



A Comparison of the $Y C_B C_R$ Color Space with Gray Scale for Face Recognition for Surveillance Applications

Jamal Ahmad Dargham^a, Ali Chekima^a, Ervin Gubin MOUNG^a and Segiru Omatu^b

jamalad@ums.edu.my, chekima@ums.edu.my, menirva.com@gmail.com, omatu@rsh.oit.ac.jp
^aFakulti Kejuruteraan, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
^bFaculty of Engineering, Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, 5-16-1, Omiya, Asahi-ku, Osaka, 535-8585, Japan.

KEYWORD

Principal Component Analysis; YCBCR color space; face recognition; surveillance applications

ABSTRACT

Face recognition is an important biometric method because of its potential applications in many fields, such as access control and surveillance. In this paper, the performance of the individual channels from the YCBCR color space on face recognition for surveillance applications is investigated and compared with the performance of the gray scale. In addition, the performance of fusing two or more color channels is also compared with that of the gray scale. Three cases with different number of training images per persons were used as a test bed. It was found out that, the gray scale always outperforms the individual channel. However, the fusion of CBxCR with any other channel outperforms the gray scale when three images of the same class from the same database are used for training. Regardless of the cases used, the CBxCR channel always gave the best performance for the individual color channels. It was found that, in general, increasing the number of fused channels increases the performance of the system. It was also found that the gray scale channel is the better choice for face recognition since it contain better quality edges and visual features which are essential for face recognition.

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. There are several reports that have reported the effect of color spaces in biometric fields. Karimi and Devroye (2007) reported that using color images with PCA-based face recognition system improves the recognition accuracy. Chaves-González *et al.* (2010) investigated the effect of color spaces for face detection application using RGB, CMY, YUV, YIQ, $Y P_B P_R$, $Y C_B C_R$, $Y C_G C_R$, $Y D_B D_R$, HSV— or HSI— and CIE-XYZ. They tested 15 colored frontal face images of different people obtained from the AR face database. They found out that using the color information can improve the accuracy of the biometric system and reported that HSV color space is the best one for skin detection with a success rate of 95.06%. Yoo *et al.* (2007) investigated the effect of color spaces



(RGB, HSV, $YC_B C_R$, and $YC_G C_R$) for face recognition. They reported that the $YC_B C_R$ color space gives the best performance (92.3% correct recognition) when tested on the color FERET database using PCA-based face recognition method. Chelali *et al.* (2015) proposed a PCA-based face recognition system that uses the segmented skin region of the face as the features for their system. Three color spaces were investigated for skin region segmentation; RGB, HSV, and $YC_B C_R$. They used three databases for their experiment, two in a controlled environment with the same illumination and background (Computer Vision Center database and LCPTS Facial database), and the third database (CALTECH database) in an uncontrolled environment that present a variation in illumination, pose and background. The best reported recognition rate (96%) is obtained when the $YC_B C_R$ color space is used. As can be seen from the review, there are evidence that using color can enhance face recognition systems and that most researchers reported that the $YC_B C_R$ color space gives better recognition rate than other color spaces. Since, we have used gray scale images obtained from color images in our previous work (Dargham *et al.*, 2015), in this paper a comparison of the performance of the $YC_B C_R$ color space with gray scale for face recognition for surveillance application is carried out. In addition, an investigation of which channel or combination of channels from the $YC_B C_R$ color space gives the best performance is carried out.

2. Surveillance Dataset and Database Preparation

The ChokePoint Dataset (National ICT Australia Limited, 2014) is one of the freely available surveillance databases under real-world surveillance conditions. It 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 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 contains 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. The ChokePoint Dataset provided the eyes coordinates for their color video images. All the frontal face images are cropped according to their given eyes coordinate and the eyes are aligned so that their eyes are in the same horizontal line. Figure 2 shows the face cropping template. Assuming E_L is the length between the right and left eye and M is the middle point between the right and left eye, then the length from the right eye to R , and from M to T , as well as from the left eye to L , is equal to E_L and the length from M to B is $2E_L$.

In our previous work (Dargham *et al.*, 2015), the gray scale database provided by the ChokePoint Dataset was used while in this paper, the color video images of the ChokePoint Dataset are used. The video frames were divided into three classes, namely; FAR, MEDIUM, and CLOSE. Each person has 18 frames per class. The first 18 frames from the 54 frames for each person are assigned to the FAR class. The next 18 frames after the FAR class frames are assigned to the MEDIUM class and the last 18 frames are assigned to the CLOSE class. A total of 4536 cropped video images with frontal face of a single person were used. Three databases representing three different session containing 1512 video images each were created. They represent 28 different persons. 14 persons are used for training and recall testing, and the other 14 different persons are used for reject testing only. The image size of the grayscale frontal face that ChokePoint Dataset provided is 96 by 96 pixels. Thus, in this work, the cropped video images will be set to 96 by 96 pixels. The grayscale and $YC_B C_R$ color space image obtained from the color video will be used in this work and the performance of both the gray scale and the $YC_B C_R$ will be compared. The color channels that are investigated are the C_B , C_R , C_B multiplied by C_R ($C_B \times C_R$), the ratio of C_B over C_R (C_B/C_R), and the differences between C_B and C_R ($C_B - C_R$).



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

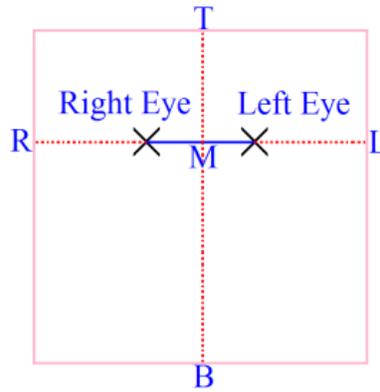


Figure 2: The face cropping template

3. The Training and Testing Database

In our previous work (Dargham *et al.*, 2015), we have investigated the effect of training data selection on face recognition in surveillance application by using the provided gray scale images of the ChokePoint Dataset as our data. Eight cases of selecting training images were proposed. It was found that, Case 1 (using FAR class) gives the lowest performance while Case 8 gives the best performance followed by Case 6. Thus, in this work Case 1, Case 6, and Case 8 selection methods will be used to create the training database. Table 1 shows the training database criteria while Table 2 shows the training database details for each cases.

Table 1: The proposed training database criteria

Case	Selection criteria
1	Three images per person randomly selected from the same class and from the same database session will be used
6	Nine images per person from all the database session with 1 mean image per class will be used.
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.

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	3	1 class, 1 session	3
6	126 images	14	9	3 classes, all session	1
8	252 images	14	18	3 classes, all session	2

Two testing databases were created. The first database, Client test database, has 756 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 756 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) (Dargham *et al.*, 2012; Turk *et al.*, 2004) as shown in Figure 3. It consists of an offline training phase and an online testing phase. During the training phase, the entire set of training images are projected, using PCA, to a lower dimensional feature space. This projection will produce sets of feature vectors having much smaller dimensions compared to the original image dimensions. The feature vectors, a lower dimensional representation of the grayscale or $YC_B C_R$ 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.

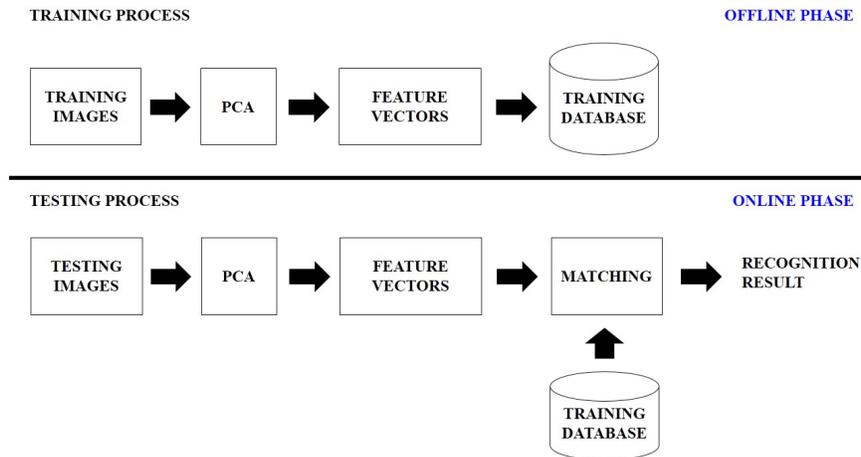


Figure 3: Block diagram of the PCA based face recognition system

5. The Training Phase

For the training and testing phases, the Euclidean distance, a popular distance measurement in biometric (Turk and Pentland, 1991; Zhao *et al.*, 2003; Yoo *et al.*, 2007), is used to calculate the distance between two vectors. 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 proposed. Let f be the feature vector from the training database, then, the threshold

value, t , is the largest calculated Euclidean distance between any two feature vectors from the training database, denote as $\max\{\|f_M - f_N\|\}$, divided by a tunable parameter denoted as $Tcpara$ as shown in equation (1).

$$t = \frac{\max\{\|f_M - f_N\|\}}{Tcpara} \quad (1)$$

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 used during the testing phase where it will be compared with a Euclidean distance value generated by the test person image during the matching task.

6. The Testing Phase

Let $i = 1, 2, \dots, 42$ and $M = 1, 2, \dots, Z$ and Z is the size of the training database. Let $T_R = \{f_{R1}, f_{R2}, \dots, f_{R42}\}$ 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 and denoted as $E = \{e_1, e_2, \dots, e_{42}\}$ where E is a set that contain all the 42 minimum Euclidean distances 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 $L = \{l_1, l_2, \dots, l_{42}\}$. 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. For a **Recall** test, if a test person T_i is correctly matched to the same person P_i in the training database, it is a **Correct Recall**. If a test person T_i is incorrectly matched with person P_j , where i and j are not the same person, it is a **False Acceptance**. If a test person T_i is of a person P_i in the training database but rejected by the system, it is a **False Rejection**. For a **Reject** test, if a test person T_i , from the Imposter test database is rejected by the face recognition system, it is a **Correct Reject**. If a test person T_i from the Imposter test database is accepted by the program, it is a **False Acceptance**. Let NCR = number of correct recall, NCJ = number of correct reject, and NT = number of test, then **Correct Recall Rate** = NCR/NT and **Correct Reject Rate** = NCJ/NT . For this work, the threshold tuning parameter, $Tcpara$, was set so that the system has equal correct rates for both recall and reject. This classification rate is defined as the **Equal Correct Rate (ECR)**.

7. Data Fusion Strategy

The fusion strategy fused the outputs from two or more face recognition systems. The logic operator chosen for the data fusion strategy is the OR operator. For the recall test, if at least one system give a correct matching and the other system(s) give wrong matching or not found, then a correct match is found. Otherwise, the fusion system will give a no match is found. For the reject test, if at least one system give a correct reject and the other system(s) give false acceptance, then the fusion system gives a correct reject. Otherwise, the fusion system will give a false acceptance. The use of AND operator will make the fusion strategy become more strict, for example, for a recall test, all the individual system need to give a correct match found in order for the fusion system to give a correct match found. Thus, the OR operator were selected since it require at least one system to give a TRUE statement for the fusion system to give TRUE statement too.

8. Results and Discussion

Each experiment for each of the cases was carried out 10 times and the average of the recognition rates from the 10 experiment runs is used for comparison. Figure 4 shows the performance of the gray scale with individual channels of the $YC_B C_R$. As can be seen, from Figure 4, the grayscale always outperform the other five $YC_B C_R$ channels, giving 76.41%, 100%, and 100% rate in case 1, 6, and 8 respectively. Comparing the color channels, it was found that the $C_B \times C_R$ channel always outperform the other channels for all the cases tested. From the results of case 1, it can be seen that the C_B , C_R , C_B/C_R , and $C_B - C_R$ channels are not suitable when only three images from one class are used for training as all of them give an equal error rate below 50%. As can be seen from Figure 4, the performance of all channels improves as the number of images per person increases from 3 images in case 1, to 9 images in case 2 and then to 18 images in case 3.

The fusions of several channels from the $YC_B C_R$ color space are also carried out to investigate whether or not the fusion strategy can compete with the gray scale performance. The proposed fusion strategies are shown in Table 3. The average rate of the recall and reject rates given by the fusion strategy is used for comparison and is shown in Figure 5.

As can be seen from Figure 4 and Figure 5, the fusion of two or more channels always gives better performance than the individual channel regardless of the number of images used for training or the data set from which they are selected. As can be seen from Figure 5, the fusion of the two best channels slightly outperforms the gray scale for case 1. In addition, as the number of fused channels increases so does the performance of the system albeit not at the same rate.

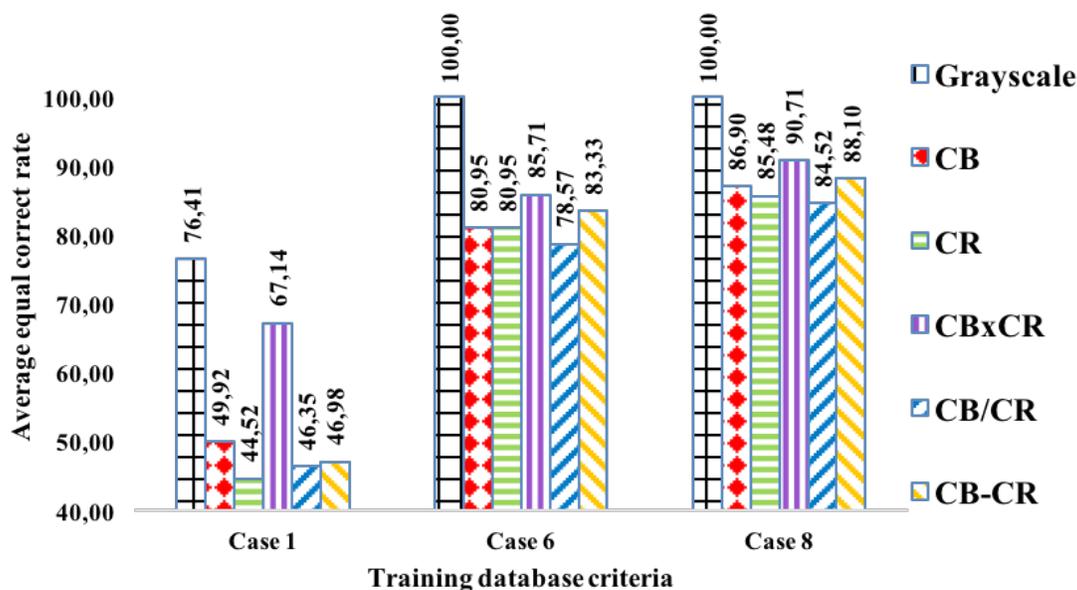


Figure 4: Average equal correct rate for the individual color channels compared to that of the gray scale

Table 3: The proposed fusion strategy

	Case 1	Case 6	Case 8
Fusion of the two highest rate	$C_B, C_B \times C_R$	$C_B \times C_R, C_B - C_R$	$C_B \times C_R, C_B - C_R$
Fusion of the three lowest rate	$C_R, C_B/C_R, C_B - C_R$	$C_B, C_R, C_B/C_R$	$C_B, C_R, C_B/C_R$
Fusion of the three highest rate	$C_B, C_B \times C_R, C_B - C_R$	$C_B, C_B \times C_R, C_B - C_R$	$C_B, C_B \times C_R, C_B - C_R$
Fusion of the one highest rate and two lowest rate	$C_R, C_B \times C_R, C_B/C_R$	$C_R, C_B \times C_R, C_B/C_R$	$C_R, C_B \times C_R, C_B/C_R$
Fusion of all the five channels	All channels	All channels	All channels

It can also be seen that the fusion of channels that individually gives good performance will give higher performance than fusing those with lower performance. This suggests that the performance of the individual channel is an important criterion when selecting channels for fusion. In Case 6 and 8 however, the gray scale outperforms all the proposed fusion strategies. This shows that using one image and one mean image from each class from all the database sessions gives the best performance in terms of accuracy and process time since converting RGB to gray scale is faster than converting RGB to $Y C_B C_R$.

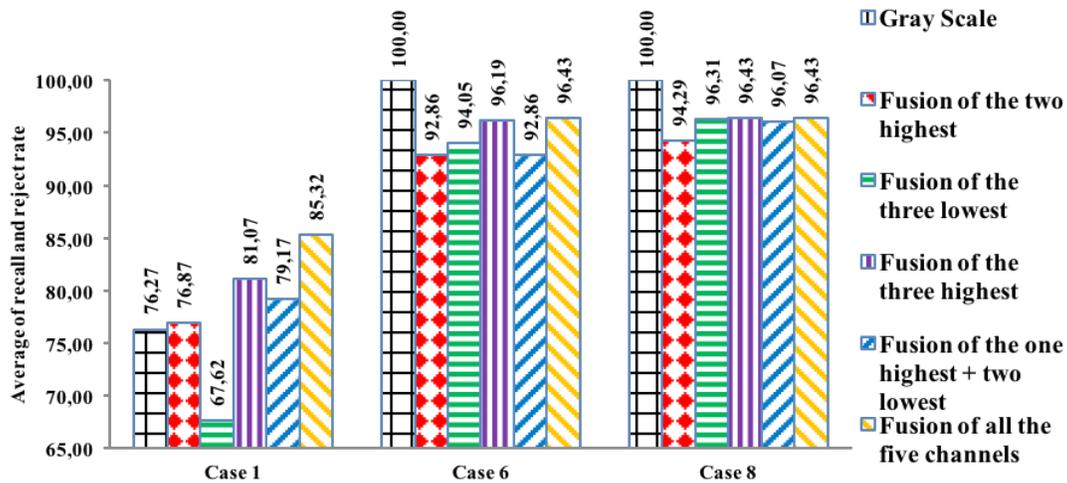


Figure 5: Recognition performance of the data fusion

The results of the individual channels and their fusions, shown in Figure 4 and Figure 5, are new results which were not reported in our previous work (Dargham *et al.*, 2015). However, the results of the gray scale shown in Figure 4 and in (Dargham *et al.*, 2015) are for the same cases. The reason that they had different recognition rates is because of the different training data set used. Nevertheless, the same performance pattern is shown in both works where Case 1 gives the lowest performance while Case 8 gives the best performance.

The results shown in figures 4 and 5 gives rise to an interesting question why some channels gives better results than the other channels. By comparing the grayscale, C_B , and C_R channels from the same image, using human eye observation, one conclusion can be made is that in any image, the important edges or visual features are clearer in gray scale compared to the C_B and C_R channels. This statement is also supported by (John Zhang, 2012) and (Rethunk, 2012). For face recognition, the facial features are important for recognition. Figure 6 shows the average number of edges obtained from the mean face image from all the database sessions for grayscale, C_B , C_R , $C_B \times C_R$, C_B/C_R , and $C_B - C_R$ by using the MATLAB function, *edge(image)*, for case 1, 6, and 8.

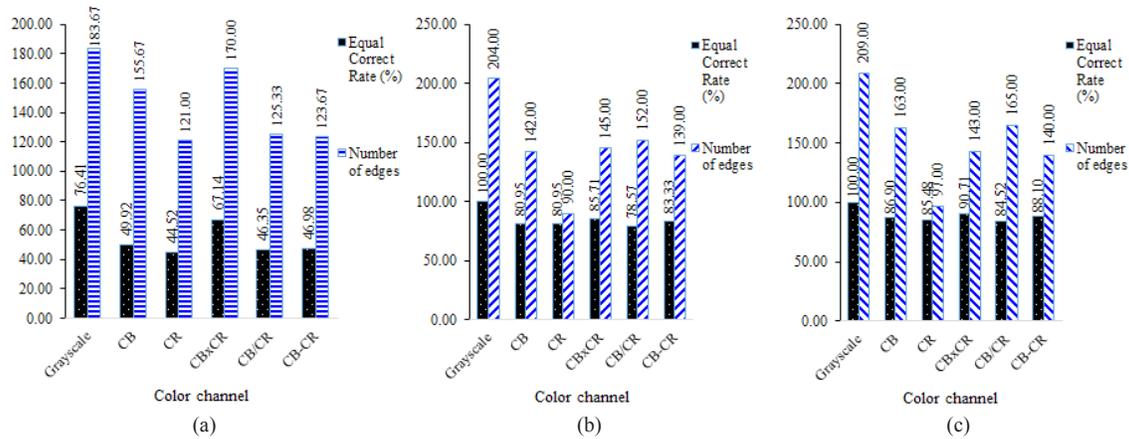


Figure 6: The equal correct rate and the number of edges for (a) Case 1, (b) Case 6, and (c) Case 8

As can be seen from figure 6, from Case 1, when only three images from one class are used for training, we see that the color spaces with higher number of edges tend to give better recognition results. Gray scale has the highest number of edges, and as shown in figure 6, it gave the best recognition performance. Among the $C_B C_R$ color spaces, $C_B \times C_R$ has the highest edges and also has the highest recognition performance as seen in figure 4. It can be seen that C_B / C_R and $C_B - C_R$ has similar number of edges, and their recognition performance is similar too. Overall, for Case 1, the number of edges chart pattern is similar to the average equal correct rate chart pattern.

Also from figure 6, for case 6 and case 8, we can see that the gray scale always has the highest number of edges compare to the other channels and gray scale also has the best recognition performance, same as in case 1. Comparing the gray scale edges in case 1, 6, and 8 also shows an increasing number of edges. This shows that using the gray scale representation, as the number of images used in the training database increases, it can increase the quality of the mean face that were used for the PCA function, thus resulting a better quality of eigenface. Also from figure 6, for case 6 and 8, by comparing the C_B and C_R , it can be seen that C_B has higher number of edges compare to C_R , and as shown in figure 4, C_B also has better performance compare to C_R , except for case 6 where both of them has the same recognition rate. For the case 6 and case 8, although the number of edges chart pattern and the average equal correct rate chart pattern for C_B and C_R shows similar pattern, the number of edges has little effect on the recognition performance. But for $C_B \times C_R$, C_B / C_R , and $C_B - C_R$, there are no direct relation that can be made by comparing the number of edges and their recognition performance shown in figure 4. This could mean that, if a mean image was used in the training database (case 6 and case 8 both uses mean images in their training database), the number of edges has no positive effect on $C_B \times C_R$, C_B / C_R , and $C_B - C_R$ recognition performance.

9. Conclusion

In this work, an investigation of the performance of different channels of the $Y C_B C_R$ color space on face recognition for surveillance applications has been carried out. It was found that, the gray scale gives better performance than any of the individual color channel regardless of the cases used. However, the fusion of $C_B \times C_R$ with any other channel outperforms the gray scale when three images of the same class from the same database were used for training (Case 1). For the $Y C_B C_R$ color space, the best performance is achieved by using the $C_B \times C_R$ channel. The C_B , C_R , C_B / C_R , and $C_B - C_R$ are not suitable to be used when the number of training images per person is small. It was also found that if the differences between individual channels performance vary significantly, the individual channel performance become an important criteria when selecting channels for fusion. It was also found that, for face recognition system where facial features are crucial for the recognition, the gray scale is the better choice since it contain better quality of edges and visual information compare to the other $C_B C_R$ channels.

10. References

- Chaves-González, J. M., Vega-Rodríguez, M. A., Gómez-Pulido, J. A., Sánchez-Pérez, J. M., 2010. Detecting skin in face recognition systems A colour spaces study, *Digital Signal Processing*, Volume 20, Issue 3, pages 806-823.
- Chelali, F. Z., Cherabit, N., Djeradi, A., 2015. Face recognition system using skin detection in RGB and YCB-CR color space, *Web Applications and Networking (WSWAN), 2nd World Symposium*, pages 1-7.
- Dargham, J., Chekima, A., Moug, E., and Omatu, S., 2015. The Effect of Training Data Selection on Face Recognition in Surveillance Application, *Advances In Distributed Computing And Artificial Intelligence Journal*, Volume 3, pages 58-66.
- Dargham, J., Chekima, A., and Moug, E., 2012. Fusion of PCA and LDA Based Face Recognition System, *International Conference on Software and Computer Applications, IPCSIT* Volume 41.
- Karimi, B. and Devroye, L., 2007. A Study on Significance of Color in Face Recognition using Several Eigenface Algorithms, *Canadian Conference Electrical and Computer Engineering (CCECE)*, pages 1309-1312.
- National ICT Australia Limited, 2014. <http://arma.sourceforge.net/chokepoint/>.
- Rethunk, 2012. Why we should use gray scale for image processing? [Online forum comment]. Message posted to <http://stackoverflow.com/questions/12752168/why-we-should-use-gray-scale-for-image-processing>
- Turk, M. and Pentland, A., 1991. Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Volume 3, pages 71-86.
- Yoo, S., Park, R. and Sim, D., 2007. Investigation of Color Spaces for Face Recognition, In *Proceedings of Machine Vision Application*, pages 106-109.
- Zhang, J., 2012. Computer Vision: If you had to choose, would you rather go without luminance or chrominance? [Online forum comment]. Message posted to <https://www.quora.com/Computer-Vision/Computer-Vision-If-you-had-to-choose-would-you-rather-go-without-luminance-or-chrominance/answer/John-Zhang>.
- Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A., 2003. Face recognition: A literature survey, *ACM Computing Surveys (CSUR)*, Volume 35, Issue 4, pages 399-458.

