 1 experiment	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
1	MEDIUM	64.29	64.29	71.43	71.43	64.29	80.95	73.81	80.95
2	MEDIUM	61.90	66.67	71.43	78.57	64.29	80.95	69.05	88.10
3	MEDIUM	61.90	61.90	73.81	76.19	64.29	80.95	69.05	85.71
4	MEDIUM	57.14	64.29	71.43	78.57	64.29	80.95	66.67	83.33
5	MEDIUM	64.29	61.90	69.05	76.19	64.29	80.95	66.67	85.71
6	FAR	59.52	61.90	66.67	76.19	64.29	80.95	71.43	85.71
7	MEDIUM	59.52	59.52	69.05	78.57	64.29	80.95	71.43	83.33
8	FAR	57.14	59.52	61.90	76.19	64.29	80.95	71.43	88.10
9	MEDIUM	61.90	64.29	73.81	73.81	64.29	80.95	71.43	88.10
10	FAR	61.90	64.29	73.81	78.57	64.29	80.95	71.43	83.33
11	MEDIUM	61.90	69.05	69.05	78.57	64.29	80.95	73.81	88.10
12	FAR	61.90	57.14	73.81	80.95	64.29	80.95	66.67	88.10
13	FAR	61.90	59.52	71.43	80.95	64.29	80.95	66.67	80.95
14	MEDIUM	61.90	59.52	76.19	76.19	64.29	80.95	69.05	88.10
15	MEDIUM	59.52	69.05	66.67	80.95	64.29	80.95	71.43	85.71
16	MEDIUM	59.52	59.52	73.81	80.95	64.29	80.95	66.67	85.71
17	FAR	61.90	64.29	71.43	76.19	64.29	80.95	66.67	85.71
18	MEDIUM	64.29	64.29	73.81	76.19	64.29	80.95	69.05	83.33
19	FAR	64.29	66.67	71.43	78.57	64.29	80.95	71.43	85.71
20	FAR	61.90	64.29	73.81	78.57	64.29	80.95	69.05	85.71

As can be seen from Table 4, the average equal correct rates of the first 10 experiments gives 60.95±2.56 standard deviation (std), 62.86±2.30std, 70.24±3.77std, 76.43±2.37std, 64.29, 80.95, 70.24±2.31std and 85.24±2.46std for Case 1, Case 2, Case 3, Case 4, Case 7, and Case 8 respectively. The MODE of the Case 1 best class is the MEDIUM class which mean on the first 10 run, the MEDIUM class gave the best performance most of the time in case 1. Also from Table 4, the average rate of all the 20 experiments gives 61.43±2.13std, 63.10±3.33std, 71.19±3.36std, 77.62±2.49std, 64.29, 80.95, 69.64±2.43std, and 85.48±2.31std for Case 1, Case 2, Case 3, Case 4, Case 7, and Case 8 respectively. The MODE of the Case 1 best class is still the MEDIUM class. This shows that the MEDIUM class gives the best performance most of the time in case 1.

Figure 3 and figure 4 shows the comparison between all cases performances for the first 10 experiments run and all the 20 experiments run respectively.
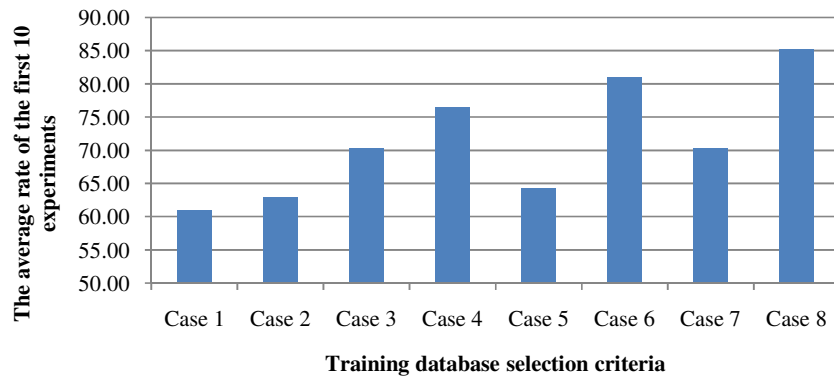
Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

64

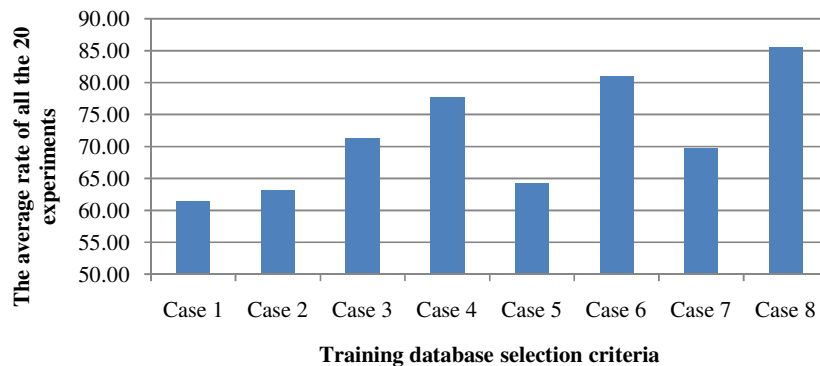**Figure 3** The average rate of the first 10 experiments

Figure 4 The average rate of the all the 20 experiments

   Comparing the Case 1 and Case 2 results shows that using one image randomly selected from each of the three classes of the same database session gives better equal correct recognition rate than using three images randomly selected from a single class of the same database session. Comparing Case 2 and Case 3 results shows that recognition performance improves as the number of images per class increases. Comparing Case 3 and Case 4 results shows that using one image per class from all the database sessions gives better recognition performance than using three images per class from the same database session. Comparing Case 2 and Case 5 recognition results shows that using one mean image from each class from the same database for training gives better result than using one image from each class from the same database for training. Comparing Case 4 and Case 6 recognition results shows that using one mean image from each class from all the database sessions gives better result than using one image from each class from all the database sessions for training. Comparing Case 5 and Case 7 recognition results shows that using one image and one mean image from each class from the same database session gives better result than using one mean image from each class from all the database sessions. Comparing Case 6 and Case 8 recognition results shows that using one image and one mean image from each class from all the database sessions gives better result than using one mean image from each class from all the database sessions. Comparing Case 8 with both Case 4 and Case 6 also shows that recognition performance improves as the number of images per class increases.

   The performance of our system is also compared with several PCA based methods as shown in Table 5. Although direct comparison between our system and their system is not meaningful as different training and testing conditions

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

65

are used. Nevertheless, our system performance of 85.48%±2.31% (std) is comparable to that of [6] with an average rate of 85.53%

Table 6 Comparison of several PCA based face recognition in surveillance application.

Methods	Recognition rate
Vera-Rodriguez et al, 2010	CLOSE = 91.47% correct recall rate MEDIUM = 85.53% correct recall rate FAR = 64.31% correct recall rate
YuxiPeng, 2011	1.0 meter distance = 8.5% correct recall rate
Our proposed method (Case 8)	85.48%±2.31% (std) correct recall and reject rate. (The average Best Score from 20 experiments run)

8. Conclusion

In this work, an investigation of the effect of the method of selecting training images on theperformance of face recognition in surveillance application has been carried out. It was found that using three images from MEDIUM distance class for training gives better performance than FAR and CLOSE distance class regardless of the database session used.However, using one image from each class from the same session for training gives better recognition performance compared to the one that uses three MEDIUM class images from the same session for training.It was also found that as the number of images per class increase, the recognition performance also increases. Lastly, it was found that by using one image per class and one mean image per class from all the database sessions gives the best recognition performance.Thus, the best choice for training data is to select them from all available classes from all available database sessions. In our further work, we intend to study whether using YC_BC_R colour space will improve the recognition rate for surveillance application.

9. References

Anlai Sun, Wei Hu, Ying Xiong, Jian Li, QingE Wu, 2014. A hierarchical target recognition method based on image processing. *Artificial Intelligence Research, Vol. 3, No. 3*.

Davis, M, Stefan Popov, Cristina Surlea, 2010. Real-Time Face Recognition from Surveillance Video.*Studies in Computational Intelligence Vol.332*.Pages 155-194.

Jillela, R.R., Ross, A.,2009.Adaptive frame selection for improved face recognition in low-resolution videos.*Neural Networks, IJCNN*.Pages 1439-1445.

Kamgar-Parsi, B., Lawson, W., Kamgar-Parsi, B., 2011.Toward Development of a Face Recognition System for Watchlist Surveillance.*Pattern Analysis and Machine Intelligence, IEEE Transactions. Vol.33*. Pages 1925 - 1937.

National ICT Australia Limited, 2014.http://arma.sourceforge.net/chokepoint/.

Vera-Rodriguez, R., J. Fierrez, P. Tome, J. Ortega-Garcia, 2010. Face Recognition at a Distance: Scenario Analysis and Applications. *Advances in Intelligent and Soft Computing, Vol.79*. Pages 341-348.

Wiliem, Arnold and Madasu, Vamsi Krishna and Boles, WageehW. and Yarlagadda, Prasad K.D.V, 2007. A feature based face recognition technique using Zernike moments. *Proceedings RNSA Security Technology Conference*.Pages 341-355.

Yoo, S., Park, R. and Sim, D., 2007.Investigation of Color Spaces for Face Recognition.*In Proceedings of Machine Vision Application*.Pages 106-109.

YuxiPeng, 2011. Face Recognition at a Distance: a study of super resolution. Master thesis, University ofTwente.

Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A., 2003 Face recognition: A literature survey.*ACM Computing Surveys (CSUR), Vol.35, Issue 4*. Pages 399-458.

Regular Issue
Vol.3 n.4
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

66