3rd International Conference on Automation, Control, Engineering and Computer Science (ACECS'16)
Proceedings of Engineering & Technology (PET)
pp. 5-10
Copyright IPCO-2016

Video-Based Face Recognition Using Hidden Markov Models and 2D Discrete Cosine Transform: Application to (VIDTIMIT database)

Beldi Makrem
Laboratoire Signal, Image and technologie de l'Information
Ecole Nationale d'Ingénieurs de Tunis
Université Tunis El Manar
Beldimakrem@yahoo.fr

Lachiri Zied
Laboratoire Signal, Image and Technologie de l'Information
Ecole Nationale d'Ingénieurs de Tunis
Université Tunis El Manar
Zied.lachiri@enit.rnu.tn

***Abstract*— This paper proposes an approach based on two-dimensional discrete cosine transform for face recognition is presented. We also offer a new model for the recognition faces under large variations of illumination in the videos. We showed the robustness of the system with respect to the variation of the luminance by the use of the technique of histogram remapping in the preprocessing phase. The face recognition method is based on Hidden Markov Models (HMM) with the use of a top-to-bottom architecture. Many of these HMM's built for each individual with robust characteristics in wide variation of illumination. Facial features are retrieved via the use of two-dimensional discrete cosine transform 2D-DCT in the parameterization phase. This method use an HMM classification model based, on local approaches with a horizontal sampling technique, and the global feature-based 2D-DCT. Has an accuracy of 95%. The results, to the best of knowledge of the authors, give the best recognition percentage compared to any other method reported so far on VIDTIMIT video database.**

Keywords— ***Face Detection; Face Identification; histogram remapping; Top-to-Bottom HMM; DCT; VIDTIMIT;***

I. INTRODUCTION (*HEADING 1*)

Over the post decades, numerous face recognition research and studies have been carried out in the field of computer vision on video based systems. In speaker verification applications, there is a need to extract information from face that is speaker specific and robust to luminance. Many algorithms and methods have been developed for detection and recognition face.

In face recognition system the first step is detection like skin regions [1] and face detection in color images using PCA [2]. Then the second step is features extraction. In the literature two feature extraction environments exist. The first is local feature extraction characterize a small area of the image. Each descriptor extracted partial information and must therefore be combined with other descriptors to provide a complete representation of the image to be analyzed. Such as Gabor wavelets [3], Local binary pattern [4]. The second is Global Feature the principle is based on the extraction of a set of attributes calculated on the entire image. Such as 2D-DWT, Eigenface [5], Fisherfaces [6] and Zernike Moments [7]. Most facial feature extraction methods undergo certain difficulties like illumination, noise and orientation.

In this paper a novel feature extraction method for 2-D facial shape to create an observation sequence and to build a top-bottom HMM model the technique is motivated by the work of Samaria and Young [8]-[9]. This method is based on discrete cosine transform and histogram remapping. Its shows robustness to pose and illumination variation, leading to significant enhancement in performance of face recognition system in VIDITMIT database.

This paper is organized as follows: in Section 2, the methodology and the methods used are presented describes the Markov Model Hidden high-low with the method of recovery (Overlap) used and we show the different steps of parameterization. Section 3 describes the application we operate the audio-visual VIDTIMIT database to apply our method. Finally, experiments are presented in Section 4.
The system of recognition adopted for this study is The MATLAB: KPMstats, KPMtools and netlab3.3.

II. PROPOSED METHOD

A. System description

In this section, a visual identity verification model is described that is used in our experiments.

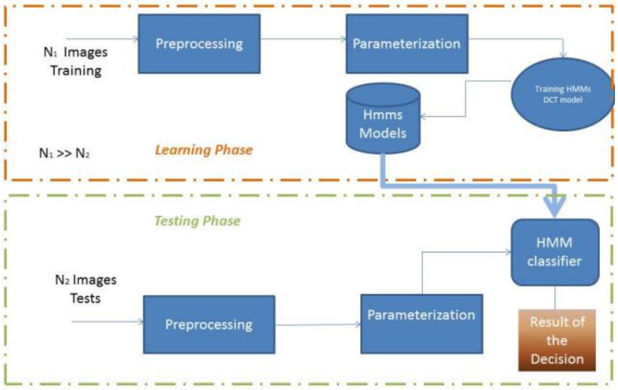

Fig. 1. Face recognition system architecture

B. A 5-State top to bottom Hidden Markov Model

Two basic HMM topologies are studied in the literature, ergodic topology [8]-[10] and the up-down topology (Top to bottom) [8]-[11]. In our work will process and study the concept of an HMM from top to bottom as these models represent facial information in a more natural way. The setting HMM is studied in Section IV with a complete set of experiences.

An observation sequence O is generated from an image of X * Y, using a sampling window X * L, where X * M pixels overlap as illustrated in Figure 2.

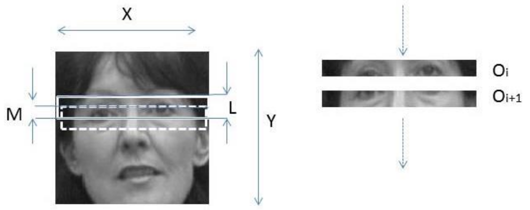

Fig. 2. Sampling Technique for top to bottom HMM

A series of vectors of pixel identification comments is generated, where each observation Oi contains pixel values in the block of lines arranged in a column vector. Each observation vector is thus a block of L lines, and there is an M line overlap between successive observations. The length of the observation T sequence can be obtained by:

$$T = \emptyset\left(\frac{Y-L}{L-M}\right) + 1 \qquad (1)$$

, 1≤ L ≤10 and 0≤ M ≤L-1

Assuming that each face is in a vertical front position, is these characteristics will occur in a predictable order, that is to say the front, then the eyes and the nose, and so on. This order suggests the use of a top to bottom model (non-ergodic), where only transitions between adjacent states of a top-down manner will be allowed. For fixed size face images, there are three HMM parameters that affect the performance of the model: N the number of state HMM, the height of the sampling window L and the amount of overlap M. using this abbreviated notation, a HMM with these parameters will be defined as follows:

H (N, L, M)

Initially, it is assumed that N = 5 is a reasonable number of states based on the intuitive argument that five facial features seem subjective when crossing the face from top to bottom as shown in Figure 3.

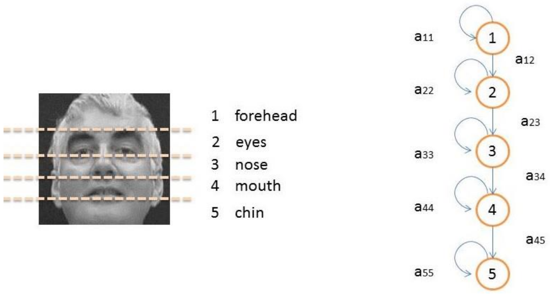

Fig. 3. top to bottom segmentation with five states HMM

C. Feature Extraction 2D-DCT

The use of discrete cosine transform [12] to the face recognition purposes is relatively recent [13]-[14]. Similar to eigenfaces from a mathematical point of view, it is against much faster, both in the learning phase and phase recognition. That being said, that each face image is represented by a vector composed of the original transform coefficients. When a face is presented to the module, its transform is calculated and a number of coefficients is selected for comparison with those in the database. This last step is carried out using a distance or with a classifier.

The first DCT coefficients are stored and when learning and used directly for the identification phase. These correspond to low and medium frequencies contained in the images.

The two-dimensional discrete cosine transformation to each face image size M * N, is defined as:

$$F(u,v) = \frac{2}{\sqrt{MN}} \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)C(u)C(v)\cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)v\pi}{2N} (1)$$

Inverse transformation is:

$$f(x,y) = \frac{2}{\sqrt{MN}} \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)C(u)C(v)\cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)v\pi}{2N} (2)$$

$$x,u = 0,1,2,...,M-1 \;, \; y,v = 0,1,2,...,N-1$$

$$C(u) = \begin{cases} \frac{1}{\sqrt{M}} & u = 0 \\ \sqrt{\frac{2}{M}} & u = 1,2,\dots,M-1 \end{cases} \qquad (3)$$

$$C(v) = \begin{cases} \frac{1}{\sqrt{N}} & v = 0 \\ \sqrt{\frac{2}{N}} & u = 1,2,\dots,N-1 \end{cases} \qquad (4)$$

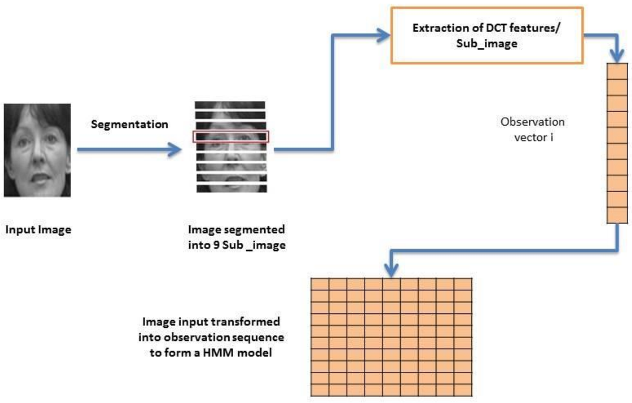
Extraction of DCT features/ Sub_image
Segmentation
Observation vector i
Input Image
Image segmented into 9 Sub _image
Image input transformed into observation sequence to form a HMM model

Fig. 4. feature extraction (The observation vector extraction process)

According to Figure 4, the observation vector (transformed by DCT) brings together the highest values in the upper left corner of the matrix (sub_image) and the lowest values in the lower right corner (high frequencies). Thus the maximum information on each sub_image is concentrated on the upper left of the matrix. We collect for each block a DCT coefficient vector in descending order by applying a zigzag course and then we insert the new vector of DCT coefficients sorted (observation vector) in the overall matrix of DCT coefficients (observation sequence) representing the model of an individual.

III. APPLICATION

A. VIDTIMIT dataset

The base VIDTIMIT [15]-[16] is a multimodal database that contains 2D facial images, video sequences with voice recordings. The VIDTIMIT contains 43 people (19 women and 24 men) recorded in 3 sessions, with a period of 7 days between sessions 1 and 2 is six days between sessions 2 and 3. The period between sessions allows changes of voice, the hair style, makeup, clothes and smelling (which can affect the pronunciation). In the other camera's zoom factor was disrupted randomly mannered after each record. There are 10 video clips per person with an average duration of 4.25 seconds, which is about 100 frames per video. The recording of video footage was made in a noisy office environment using a digital video camera PAL broadcast quality. The video of each person is stored as a numbered sequence of JPEG images with a resolution of 384 x 512 pixels (lines x columns). With a 90% quality setting was used when creating JPEG images.

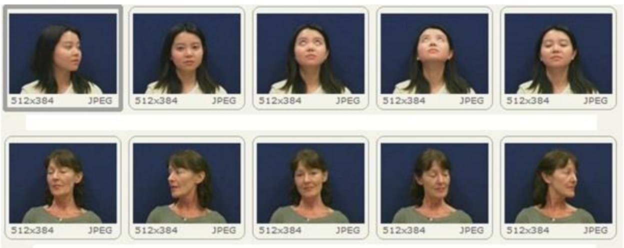

Fig. 5. Sample Dataset of Face Image from VIDTIMIT

B. Face Detection and standardization

Prior to the visual identification of face, the system proceeds to the location of her face. A Viola & Jones detector [17] is used to detect the face in each frame of the sequence. The image database has been divided into several groups: the opposite faces for learning was to detect and treat the session 1 and 2 and opposite faces for the test was extracted from the session 3, these faces having undergone treatment before the stage of parameterization.

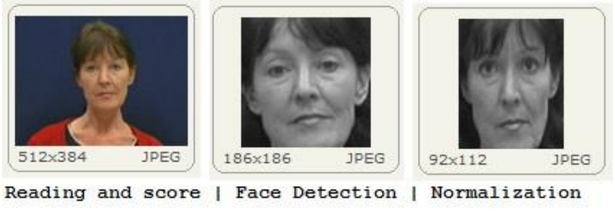

Fig. 6. VIDTIMIT sample face detection and standardization

1) Histogram remapping

The histogram remapping technique is the transformation of the intensity values of pixels of the face image by the transformed row [18]. The rank transformation is essentially a histogram equalization process which makes the histogram of the image approximates the uniform distribution. Where the value of each pixel of a two-dimensional image I (x, y) is replaced by R index (or rank) would be if the image pixels were ordered from the bottom up. For example the value of the most negative pixel is assigned to a rank 1, while the most positive value is assigned to a classification N.

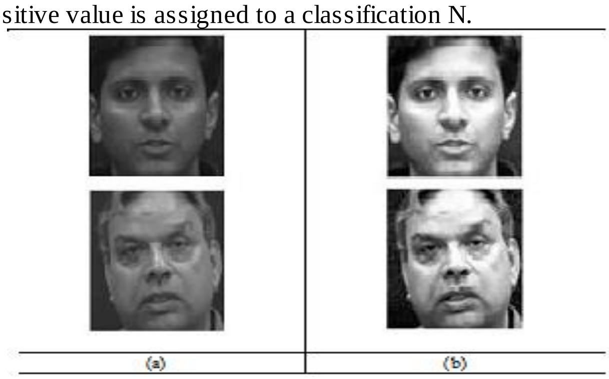

Fig. 7. (a) original image and (b) normalized image with histogram remapping (VIDTIMIT samples)

2) The illumination variation

The appearance of a face in an image varies greatly depending on the illumination of the scene during the shooting (see Figure 8). The lighting variations make face recognition task very difficult. Indeed, the change in appearance of a face of enlightenment, sometimes proves more critical than the physical difference between individuals, and may lead to misclassification of the input images. This has been observed experimentally in Adini and al [19], where the authors used a database of 25 individuals. The face recognition in an uncontrolled environment remains an open research area.

$$F(y,x) = f(y,x) + mx + \delta \tag{5}$$

$$m = \frac{-\delta}{xSize/2} \tag{6}$$

And δ = illumination delta (in pixels) δ = 0 (no change)

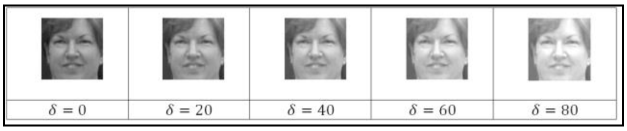

Fig. 8. Visual Effect of change δ luminance

IV. EXPERIMENTAL RESULTATS

This section presents the experiences with various H models using VIDTIMIT database. The purpose of these experiments was to study the effect of different settings on recognition rate. A model was formed for each of the 43 subjects using 40 images training, a total of 1720 learning images. The remaining 10 frames for each subject were used for testing, or a total of 430 test images.

On VIDTIMIT database created in this experiment we will study the effect of the change in luminance and by rank histogram normalizing effect (histogram remapping see Section III.B) on the recognition rate on a set of HMM model formed by 40 training images for each model and test 10 images.

By varying the size of sampling window, percentage overlap and number of DCT coefficients, experiments are carried out with the same set of images to test the efficacy the proposed scheme.

The second experiment investigated the effect of the DCT block size on the system recognition rate with a variation in the number of DCT coefficients in the feature vector. The results in Table 2 indicates that the use of 21 coefficients as optimum size for 2D DCT using 8 * 8 pixel blocks with an overlap rate set at 50% (L = 10 and M = 5) gives a better performance.

TABLE I. PERFORMANCES OF VARIATION IN DCT DIMENSIONALITY (FEATURE VECTORS)

Dimension DCT coefficients	bloc 8*8 2D-DCT	bloc 16*16 2D-DCT
7	69	69,00
15	85	84,50
21	86	82
36	83	80
49	83,25	79
61	82,25	80,20

Following the above analysis, it is reasonable to expect better results from recognition when a larger value of M is used. To determine the effect of M on performance recognition, a full set of experiments were performed with the fixed number of states N = 5 as indicated in section 3 and 21 8 * 8 coefficients 2D-DCT blocks the height of the window in the range of 2 <= L <= 10 pixels and overlap can all 0 <= M <= L - 1 pixels. The results are summarized in Table2, where the error rate is expressed as a percentage, and the overlap is in units of pixels. The performance seems to improve recognition of duplication increases, in line with expectations. A greater overlap, however, implies a greater value of T and the complexity of calculations required varies linearly in the identification process with T.

TABLE II. PERFORMANCE VARIATION OVERLAP BETWEEN WINDOWS.

Dimension	Overlap M	Bloc 8*8 2D-DCT
21	2	76
21	3	80,25
21	5	82,50
21	7	92
21	9	84

On the database VIDTIMIT create model in Section B, in this experiment we will study the effect of the change in luminance (see Section III.B.2) and the histogram normalizing effect by rank (remapping of histogram see section III.B.1) on the recognition rate over a set of HMM model consisting of 20 training images for each model and 5 test image to each model in fixed terms (number of 8 * 8 DCT coefficient block = 21 and an overlap rate of 75%). Based on these data, the method of histogram remapping to give higher face recognition rate compared with that of the method of variation of the luminance. This indicates that Rank histogram normalization provides superior reliable feature extraction capabilities to that variation of luminance. Figure9 shows the considerable influence on face recognition rate.

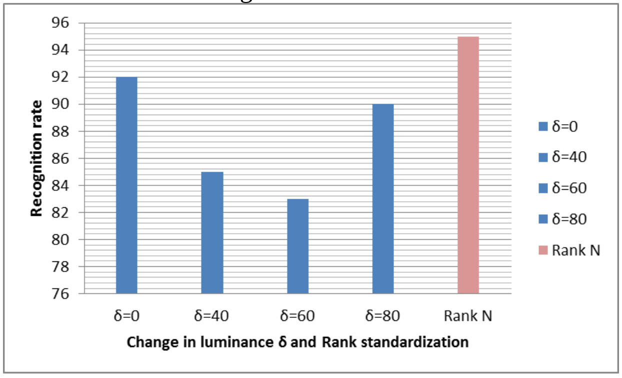

Fig. 9. Effect of the luminance variation histogram and the histogram remapping on the face recognition rate.

The highest recognition score achieved is 95% i.e. only 21 images out of 430 images is misclassified with a sampling window of 8 x 8 with an overlap of 75% using 21 significant DCT coefficients and rank normalization in pre-processing.

The comparative results of some of the recent regimes on VIDTIMIT video database are given in Table 3. Note that the extraction of the descriptors processed by the discrete cosine method in the proposed parameterization phase and utilization HMM models top to bottom in the classification phase gives a better percentage of recognition of one of the proposed methods. Results shown are taken from articles published by the authors, which are indicated by reference numbers in brackets.

TABLE III. COMPARATIVE RECOGNITION RESULTS OF SOME OF THE OTHER METHODS AS REPORTED BY THE RESPECTIVE AUTHORS ON VIDTIMIT DATABASE.

Methods	Parameterization	Training images	Test images	Identification rate in %	References
HMM	DCT	1720	430	90	Proposed
HMM	DCT using rank normalization	1720	430	95	Proposed
GMM	2D-DCT using UBM normalization	430	215	85	[20]
PCA		430	215	87	[21]
PCA	DT-CWT	430	215	91	[22]
SVM + PCA	Egenfaces	1720	430	75	[23]
SVM + LDA	Fisherfaces	1720	430	87	[23]

For the last two methods, we have just evaluated these systems developed [23] in the VIDTIMIT database with the same conditions for the pre-processing.

III. Conclusion

Challenges like face detection and reduced lighting one was reached on the VIDTIMIT database. Face detection in a video sequence was done using the viola & jones algorithm. We use a histogram remapping technique in the pretreatment phase to reduce the effect of the change in luminance. In the phase of parameterization is to use the composition in Haar wavelet for extracting observation sequences used by the HMM classifier with a downward architecture. This paper has detailed work done on face processing using a novel approach involving Hidden Markov Models. This paper has illustrates how these hybrid models can be used to extract facial bands and automatically segment a face image into meaningful regions, showing the benefits of simultaneous use of statistical and structural information. We showed how the segmented data can be used to identify different subjects. Successful segmentation and identification of face images was obtained. In this article it has obtained good results with a significant increase in the recognition accuracy was clearly observed through the experiments. A rate of 95% is obtained by recognition experiments on VIDTIMIT database that has undergone pretreatment to make it usable.

V. References

[1] J. Haddadnia, M. Ahmadi, and K. Faez, "An Efficient Feature Extraction Method with Peudo Zernike Moment in RBF Neural Network Based Human Face Recognition System", *EURASIP Journal on Applied Signal Processing JASP*, vol. 9, pp. 890-891, 2003.

[2] Qiong Liu, Guang-zheng Peng, "A Robust Skin Color Based Face Detection Algorithm", *Informatics in Control, Automation and Robotics (CAR), 2nd International Asia Conference on*, vol.2, pp. 525 – 528, 2010.

[3] Dae Young Ko, Jin Young Kim and Seong Joon Baek."A study on the implementation and robustness of face verification method under Illumination changes", *Robot and Human Interactive Communication, ROMAN 2004. 13th IEEE International Workshop*, pp. 259 – 263, 2004.

[4] Conrad Sanderson and K. Kuldip Paliwal, "likelihood normalization for face authentication in variable recording conditions", *IEEE transactions on Image Processing*, vol.1, pp. 301-304, 2002.

[5] Ojala. T, Pietikainen. M, Harwood. D., "A comparative study of texture measureswith classification based on feature distributions*". Pattern Recognition*, vol.29, pp. 51-59,1996.

[6] Vinay, K.B. , Shreyas, B.S. "Face Recognition Using Gabor Wavelets", *Fortieth Asilomar Conference on:Signals, Systems and Computer*, pp:593-597, oct 2006.

[7] H. Yu, J. Yang," A Direct LDA Algorithm for High-Dimensional Data - with Application to Face Recognition", *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.

[8] F. Samaria, "Face recognition using hidden Markov models", *Ph.D. Thesis*, University of Cambridge, 1994.

[9] B. Menser and F. Muller, "Face Detection in Color Images Using Principal Component Analysis", *Image Processing and Its Application,. Seventh International Conference on (Conf. Publ. No. 465),IET,Volume:2*, pp. 620 - 624 ,1999.

[10] Kumar S. A. S, Deepti D. R. And Prabhakar B, "Face Recognition Using Pseudo-2d Ergodic HMM", *Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings. IEEE International Conference on*, vol. 2, 2006.

[11] F.s. Samaria and A. Harter, "Paramétrisation of stochastic model for human face identication", *in Proceedings of the Second IEEE Workshop on Application of Computer Vision*.1994.

[12] Ziad M. Hafed and Martin D. Levine," Face recognition using discrete cosine transform". *International Journal of Computer Vision*, 43(3):167–188, July – August 2001.

[13] V.V. kohir, U. B. Desai, "Face recognition using a DCT-HMM approach", *Proc. IEEE Workshop on Applications Of computer Vision (WACV'98)*, pp.226-231,1998.

[14] C. Sanderson, "Automatic Person Verification Using Speech and Face Information", PhD Thesis, School of Microelectronic Engineering, Griffith University, Brisbane, Australia, 2002.

[15] C. Sanderson, "*Automatic Person Verification Using Speech and Face Information"*, PhD Thesis, School of Microelectronic Engineering, Griffith University, Brisbane, Australia, 2002.

[16] C. Jankowski, A. Kalyanswamy, S. Basson and J. Spitz," NTIMIT: A Phonetically Balanced, Continuous Speech Telephone Bandwidth Speech Database", *Proc. International Conf. Acoustics, Speech and Signal Processing*, Albuquerque, Vol. 1, pp. 109-112, 1990

[17] Viola, Paul and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume: 1, pp.511–518,2001.

[18] Štruc, V., Žibert, J. in Pavešiæ, N.,"Histogram remapping as a preprocessing step for robust face recognition", *WSEAS transactions on information science and applications*, vol. 6, no. 3, pp. 520-529, 2009.

[19] Y. Adini, Y. Moses, S. Ullman, Face recognition: "The problem of compensating for changes in illumination direction", *IEEE Trans. Patt.Anal. Mach. Intell. 19*, pp.721–732, 1997.

[20] Conrad Sanderson and K. Kuldip Paliwal, " likelihood normalization for face authentication in variable recording conditions", IEEE transactions on Image Processing, vol.1, pp. 301-304, 2002.

[21] Dae Young Ko, Jin Young Kim and Seong Joon Baek."A study on the implementation and robustness of face verification method under Illumination changes", *Robot and Human Interactive Communication, ROMAN 2004. 13th IEEE International Workshop*, pp. 259 – 263, 2004.

[22] Gunawan Sugiarta Y B, Riyanto Bambang, Hendrawan and Suhardi, "Feature Level Fusion of Speech and Face Image based Person Identification System," *Proceedings of the Second International Conference on Computer Engineering and Applications*, pp. 221-225, 2010.

[23] Becker, B.C., Ortiz, E.G., "Evaluation of Face Recognition Techniques for Application to Facebook," *in Proceedings of the 8th IEEE International Automatic Face and Gesture Recognition Conference*, 2008.