International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Detection of Human Head Direction Based on Facial Normal Algorithm

Lam Thanh Hien
Vice-Rector of Lac Hong University
email: lthien@lhu.edu.vn

Do Nang Toan
Associate professor in Computer Science of Vietnam Academy of Science and Technology

Tran Van Lang
Associate Professor in Computer Science of Vietnam Academy of Science and Technology

Abstract: **Many scholars worldwide have paid special efforts in searching for advance approaches to efficiently estimate human head direction which has been successfully applied in numerous applications such as human-computer interaction, teleconferencing, virtual reality, and 3D audio rendering. However, one of the existing shortcomings in the current literature is the violation of some ideal assumptions in practice. Hence, this paper aims at proposing a novel algorithm based on the normal of human face to recognize human head direction by optimizing a 3D face model combined with the facial normal model. In our experiments, a computational program was also developed based on the proposed algorithm and integrated with the surveillance system to alert the driver drowsiness. The program intakes data from either video or webcam, and then automatically identify the critical points of facial features based on the analysis of major components on the faces; and it keeps monitoring the slant angle of the head closely and makes alarming signal whenever the driver dozes off. From our empirical experiments, we found that our proposed algorithm effectively works in real-time basis and provides highly accurate results.**

Keywords: **Head Direction, Facial Normal Model, Novel Algorithm, Head Detection, Face Model**

1. INTRODUCTION

Human head direction, an indicator of the human visual focus of attention, has numerous applications in our daily life, such as such as human-computer interaction, teleconferencing, virtual reality, and 3D audio rendering [1-3]. Smith et al. [4], Benfold & Reid [5], and Chamveha et al., [6] claimed that attention focus of humans can be inferred from head direction. Therefore, several scholars have paid special efforts in proposing advance approaches to estimate head direction [6].

For instance, Murase & Nayar [7] recognized the face orientation in a certain space by using parametric Eigen face model which is based on Principal Component Analysis and each pixel is treated as a random variable. However, it requires a large sample size which takes time to collect the data and analyze them. Wang & Brandstein [8] successfully classified existing methods for head pose estimation from video signals into feature-based and appearance-based approaches. Particularly, the feature-based approaches proposed by Ballard & Stockman [9], Horprasert et al. [10], and Matsumoto & Zelinsky [11] mainly consider some specific facial features, such as eyes, nostrils, and mouth, in constructing a head direction model. However, Canton-Ferrer et al. [12] claimed that these approaches fail to deal with multi-ocular analysis of face of head images; and, their performance of these approaches significantly depends on the environment lighting conditions, the camera angles, the face orientation towards cameras, etc. Hence, the feature-based methods fail to be implemented in some practical applications because they are quite sensitive to the selection of the points of the facial features and need near-frontal views with high resolution images [2, 13, 14] or need special equipment such as depth cameras [15,16]. On the contrary, appearance-based approaches by Voit et al. [17] and Zhang et al. [18] directly use pixel values of an image as an input for image features; hence, they are successfully employed to deal with low resolution images [6].

Nonetheless, Chamveha et al. [6] pointed that the dataset used for training estimators in the employed neural networks significantly affects the performance of the appearance-based methods. As head appearances keep incessantly changed due to the head orientation, illumination or viewing angle, collecting ground-truth dataset for the training becomes crucial for their successful applications, which is a labor-intensive, time-consuming, and costly task. Moreover, the use of head orientation estimation as an input leads to the certain limitation of output angle resolution [2]; for instance, Rae & Ritter [19] suggested using a step of 25^o for the angle estimation.

Instead, Hien et al. [20] proposed a human face model based on spatial approach to detect human face and its components, including eyes, mouth, and nose. Their approach was successfully applied to alarm drowsy drivers by detecting if their eyes are continuously closed in predetermined duration measured in seconds or frames. In order to do that, a camera is mounted in front of the driver at a fixed angle and the driver sits at a fixed position and looks straight ahead. However, in practice, this assumption is hardly satisfied because drivers sometimes adjust their positions, look around, shake/ incline their heads, etc. to make them more comfortable. In these cases, the model fails to correctly detect the eyes to keep track if the drivers are drowsy. To overcome this shortcoming, this paper aims at proposing a novel algorithm based on the normal of human face to recognize human head direction which acts as an important input in their human face model.

To achieve the objective, this paper is organized as the following. Section II briefly presents the facial normal model, which provides fundamentals for our proposed algorithm in detecting the slanting angle of heads constructed in Section III. Experimental results are elucidated in Section IV. Some conclusions make up the last section.

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

2. Facial Normal Model

The facial normal model proposed by Gee & Cipolla [21] includes five facial features, including two far corners of eyes, two points of mouth corners, and the tip of nose. They assumed that the four points of eyes and mouth corners make up a plane called facial plane. The normal of facial plane at the nose tip is called facial normal as shown in Fig. 1.

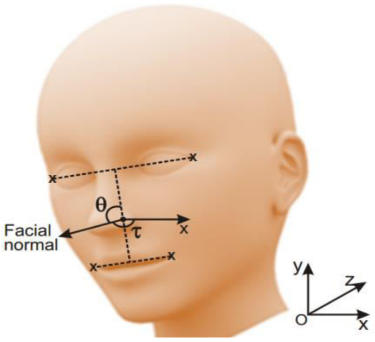

Fig.1. Facial Normal Model

Assume that a coordinate system *Oxyz* is located at the center of the camera as shown in Fig. 1, where *Ox* and *Oy* axes are aligned along the horizontal and vertical directions in the image, and *Oz* axis is aligned along the normal to the image plane. From the two points of far corners of eyes and two points of far corners of mouth, we can easily find their midpoints which are then joined up to create the symmetric axis of the facial plane. To estimate the direction of facial normal in 3D space, the model needs two predetermined ratios, namely as $R_m = L_m / L_f$ and $R_n = L_n / L_f$ where L_m, L_n, and L_f are accordingly measured as plotted in Fig. 2.

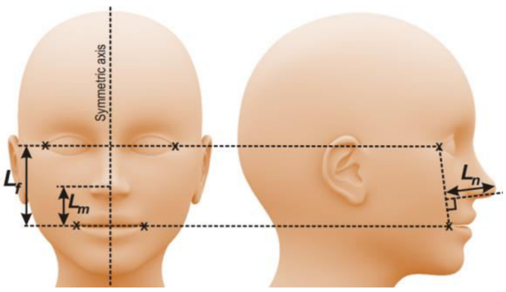

Fig.2. Fundamental Parameters L_m, L_n, and L_f

Because length ratios along the symmetric axis are preserved, with the model ratio R_m, we can easily locate the nose base along the axis. Then, the facial normal in the image is determined by joining the nose base and the nose tip. As a consequence, the tilt direction can be easily obtained by measuring the angle τ between the facial normal in the image and the *Ox* axis. Moreover, we also need to measure the slant angle σ which is defined as the angle between the optical axis and the facial normal in 3D space. Basically, the slant angle σ can be computed from the model ratio R_n and the angle θ from the image [21].

Thus, in the coordinate system *Oxyz*, the facial normal $\hat{n}$ is determined by

$$\hat{n} = \left[\sin\sigma\cos\tau, \sin\sigma\sin\tau, -\cos\sigma\right]. \tag{1}$$

3. Our Proposed Algorithm

As reviewed in Section II, two model ratios R_m and $R_n$are prerequisite in identifying the facial normal based on the approach by Gee & Cipolla [21]. However, different faces result in different values of the ratios. Thus, the approach fails to effectively perform in some practical cases. To overcome the existing drawback, we propose a novel approach by optimizing a simple face model in combination with the aforementioned facial normal model.

A. Mapping the Model to Image Coordinates Based on Facial Normal

- Face Model and Transformational Operations

Our 3D face model also includes five facial features as discussed in the facial normal model by Gee &Cipolla [21]. Our face model is also located on the coordinate system *Oxyz* where *Oxy* plane is the facial plane; the center O is located at the nose base; and the nose tip is on the *Oz* axis. Particularly, the five critical feature points namely as are plotted in Fig. 3.

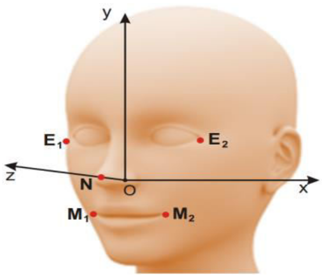

Fig.3. Proposed Face Model

Our face model is constructed based on five transformational operations $P = < P_1, P_2, P_3, P_4, P_5 >$, where

$$P_1(x): \begin{cases} E_1.x = E_1.x * x \\ E_2.x = E_2.x * x \end{cases} \qquad P_2(x): \begin{cases} E_1.y = E_1.y * x \\ E_2.y = E_2.y * x \end{cases}$$

$$P_3(x): \begin{cases} M_1.y = M_1.y * x \\ M_2.y = M_2.y * x \end{cases} \qquad P_4(x): \begin{cases} M_1.x = M_1.x * x \\ M_2.x = M_2.x * x \end{cases}$$

$$P_5(x): N.z = N.z * x$$

- Calculation of Normalized Coordinates

The next step is to rotate the face model based on the obtained facial normal. Let $n^1 = (n_1^1, n_2^1, n_3^1)$ and $n^2 = (n_1^2, n_2^2, n_3^2)$ be respectively the normal of the original face model and the calculated facial normal. Then, the normalization of the model coordinates is conducted based on the following steps:

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Step 1: Identify rotating axis

$$n = (n_2^1 * n_3^2 - n_3^1 * n_2^2, n_3^1 * n_1^2 - n_1^1 * n_3^2, n_1^1 * n_2^2 - n_2^1 * n_1^2) \quad (2)$$

Step 2: Normalize the axis vector

$$u = (u_1, u_2, u_3) = \frac{1}{\|n\|} * n \quad (3)$$

Step 3: Calculate the rotating angle

$$\alpha = \arccos\left(\frac{n^1 \cdot n^2}{\|n^1\| * \|n^2\|}\right) \quad (4)$$

Step 4: Calculate the rotating matrix

$$Q = \begin{pmatrix} q1 & q2 & q3 \\ q4 & q5 & q6 \\ q7 & q8 & q9 \end{pmatrix} \quad (5)$$

where:

$q1 = \cos\alpha + u_1 * u_1 * (1 - \cos\alpha)$

$q2 = u_1 * u_2 * (1 - \cos\alpha) - u_3 * \sin\alpha$

$q3 = u_1 * u_3 * (1 - \cos\alpha) + u_2 * \sin\alpha$

$q4 = u_2 * u_1 * (1 - \cos\alpha) + u_3 * \sin\alpha$

$q5 = \cos\alpha + u_2 * u_2 * (1 - \cos\alpha)$

$q6 = u_2 * u_3 * (1 - \cos\alpha) - u_1 * \sin\alpha$

$q7 = u_3 * u_1 * (1 - \cos\alpha) - u_2 * \sin\alpha$

$q8 = u_3 * u_2 * (1 - \cos\alpha) + u_1 * \sin\alpha$

$q9 = \cos\alpha + u_3 * u_3 * (1 - \cos\alpha)$

Any point $A = (A_x, A_y, A_z)$ in the face model can be then transformed into its new coordinate determined by $A_{new} = A \times Q$. Consequently, the model coordinates on projection plane can be efficiently normalized by shifting them to the coordinate center and converting the length of the coordinate vector into unit.

- Calculate the Coordinate Solution

Based on the five critical facial features and the normalization process mentioned above, the coordinate solution is computed by

$$y_0 = normalize(\{E_1^0.x, E_1^0.y, E_2^0.x, E_2^0.y, M_1^0.x, M_1^0.y, M_2^0.x, M_2^0.y, N^0.x, N^0.y\}) \quad (6)$$

From the above steps, we have a mapping $y = f(x), R^5 \to R^{10}$ where $x = (p_1, p_2, p_3, p_4, p_5)$ is the key parameter for the above five transformational operations; thus, the normalized coordinate of the model after appropriate projection is determined by

$$y = f(x) = (x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_5, y_5) \quad (7)$$

$$y = normalize(E_1.x, E_1.y, E_2.x, E_2.y, M_1.x, M_1.y, M_2.x, M_2.y, N.x, N.y) \quad (8)$$

B. Optimizing the Face Model

With the above constructed projection, we now need to identify the value of x such that $f(x) = y_0$. To solve this problem, we propose a novel algorithm to optimize the parameters in the above five transformational operations so that their results well fit the coordinate solutions constructed above. With a set of input parameters, the algorithm will minimize the squared sum of each deviation between coordinate solution (y_0) and the normalized coordinate*(Y)*, which can be expressed as

$$E = \sum_{i=1}^{10} [y(i) - y_0(i)]^2 \to \min, \quad (9)$$

where $y(i)$ and $y_0(i)$ are the i^{th} component in the normalized vector $y = f(x)$; and the coordinate solution vector y_0, respectively.

Our proposed process for the optimization is departed from Cootes et al. [22]. Particularly, let $r(x) = f(x) - y_0$; the objective function (9) can be now rewritten as $E = r^T r$. With the first degree Taylor disposition, we have

$$r(x + \partial x) = r(x) + \frac{\partial r}{\partial x}\partial x \quad (10)$$

Thus,

$$\partial x = Rr(x) = -\left(\left(\frac{\partial r}{\partial x}^T \frac{\partial r}{\partial x}\right)^{-1} \frac{\partial r}{\partial x}^T\right) r(x) \quad (11)$$

The Jacobi matrix $\frac{\partial r}{\partial x}$ is estimated from a set of samples. Specifically, each component in the matrix is estimated by computing a large amount of the functional deviation against its relevant argumentative deviation which is first specified in a given range. Thus, the proposed algorithm is designed as the following.

- Inputs: Face model and the image coordinates of the five points of the critical facial features.
- Outputs: Facial normal vector.
- Some prior calculation:
  - ✓ $R = -\left(\left(\frac{\partial r}{\partial x}^T \frac{\partial r}{\partial x}\right)^{-1} \frac{\partial r}{\partial x}^T\right)$
  - ✓ y_0
  - ✓ Initial values for x
  - ✓ Initial parameter series $K = \{1, 0.5, 0.25, 0.125, 0.0625\}$
- Iterative loops:

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 1, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

- ✓ Calculate the deviation vector $r(x) = f(x) - y_0$
- ✓ Calculate $E = r^T r$
- ✓ Calculate $\partial x = Rr(x)$
- ✓ For each value of $k \in K$, calculate
  - $x' = x + k * \partial x$
  - $r'(x) = f(x') - y_0$
  - $E' = r'^T r'$

If $E' < E$, replace $x = x'$ and move to the next loop; otherwise, move to the next value of k.

The optimization algorithm proposed by Cootes et al. [22] has been successfully employed on their active appearance model (AAM) whose processing time is adjacent to the real time. However, their approach is built on a quite complicated mapping function and depends on the size of the dataset. Meanwhile, our proposed algorithm works with a simpler mapping function which takes a constant time frame due to the fixed size of the dataset.

5. Experimental Results

We now conduct practical experiments to test the quality of the proposed algorithm based on the data obtained from 3D face model with different projection parameters. The 3D face models are first transformed based on some predetermined model parameters defined previously. Then, they are rotated and evaluated at different rotating angle as shown in Fig. 4.

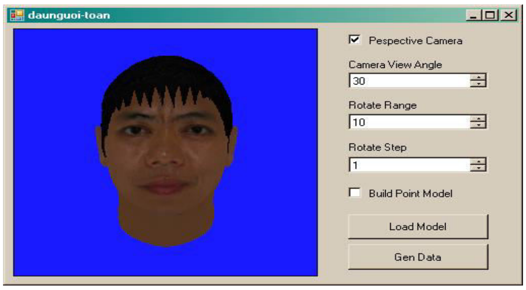

Fig.4. Interface ForObtaining Data From 3D Face Model

Based on our obtained experimental results, we found that the deviated angle between the calculated normal and the original normal is approximately 16.35^o with the maximum value of 41.11^o and minimum value of 5.17^o, and more than 93% of the observations have the deviated angle less than 20^o.

Moreover, from these positive primary results, we develop a computational program to recognize the head direction of driver and integrate with the system for detecting driver drowsiness proposed by Hien et al. [21] to make effective alarming signals when the driver is drowsy. The program is written in Visual C++2008 with the support from the open source computer vision library provided by Intel.

The program intakes data from either video or webcam, and then automatically identify the critical points of facial features based on the analysis of major components on the faces. It keeps monitoring the slant angle of the head closely and makes alarming signal whenever the driver dozes off as shown in Fig. 5 and Fig. 6.

Practically, we conducted our experiments in Lac Hong University with 11 observations and Institute of Information Technology – Vietnam Academy of Science and Technology with 26 observations in September 2014. The input data for our experiments were collected from a video camera. Basically, among the 11 observations taken in lac Hong University, we had 10 correct estimations compared to our perception, accounting for 90.9% of the sample; and we had 24 correct estimations out of the 26 observations (92.3%) taken from the Institute. More importantly, we found that our computational program effectively works in real-time basis and provides highly accurate results based on the specified conditions in our experiments. Certain errors also exist in some cases where the program fails to correctly identify the critical points of the facial features.

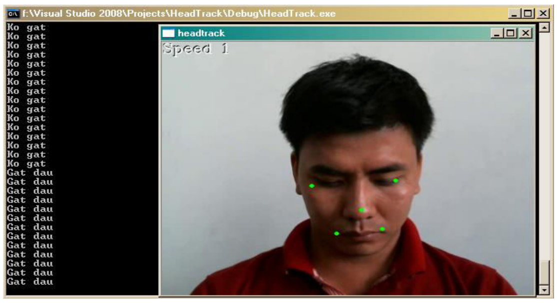

Fig.5. Experimental Results Provided by our Computational Program

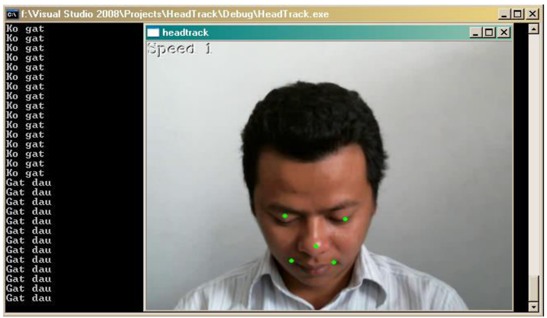

Fig.6. Experimental Results Provided by our Computational Program

6. Conclusion

Estimating human head direction has been an interesting research topic in the digital signals and image processing with numerous practical applications in recognizing human faces, monitoring human activities, human-machine interactions, etc. However, some problems are still left unsolved, especially the optimizing the system performances. Hence, departing from a well-known facial normal model, we proposed a novel algorithm to efficiently estimate human head direction by optimizing a simple face model incorporated with the facial normal model to overcome the avoidable differences in facial features among various faces of different people. This obviously fulfills the current literature in terms of efficiency in recognizing human head direction. Future researches should further consider our proposed algorithm in combination with other systems to provide advance digital solutions in practice.

ACKNOWLEDGMENT

This paper was conducted with the special supports from Project VAST01.04/13-14 funded by Vietnam Academy of Science and Technology and Project B2014-TN07-01 funded by Thai Nguyen University.

REFERENCES

[1] L. Zhao, G. Pingali, and I. Carlbom. (2002). Real-time head orientation estimation using neural networks. *Proc. 2002 Int. Conf. Image Process.*, 1, pp. I-297 – I-300.

[2] C. Canton-Ferrer, J. R. Casas, and M. Pardas. (2008). Head Orientation estimation using particle filtering in multiview scenarios. *Lect. Notes Comput. Sci.* 4625, pp. 317-327.

[3] C. Wang, S. Griebel, and M. Brandstein. (2000). Robust automatic video-conferencing with multiple cameras and microphones. *Proc. IEEE Int. Conf. Multimedia Expo.* 3, pp. 1585–1588.

[4] K. Smith, S. O. Ba, J. M. Odobez, and D. Gatica-Perez. (2008). Tracking the visual focus of attention for varying number of wandering people. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, pp. 1212-1229.

[5] B. Benfold, and I. Reid. (2009). Guiding visual surveillance by tracking human attention. *Proc. 20th British Machine Vision Conf. (BMVC2009)*, pp. 14.1-14.11.

[6] I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Ohabe, Y. Sta, and A. Sugimoto. (2013). Head direction estimation from low resolution images with scene adaptation. *Comput. Vision Image Understanding.* 117, pp. 1502-1511.

[7] H. Murase, and S. K. Nayar (1995). Illumination planning for objects recognition using parametric eigenfaces. *IEEE Trans. Pattern Anal. Machine Intell.* 16 (12), pp. 1219-1227.

[8] C. Wang, and M. Brandstein. (2000). Robust head pose estimation by machine learning. *Proc. IEEE Int. Conf. Image Process.* 3, pp. 210–213.

[9] P. Ballard, and G.C. Stockman. (1995). Controlling a computer via facial aspect. *IEEE Trans. Syst. Man, Cybern.* 25(4), pp. 669–677.

[10] T. Horprasert, Y. Yacoob, L. S. Davis. (1996). Computing 3-D head orientation from a monocular image sequence. *Proc. Int. Conf. Autom. Face Gesture Recognit.* pp. 242–247.

[11] Y. Matsumoto, and A. Zelinsky. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.* pp. 499–504.

[12] C. Canton-Ferrer, J. R. Casas, and M. Pardàs. (2005). Fusion of multiple viewpoint information towards 3D face robust orientation detection. *Proc. IEEE Int. Conf. Image Process.* 2, pp. 366–369.

[13] A. Puri, H. Kannan, and P. Kalra. (2012). Coarse head pose estimation using image abstraction. *Proc. 9th Conf. Comput. Robot Vision (CRV2012).*

[14] D. Huang, M. Storer, F. D. L. Torre, and H. Bischof. (2011). Supervised local subspace learning for continuous head pose estimation. *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR2011).* pp. 2921-2928.

[15] G. Fanelli, T. Weise, J. Gall, and L. V. Gool. (2011). Real time head pose estimation from consumer depth cameras. *Proc. 33rd Annu. Symp. German Assoc. Pattern Recognit. (DAGM2011).* pp. 101–110.

[16] G. Fanelli, T. Weise, J. Gall, and L. V. Gool. (2011). Real time head pose estimation with random regression forests. *Proc. IEEE Conf. Computer Vision Pattern Recognit. (CVPR2011).* pp. 617–624.

[17] M. Voit, K. Nickel, and R. Stiefelhagen. (2007). Neural Network-based head pose estimation and multi-view fusion. *Lect. Notes Comput. Sci.* 4122, pp. 291–298.

[18] Z. Zhang, Y. Hu, M. Liu, and T. Huang. (2007). Head pose estimation in seminar rooms using multi view face detectors. *Lect. Notes Comput. Sci.* 4122, pp. 299–304.

[19] R. Rae, and H.J. Ritter. (1998). Recognition of human head orientation based on artificial neural networks. *IEEE Trans. Neural Networks*. 9, pp. 257–265.

[20] L. T. Hien, T. V. Lang, H. M. Toan, and D. N. Toan. (2012). Modeling the human face and its application for detection of driver drowsiness. *Int. J. Comput. Sci. Telecommun.* 3(11), pp. 56-59.

[21] A. Gee, and R. Cipolla. (1994). Determining the gaze of faces in images. *Image Vision Comput.* 12(10), pp. 639-647.

[22] T. F. Cootes, G. J. Edwards, and C. J. Taylor. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Machine Intell.* 23(6), 681-685.

AUTHOR'S PROFILE

Lam Thanh Hien received his MSc. Degree in Applied Informatics Technology in 2004 from INNOTECH Institute, France. He is currently working as a Vice-Rector of Lac Hong University. His main research interests are Information System and Image Processing.
email: lthien@lhu.edu.vn

Do Nang Toan is an Associate professor in Computer Science of Vietnam Academy of Science and Technology. He received BSc. Degree in Applied Mathematics and Informatics in 1990 from Hanoi University and PhD in Computer Science in 2001 from Vietnam Academy of Science and Technology. He is currently working as a Head of Department of Virtual reality technology at Institute of Information Technology, Vietnamese Academy of Science and Technology and as Dean of Faculty of Multimedia Communications, Thai Nguyen University of Information and Communication Technology. His main research interests are Pattern recognition, Image processing and Virtual reality.

Tran Van Lang is an Associate Professor in Computer Science of Vietnam Academy of Science and Technology. He received BSc. Degree in Mathematics in 1982 and PhD in Mathematics – Physics in 1995 from HCM City University of Natural Science, Vietnam. He also worked at Dorodnitsyn Computing Center, Russian Academy of Science. He is currently working as Associate Professor in Computer Science at a research institute of Vietnam Academy of Science and Technology; and as Dean of Faculty of Information Technology, Lac Hong University. His main research interests are High Performance Parallel and Distributed Computing, Bioinformatics, Scientific Computation and Fuzzy Computing Methods. He is IEEE Member with No. Number is 90579642.