Multimodal 2D-3D Face Recognition

Gawed M. Nagi, Rahmita Rahmat, Muhamad Taufik, and Fatimah Khalid

***Abstract*—Up to date, many advances have been made to 2D face recognition (2D FR) due to its broad range of applications in security and commercial areas as well as in smart devices. However, 2D FR is still quite vulnerable under unconstrained conditions of the image acquisition process. To overcome 2D FR limitations, researchers shift to 3D face recognition technology but this technology is computationally expensive and inapplicable to real-world face recognition systems. Multimodal 2D-3D face recognition can combine the strength of both 2D and 3D modalities. In this paper a multimodal 2D-3D face recognition approach has been proposed based on geometric and textural characteristics of 2D and 3D modalities. The conducted experiments show that the proposed approach achieved promising results with illumination and head pose variations. The performance is evaluated using the landmark Bosphorus facial database**

***Index Terms*—2D-3D face recognition, geometric invariants, local binary pattern (LBP), k-nearest neighbor (kNN).**

I. Introduction

In 2D face recognition systems, three challenges have consistently caused problems to the robustness and reliability of such systems. These challenges namely head pose (viewpoint) differences, illumination changes, and facial expression variations as reported by many researchers [1], [2], and such challenges have consequently motivated many researchers to trend towards 3D face recognition. On the other hand, 3D face images acquisition by laser scanners and digitizers is still quite expensive and impractical for most real-time applications even these devices gradually become cheaper and faster. Thus, 2D-3D face recognition can exceed the limitations of both 2D face sensitivity against pose and illumination, and the inapplicability of 3D capturing process. Thus the main objective of multimodal 2D-3D face recognition approaches is to offer a robust system with satisfactory performance while keeping the system practical.

II. Related Work

As a new trend in face recognition area, Multimodal 2D-3D face recognition has recently received more attention and a great deal of research effort has been dedicated to this direction of face recognition (FR). The efficient fusion of both sources of information; texture (2D) and shape (3D) can increase the overall performance of FR which is critical in many security and commercial recognition systems [3], [4].

Many approaches have been proposed to utilize the information of both 2D and 3D modalities for recognition. These approaches have demonstrated that the performance of multimodal systems outperforms significantly the performance of using either 2D or 3D alone [3][5].

Some of multimodal systems do the fusion in early stages while others do the fusion with the last stages of recognition. As example of the fusion of 2D and 3D modalities on the score or decision level, Hüsken et al. [3] developed a successful 2D+3D face recognition technique called hierarchical graph matching (HGM). HGM algorithm is applied to both 2D (texture) and 3D (depth) images and the fusion on the score level (the weighted sum of the matching scores) of both modalities shows a higher recognition rate. 96.8% of recognition rate is stated on FRGC dataset.

Lu *et al*. [4] proposed an approach that integrated the shape and texture modalities for recognition in which five 2.5D scans of the subject with different views are registered by Iterative Closest Point (ICP) and merged to construct 3D face models for enrollment. A set of feature points is extracted for the purpose of alignment between the 2.5D test scan and 3D model and to perform surface matching using a point-to-plane distance metric. During the surface matching stage, for each test scan, an intensity image is dynamically generated and another matching process namely appearance -based matching is performed using the linear discriminant analysis (LDA). Finally, the matching scores obtained by the two matching components are combined to make the final decision. A multimodal recognition rate of 99% and 77% is achieved for neutral faces and for smiling faces, respectively using a database of 200 gallery and 598 test faces. It is noted that the recognition rate is dropped down with expressed faces (smiled) as well as other facial expressions are not included.

A mixed 2D-3D information approach for face recognition is proposed by Tang et al. [6] in which a new method namely HaarLBP is presented for 2D faces representation in which the faces decomposed into four regions using 2D Haar wavelets and then local binary patterns (LBP) technique is applied to extract face features. The 3D morphable model (3DMM) is employed to create the virtual 3D faces. Then, five geometrical features from the virtual 3D faces are extracted to assist the face recognition. The 2D HaarLBP feature is integrated with the five geometric features of virtual 3D faces using a linear weighted scheme and the nearest-neighbor classifier (NN) is employed to perform the recognition task. The fusion results of 92.5% and 93.0% are stated on ORL and JAFFE2, respectively.

As attempt to build practical and robust face recognition systems, a few recent approaches have been proposed to use 3D images for enrollment as reference data while performing the identification or authentication using 2D images as probe data. Among those, Riccio & Dygelay [7] defined an approach in which both a 3D model and a frontal 2D image are acquired to enroll a person, while during testing in the form of verification and identification only a 2D image is

Manuscript received March 1, 2013; revised April 25, 2013.

The authors are with Universiti Putra Malaysia (UPM), Malaysia Gawed. M. Nagi (e-mail: gnagi2000@yahoo.com)

DOI: 10.7763/IJFCC.2013.V2.253

needed. They computed geometric invariants only from a 2D image of the face, but other invariants required 3D Point coordinates to be verified.

III. Approach Description

In this work, the interest lies in defining an approach which combines both facial texture information and shape geometry of 2D and 3D facial images to perform the face recognition with the presence of head pose variations. This approach follows the coarse-to-fine recognition scheme. In the initial coarse phase, the facial texture and geometric invariants of 2D facial images are used to shortlist the candidates in the enrolled facial database into 10 subjects who are roughly matching the 2D test facial image. In the recognition phase, only 3D geometric invariants of short-list candidates are involved to complete the recognition task.

Both a 3D model and a frontal 2D image are acquired to enroll a person, while during testing only a 2D image, different from that used for enrollment, is needed. There are invariants computed from a 2D image of the face and yet other invariants require 3D Point coordinates for verification. During the enrollment of a face, a configuration of non coplanar 3D points on the 3D face model is specified and the Gramian Matrix B (3x3 matrix) is computed and stored (in a file) as a biometric key for the subject. When the identity of a person needs to be identified, the B matrix is loaded from the corresponding files of short-list candidates and the 2D points are located on the 2D test image. Then, the function f_B is used to check if those points of 2D configuration are compatible with the matrix B. If the face image belong to the same person the value of the function f_B is 0 (in an ideal case) or very small in practical applications, otherwise a higher value for f_B is obtained which points out that the face image does not match with the identity represented by the matrix B. The same applies for function f_b.

A. 2D Facial Features Extraction

For 2D facial images, two types of facial features have been extracted namely geometric and textural features as shown in Fig. 1.

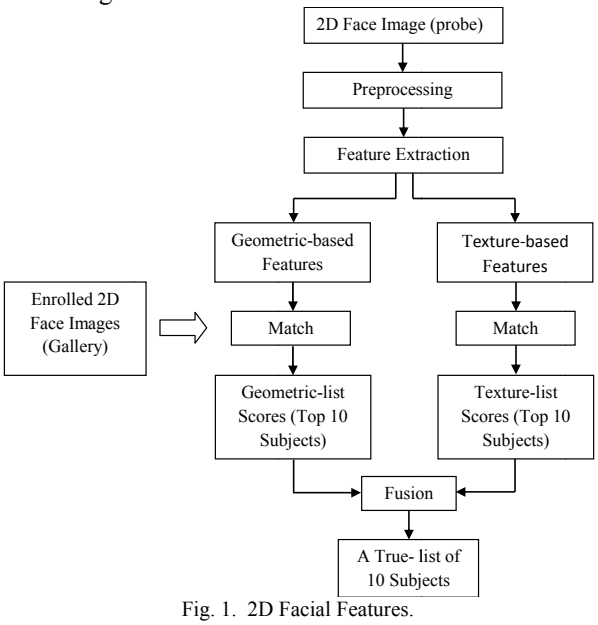

Fig. 1. 2D Facial Features.

B. Facial landmarks Selection

Bosphorus database [8] has been selected as a standard facial database to carry out the experiments in which 22 facial points are marked manually on both 2D and 3D facial images as indicated in Fig. 2. These points are used for geometric calculations; on the one hand they are a subset of the feature points defined in the MPEG-4 standard [9], and on the other hand, they label the most dominant facial features.

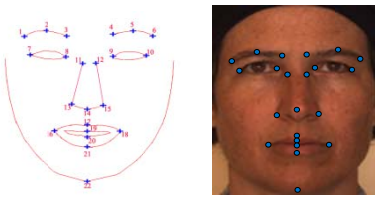

Fig. 2. 22 Facial Landmarks of Bosphorus Database

C. 2D Geometric Invariants (Cross ratios)

Among all geometric invariants, the cross ratio is considered the simplest and the most popular projective invariant. Due to its perspective invariance, the cross ratio is used widely in computer vision particularly for object recognition [10]. Geometric invariants like cross ratios are applied to face recognition to minimize most of the disadvantages of the appearance and hyper based approaches of face recognition.

The cross ratio can be defined as follows. Let P_1, P_2, P_3, and P_4 be four collinear points on the Euclidean plane and the Euclidean distance between two points P_i and P_j denotes as $\delta_{i,j}$. One definition of the cross ratio (R) is the following:

$$\Re_1(P_1,P_2,P_3,P_4) = \frac{\delta_{1,3}\delta_{2,4}}{\delta_{1,4}\delta_{2,3}} \tag{1}$$

In general, the distance between two points in the plane with coordinates (x_1, y_1) and (x_2, y_2) can be defined as an Euclidean distance and computed as follows:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

Additionally, The Cross ratio of five coplanar points P_1, P_2, P_3, P_4 and P_5 is defined as follows:

$$v\Re_2(P_1,P_2,P_3,P_4,P_5) = \frac{M(1,2,4).M(1,3,5)}{M(1,2,5).M(1,3,4)} \tag{3}$$

where

$$M(1,2,4) = \begin{vmatrix} x1 & x2 & x4 \\ y1 & y2 & y4 \\ 1 & 1 & 1 \end{vmatrix} = \begin{vmatrix} x1 & x2-x1 & x4-x1 \\ y1 & y2-y1 & y4-y1 \\ 1 & 1 & 0 \end{vmatrix}$$

$$= \begin{vmatrix} x2-x1 & x4-x1 \\ y2-y1 & y4-y1 \end{vmatrix}$$

$$= (y2-y1)(y4-y1)\left(\frac{x2-x1}{y2-y1} - \frac{x4-x1}{y4-y1}\right) \tag{4}$$

For 2D face images, 20 cross ratios of collinear and

coplanar points are selected according to their invariance where those ratios show a small variance.

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N} \quad (5)$$

where μ denotes the mean, X refers to the value of one of the possible cross ratios, and N indicates the number of cross ratios.

At the enrolment stage, these calculations are taken place off-line and stored into a vector called feature vector. Fig. 3 shows the selected 20 cross ratios of collinear and coplanar facial points.

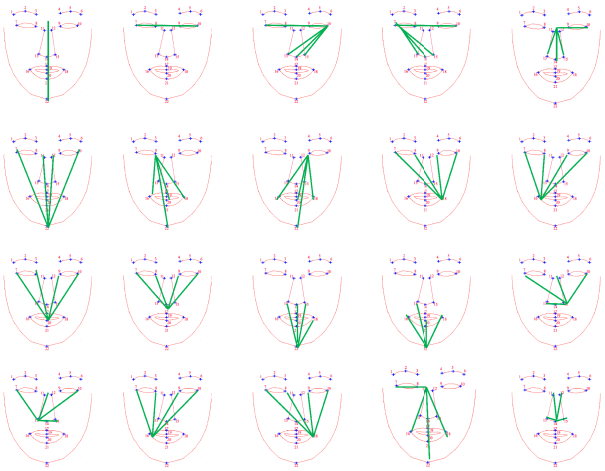

Fig. 3. 20 cross ratios of collinear and coplanar facial control points

D. Facial Textural Representation

Face representation plays the most important role in face recognition process. Hence an efficient representation results in a successful recognition method. Facial texture is one of the key facial properties that can be used for encoding face images for the purpose of discriminating faces. It is common that facial texture is integrating with other facial characteristics such as shape with texture [11] and depth with texture [12], [13] to identify or authenticate faces.

E. Uniform LBP

In this proposed approach, uniform LBP, a commonly used extension of LBPs, is applied to represent 2D facial images together with geometric invariants. In particular, LBP is a simple and popular texture descriptor that summaries the local structure of an image efficiently, therefore it has been adopted by a large range of applications.

Uniform patterns are proved to be dominant in real textures [13] and therefore they are employed herein to represent facial texture. Furthermore, uniform LBP reduces the histogram size of LBP while maintaining the discrimination information [14].

F. K-Nearest Neighbors

The k-Nearest Neighbors (k-NN) is one of the simplest classification algorithms which largely used to address several classification problems based on the closest training samples in feature space [15]. In this part, k-NN with Euclidean distance is used to classify the input face image based on LBP feature vector.

G. 3D Model Invariants

For an object in 3D space, two invariant representations can be defined named affine-invariant and rigid-invariant representations which are invariant to transformations based on basis points. In affine-invariant representation, three independent vectors p_i, p_j, p_k in R_3 represent the basis of an affine coordinate system and each vector can be defined as a combination of the basis $P_l = b_1^l p_i + b_2^l p_j + b_3^l p_k$ where the vector $b^l = (b_1^l, b_2^l, b_3^l)$ represents an affine invariant of the point P_l. In the rigid-invariant representation, the Euclidean metric information on the basis points can be represented by the inverse Gramian matrix ($B = G^{-1}$) and the representation is invariant to rotation. The Gramian of the basis points P_i, P_j, P_k represents a 3 × 3 matrix namely G [16], [7]:

$$G = \begin{pmatrix} P_i^T P_i & P_i^T P_j & P_i^T P_k \\ P_i^T P_j & P_j^T P_j & P_j^T P_k \\ P_i^T P_k & P_j^T P_k & P_k^T P_k \end{pmatrix} \quad (6)$$

Based on four and five 3D control points, two model-based projective invariant functions namely rigid model-based projective invariant function (f_B) and the affine model-based projective invariant function (f_b) respectively can be defined as follows.

In a 3D facial model, given four points of image coordinates (x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3) and assuming $x_0 = 0$, $y_0 = 0$ otherwise $x_1 = (x_1 - x_0)$ and $y_1 = (y_1 - y_0)$. Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$.

$$f_B = \frac{|x^T B y| + |x^T - y^T B y|}{|x||B||y|} \quad (7)$$

where B refers to the inverse Gramian matrix of four points. Fig. 4 shows examples of four control points configuration of 3D face models used to compute B matrix.

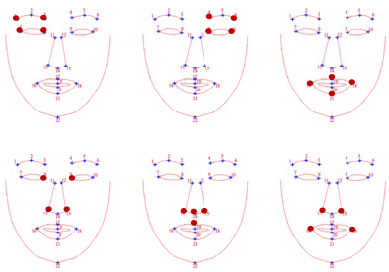

Fig. 4. Examples of four points configuration on 3D face models

The other function namely affine model-based projective invariant function is defined with five non-coplanar 3D points of image coordinates (x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) and assuming $x_0 = 0$, $y_0 = 0$ otherwise $x_1 = (x_1 - x_0)$ and $y_1 = (y_1 - y_0)$. Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$.

$$f_b(x, y) = \frac{|x_4 - \sum_{i=1}^{3} b_i x_i|}{|x||b|} + \frac{|y_4 - \sum_{i=1}^{3} b_i y_i|}{|y||b|} \quad (8)$$

Fig. 5 shows examples of five control point's configuration of 3D face models used to compute b vectors.

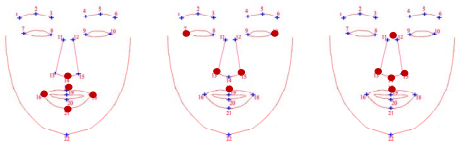

Fig. 5. Examples of five points configuration on 3D face models

For all the views of the object, the value of the functions fB and fb is zero (ideal) or close to zero. The recognition stage can be summarized by the following steps:

1) For each enrolled 3D face model, the coefficients of fB function are computed during the enrollment phase (we compute the Gramian matrix B which returns a 3x3 matrix) and stored as a biometric key for the subject.
2) For each of the true-list subjects, who selected from the whole enrolled facial dataset by the previous stage using geometric-texture features, the corresponding 3D invariants are checked.
3) To check the 3D geometric invariants of candidate subjects, the corresponding B matrix is loaded, then the 2D points on the input images (e.g. the center of the eyes, the tip of the nose, the corner of the mouth) are obtained and normalized by subtracting the coordinates of the first point:

$$xp = [\, x(2)\ x(3)\ x(4)\,] - x(1);\quad yp = [\, y(2)\ y(3)\ y(4)\,] - y(1);$$

4) Finally, the function fB and fb are checked to make the decision
5) If the value of these functions is zero or close to zero, the 2D image is considered as a view of the model represented by the B matrix. In other words, the 2D test image belongs to the subject represented by the B matrix we have computed in the enrollment phase, otherwise the image is not identified.

H. Datasets

These experiments have been performed on landmarked Bosphorus database. Bosphorus database is a 2D-3D face database with a variety of facial expressions and facial occlusions. Furthermore, Bosphorus database is rich with head pose variations which is a concern of this work.

IV. Experimentations

Experimental results obtained using Bosphorus database are encouraging. Comparing with traditional 2D face recognition methods, the proposed asymmetric face recognition method provides better performance; while compared with 3D shape based ones, it reduces the high online cost and the inconvenience of data acquisition and computation.

A. Results

As the main purpose of the combination 2D and 3D modalities is to overcome the pose variation problem in face recognition systems. Several experiments with different in-plan and out-of-plan facial images are carried out to demonstrate the robustness of the proposed approach. Table 2 shows the recognition rate using 2D as probe and 3D as gallery with frontal facial images. Based to the result of the model-based projective invariant functions *fB* and *fb*, the query image can be decided either identified where this result is zero or close to zero (i.e. ≤ 0.03 according to observation of the results) otherwise not identified. Fig. 7 illustrates the performance of the recognition of frontal images.

TABLE II: The Recognition Rate with Frontal Facial Images

No of images	2D-2D Matching	2D-3D Matching
50	87	90
100	83	88
200	83	85
300	80	85

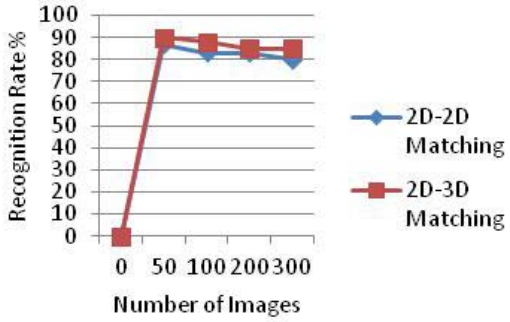

Fig. 7. The performance of the Recognition of Frontal Images

With respect to head pose, we have taken into consideration in-plan rotations with 30 degree left, 30 degree right, 45 degree left, 45 degree right, 60 degree left, and 60 degree right. Table 3 shows the performance of 2D-3D face recognition with pose variations using 300 subjects of Bosphorus database.

TABLE III: The Performance of 2D-3D Face Recognition with Pose Variations

In-plan View	Recognition Rate
30 degree left/right view	91.2
45 degree left/right view	86
60 degree left/right view	82.4

V. Conclusion and Future Work

In this paper, we investigated the advantage of integrating 2D and 3D modalities to increase the performance of face recognition with pose variations. First, 2D geometric invariants and texture information are combined to shorten the candidates list namely the true-list. Then, the B matrix of corresponding 3D models is loaded which computed and stored during the enrollment phase. Afterwards, the 2D points on the input image are applied to the model-based projective invariant functions and the decision is made depend on the result of these functions (i.e. zero or closed to zero there a match, otherwise no match). Experimental results show that the proposed approach can enhance the recognition rate with the presence of pose variations. Other geometric invariants such conic invariants and angles are shifted to be investigated in future work and occluded faces need further investigation as well.

References

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 4, 2007, pp. 1885- 1906.

[2] S. Z. Li and A. K. Jain, *Handbook of face recognition*, (2nd ed.), 2011, Springer.
[3] M. Hüsken, M. Brauckmann, S. Gehlen, and C. Von D. Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),* 2005, pp. 174–174.
[4] X. Lu, A.K Jain, and D. Colbry, "Matching 2.5D Scans to 3D Models," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, 2006, pp. 31-43.
[5] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, 2006, pp.1-15.
[6] H. Tang, Y.Sun, B. Yin, and Y. Ge, "Mixed 2D-3D information for face recognition," *Transactions on Edutainment V,* LNCS 6530, pp. 251–258, Berlin Heidelberg, Germany: Springer-Verlag, 2001.
[7] R. Riccio, and J. Dugelay, "Geometric invariants for 2D/3D face recognition," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1907–1914, 2007.
[8] A. Savran, N. Aly¨uz, H. Dibeklio˘glu, O. C¸eliktutan, B. G¨okberk, L. Akarun, and B. Sankur, "Bosphorus database for 3D face analysis," in *Proc. of the First European Workshop on Biometrics and Identity Management Workshop (BioID*, 2008.
[9] F. Lavagetto and R. Ockaj, "The facial animation engine: Toward a high-level interface for the design of MPEG-4 compliant animated faces," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 9, no. 2, pp.277-289, 1999.
[10] K. Kanatani, "Computational cross ratio for computer vision," in *Proc. of the IPPR Conference on Computer Vision*, *Graphics*, *and Image Processing: Image Understanding,* pp. 371-381, 1994.
[11] C. Liu and H. Wechsler, "A shape-and texture-based enhanced Fisher classifier for face recognition," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 598-608, 2001.
[12] C. B. Kader and P. Griffin, "Comparing and combining depth and texture cues for face recognition," *Image and Vision Computing*, vol. 23, pp. 339–352, 2005.
[13] P. Xiong, L. Huan, and C. Liu, "Real-time 3D face recognition with the integration of depth and intensity images," *Image Analysis and Recognition*, Berlin Heidelberg, Germany: Springer-Verlag, 2011.
[14] F. Bianconi and A. Fernandez, "On the occurrence probability of local binary patterns: a theoretical study," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 3, pp. 259-268, 2011.
[15] N. Bhatia and Vandana, "Survey of nearest neighbor techniques," *International Journal of Computer Science and Information Security*, vol. 8, no. 2, 2010.
[16] D. Weinshall, "Model-based invariants for 3-D vision," *International Journal of Computer Vision*, vol. 10, no. 1, pp. 207-231, 1993.

Gawed M. Nagi received his B.Sc. in computer science from Baghdad University, IRAQ and his Masters degree from New Mexico State University (NMSU), USA in 1996 and 2002 respectively. Currently he is a PH.D. student in the faculty of Computer Science and Information Technology at University Putra Malaysia (UPM). His research interests include Image processing, machine vision and Computer Graphics.

Rahmita O. K. Rahmat received the B.Sc. and M.Sc. degrees in Science Mathematics from University Science Malaysia (USM), in 1989 and 1994, respectively. During 1989 to 1990 she work as a research assistant in department of physics in University Science Malaysia experimenting on Ozone layer measurement at the Equatorial region, before working as tutor in Universiti Putra Malaysa (UPM).

She received her PhD in Computer Assisted Engineering from University of Leeds, U.K. At this moment she is working in Faculty of Computer Science and Information Technology as lecturer. Among her focus research area are Computer Graphics and Applications, Computer Assisted Surgery and Computational Geometry.

Fatimah Khalid obtained the B. Sc. in Computer Science from University Technology Malaysia (UTM) in 1992. During 1993 to 1995 she works as a System Analyst at University Kebangsaan Malaysia (UKM) and continued her Masters Degree from UKM. After getting her Masters Degree in 1997, she started involved in teaching at Sal College until 1999 and continued at the University Putra Malaysia in June 1999. She received her PhD in System Science and Management from the National University of Malaysia in 2008. At this moment, she is working in Faculty of Computer Science and Information Technology as a lecturer. Among her focus research area are Computer Vision and Image Processing, Content based Retrieval System and Computer Graphic Applications.

Muhamad T. Abdullah is a lecturer in the Department of Multimedia, Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM). He obtained his first degree in Computer Science from UPM in 1990, masters degree in Computer Science from Universiti Teknologi Malaysia (UTM) in 1992, and doctor of philosophy from UPM in 2006. He joined UPM in 1990 as a tutor.

He became a lecturer at Department of Computer Science in 1992. His research interest includes Information Retrieval, and Cross Language Information Retrieval.