International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 4, April 2015
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 107-113

Detecting Human Face with Ridge-Valley-Normal Model

Lam Thanh Hien[1], Do Nang Toan[2] and Ha Manh Toan[3]

[1] Lac Hong University, Dongnai, Vietnam
[2] Information Technology Institute, VNU
[3] Institute of Information Technology, Vietnamese Academy of Science and Technology

lthien@lhu.edu.vn

ABSTRACT

Detecting human face has been a popular but complex research topic which has successfully attracted the special attention of academic scholars and experts due to its widespread applications in practice. However, several existing methods fail to work with large-slanting angles of faces. To overcome such drawback, this study further develops the application of the ridge & valley characteristics on human face and incorporates them with the well-known facial normal model to effectively detect human face. Our proposed algorithm consists of two consecutive phases namely "learning phase" and "detecting phase". It is found effective in the experimental test with a data set with numerous images collected from different sources. Moreover, the novel searching procedure not only results in a new set of faces but also displays an examined face corresponding to which face model, from which the actual direction of human face detected can be easily calculated.

Keywords: *Human Face, Ridge Valley Characteristics, Facial Normal, Slant Angle, Facial Features.*

1. INTRODUCTION

Detecting human face has been a popular research topic that provides numerous useful results with high quality and speed. Its complicacy and practical applications have successfully attracted the special attention of academic scholars and experts; for example, the detection of human face direction has been widely used in human-computer interaction, teleconferencing, virtual reality, 3D audio rendering [1-4], driver's drowsiness [5], eye gaze classification [6].

It is quite interesting that the existing methods usually depart from a model. For instance, Murase & Nayar [7] proposed parametric Eigenface model based on Principal Component Analysis to recognize the face direction in a certain space. As their approach treats each pixel as a random variable, a large sample size is needed, i.e. considerable time in collecting and analyzing the data is one of its obvious shortcomings. The face direction models proposed by Ballard & Stockman [8], Horprasert et al. [9], and Matsumoto & Zelinsky [10] consist of some specific facial features, such as eyes, nostrils, and mouth, which fail in dealing with multi-ocular analysis of face of head images [11]. Du et al. [12] proposed a model based on the ridge-valley characteristics of human faces, which provides critical ideas in the facial model proposed in this paper. The human face model proposed by Hien et al. [13] was successfully applied to alarm drowsy drivers by detecting if their eyes are continuously closed in predetermined duration measured in seconds or frames. However, Canton-Ferrer et al. [11] claimed that environment lighting conditions, the camera angles, the face orientation towards cameras significantly affect the performance of the models. We have practically observed that several existing methods fail to work with large-slanting angles of faces as shown in Fig. 1.

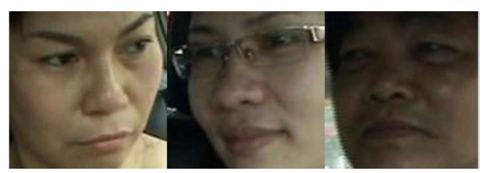


Fig. 1. Examples of slanting faces

To overcome such drawback, this paper aims at proposing a human face model based on the ridge-valley characteristics of human faces in combination with the face normal in order to not only improve its performance in detecting slanting faces but also compute their slant angles.

To achieve the objective, this paper is organized as the following. Section 2 briefly presents fundamental concepts employed in this paper while Section 3 provides details about our proposed model with specific algorithms. Experimental results are elucidated in Section 4. Some conclusions make up the last section.

2. SOME PRIMARY CONCEPTS

2.1 Ridge – Valley Characteristics

Normally, in a certain space, "Ridge" consists of the highest points whereas "Valley" consists of the lowest ones. Thus, in a particular image of a face, Du et al. [12] suggested that the nostril and the highest points on cheeks be ridges while the orbits are valleys.
Specifically, let *I(x,y)* denote a mapping function, we can accordingly present its graphs as a plane named $z = I(x,y)$. Thus, in a 3D space with a coordinate system *Oxyz,* the plane z is actually a curved plane represented by a function $G(I) = \{[x, y, I(x, y)] \in R^3\}$. Mathematically, the ridges and valleys locate at certain points where the derivatives $\frac{\partial^2 I}{\partial x^2}$ and $\frac{\partial^2 I}{\partial y^2}$ achieve their extremums. However, in practice, computer digitally codes a particular image as an intermittent function; thus, searching for the extremums becomes difficult and leads to the possible loss of information. Hence, Du et al. [12] suggested the two following formulas to lessen the ties for the ridges and valleys.

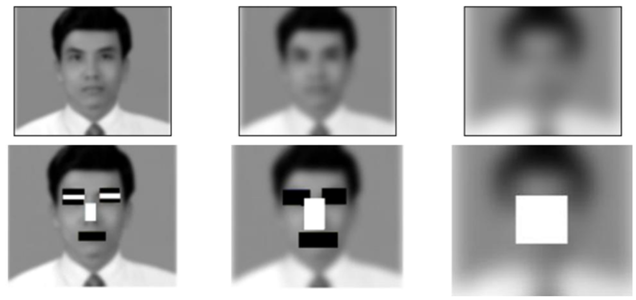
$$R_P = \{(x,y) \mid LoG(x,y) \geq \delta \wedge LoG(x,y) > 0\} \quad (1)$$

$$V_P = \{(x,y) \mid LoG(x,y) \geq \delta \wedge LoG(x,y) < 0\} \quad (2)$$

where R_P and V_P are respectively the sets of ridges and valleys; *LoG(x,y)* is the value of Laplacian-of-Guassian at the point *(x,y)*. The performance of LoG(x,y) is illustrated in Fig. 2 where δ is set at the value of $\sqrt{2}$.

Fig. 2. Ridges and Valleys based on LoG with $\delta = \sqrt{2}$

In order to improve the performance in searching for the ridges and valleys in images, Du et al. [12] proposed searching for them based on different levels (σ). Their approach shows that low levels are suggested to search for ridge and valley characteristics of small features such as eyes, nostril, etc. Higher levels are more appropriate for larger ridges and valleys. At the highest level, the whole face becomes a large ridge area, indicating that higher level performs better for larger features with fewer details. Obviously, the motive of their approach is similar to the fact that we can only recognize the body shape of a person standing far away from us; however, the closer we get to him/her, the better we identify the details on his/her face. As shown in Fig. 3, the white rectangular shapes represent the ridges detected whereas the black ones represent the valleys detected. Practically, though more ridges and valleys can be detected on an image, we only consider those on the frame containing human face.



a) After smoothed at $\sigma = 10$ b) After smoothed at $\sigma = 20$ c) After smoothed at $\sigma = 60$

Fig. 3. Searching for Ridges and Valleys at different level

2.2 Construction of Ridge – Valley Structure

Based on the obtained ridges and valleys, a structure of ordinal relationships is accordingly constructed. The areas extracted at the highest level σ_n become the root nodes in the structure whereas areas extracted at the lowest level σ_1 added to the structure are considered as the leaves and others are considered as branches.
The key task in the construction is to connect obtained areas extracted at certain level to other ones at the next level. If more than 2/3 of area *R* at level σ_i is concealed by another area *R'* at level σ_{i+1}, *R* is then considered as the branch of *R'*. With this procedure, we can easily construct a full structure from the extracted areas. Fig. 4 demonstrates the structure constructed from the extracted areas obtained from Fig. 3.

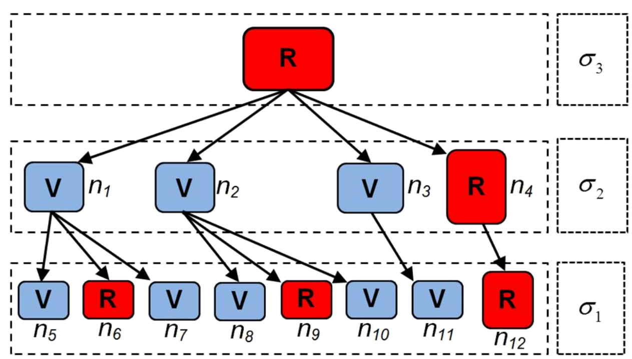


Fig. 4. A typical structure constructed from Fig. 3.

Each of the nodes on the structure contains a lot of important information about its specific description, such as position, size, its characteristic (ridge/valley), the chart of grey distribution, etc. The information, crucial for the further development of the model, can be structurally presented in Table 1. Particularly, the tags of the nodes are shown in the first column named “Node”; the examined levels are respectively shown in column “Level”; the column “Type of Node” is a binary field about the type of each node, including only two options: Ridge and Valley; the chart of grey distribution of relevant node considered as is stored in the column “Accompanied Information”; the coordinate and size of each node are shown in column “Coordinate & Size” while the last column “Master Node” shows the node directly relating to another one at its adjacent lower level with a convention that relevant cells of nodes at the level next to the highest level receive a null value.

Table 1: Important information about nodes in a node space

Node	***Level***	***Type of Node***	***Acc. Inf.***	***Coordinate & Size***	***Master Node***
n_1	σ_2	V/Sq	I_1V	$n_1(x_1, y_1, w_1, h_1)$	
n_2	σ_2	V/Sq	I_2V	$n_2(x_2, y_2, w_2, h_2)$	
n_3	σ_2	R/Rv	I_3R	$n_3(x_3, y_3, w_3, h_3)$	
n_4	σ_2	V/Rh	I_4V	$n_4(x_4, y_4, w_4, h_4)$	
n_5	σ_1	V/Rh	I_5V	$n_5(x_5, y_5, w_5, h_5)$	n_1
n_6	σ_1	R/Rh	I_6R	$n_6(x_6, y_6, w_6, h_6)$	n_1
n_7	σ_1	V/Rh	I_7V	$n_7(x_7, y_7, w_7, h_7)$	n_1
n_8	σ_1	V/Rh	I_8V	$n_8(x_8, y_8, w_8, h_8)$	n_2
n_9	σ_1	R/Rh	I_9R	$n_9(x_9, y_9, w_9, h_9)$	n_2
n_{10}	σ_1	V/Rh	$I_{10}V$	$n_{10}(x_{10},y_{10}, w_{10}, h_{10})$	n_2
n_{11}	σ_1	V/Rh	$I_{11}V$	$n_{11}(x_{11},y_{11}, w_{11}, h_{11})$	n_3
n_{12}	σ_1	R/Rv	$I_{12}R$	$n_{12}(x_{12},y_{12}, w_{12}, h_{12})$	n_4

2.3 Face Direction and Its Determination

The facial normal model proposed by Gee & Cipolla [14] 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. 5 [5].

Assume that a coordinate system *Oxyz* is located at the center of the camera as shown in Fig. 5, 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. 6 [5].

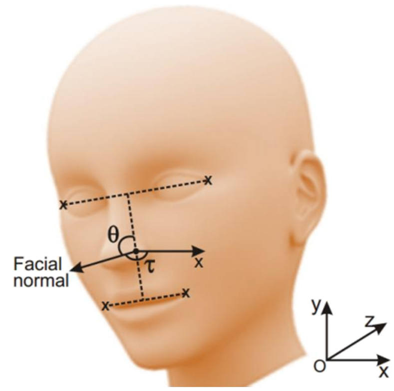


Fig. 5. Facial Normal Model

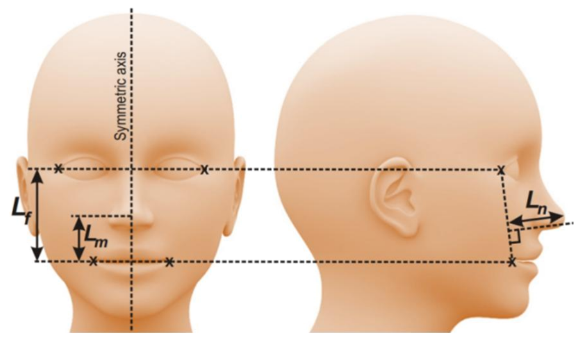


Fig. 6. Fundamental parameters L_m, L_n, and L_f

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

$$\hat{n} = [\sin\sigma\cos\tau, \sin\sigma\sin\tau, -\cos\sigma]. \quad (3)$$

3. DETECTION OF HUMAN FACE WITH RIDGE-VALLEY-NORMAL MODEL

3.1 Ridge – Valley – Normal Model (RVNM)

Based on the data set presented in Table 1, we establish a net frame in the form of scalar graphs similar to the data structure in 3D graphic. Specifically, nodes at the highest level are first added into the graph and connected with scalar sides in order to create a fundamental frame as shown in Fig. 7. The size of the fundamental frame depends on the Euclide distance among the nodes.

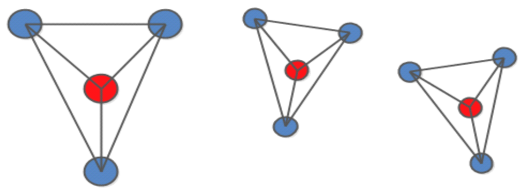


Fig. 7. Fundamental frames of nodes at the highest level

Nodes at lower levels are then inserted into the fundamental frame and connected with their master nodes. For nodes of the same master node, if a ridge node lies vertically between two valley nodes, we draw extra lines connecting two frontiers of the ridge node with the two valleys as shown in Fig. 8. Obviously, this procedure results in a lozenge located inside either the ridge or valley areas of the master node.

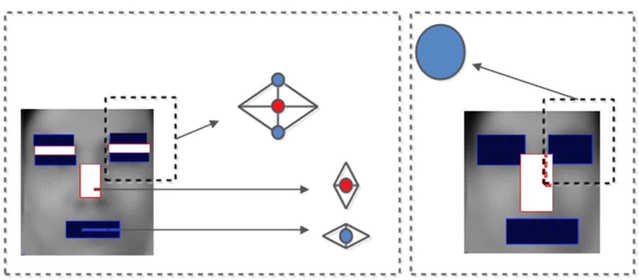


Fig. 8. Nodes created from extracted ridge/valley areas at different levels

After finishing inserting all nodes of the same level, we can erase the master node and its related sides to simplify the net frame as shown in Fig. 9. Consequently, with each ridge and valley area extracted at the highest level, a net frame is constructed with different height indicating the key characteristic of the node which is either ridge or valley.

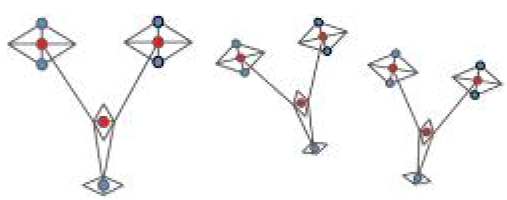


Fig. 9. Typical net frames

The above mentioned procedure results in a net frame where the position and key characteristics of the nodes on the net relate closely to the eyes, nostril, and mouth on a face. As discussed in Eq. (3), the face direction strictly relates to the position of the eyes, nostril, and mouth; thus, it also closely relates to the nodes on the net frame. This model of frame with the ties between the face direction and the nodes shown on the graphs is now called ridge-valley-normal model (RVNM).

3.2 Proposed Algorithm to Detect Human Face Based on Ridge-Valley-Normal Model

This section presents a novel algorithm to detect human face based on the RVNM discussed in Section 3.1. Particularly, the detection of human face based on the RVNM includes two typical phases. The first one is called "learning phase" which contains a collection of face images and a collection of non-face images to create a data set with a series of models of typical faces. The second one is called "detecting phase" which examines if there is any area in the image meeting the characteristics of any model in the data set. Thus, our proposed algorithm is developed as the following.

3.2.1 Establishing Data Set

We need to have a collection of face images and a collection of non-face images to create a data set with a series of models of typical faces. Each of the face is modeled with some crucial information about its facial normal and positions of the facial features including eyes, nostril, and mouth.

To have the series, we propose the following procedure.

- From the collections of input images $I = \{i_1, i_2, ..., i_n\}$, identify ridge & valley characteristics at different levels in order to establish a net frame. Hence, each of the input images is accompanied with a set of graphs $G = \{g_1, g_2, ..., g_n\}$.
- Cluster the set of graphs G based on the Euclide distance between graph g_i and its respective angle into K clusters $\left(K = \{G_1^*, G_2^*, ..., G_m^*\}\right)$. Each cluster G_j^* contains a set of net frame graphs g_k.
- For each $G_j^* \in K$, compute its expectation and variance. Besides, by employing Radial Basis Function (RBF) cited in [15], we can also establish a relationship between the nodes and its expected angle based on the characteristics of facial features determined in each input image. Specifically, RBF helps to identify the relationship between facial features and their relevant nodes in the net frame; and the angle is calculated by Eq. (3).
- Consequently, each G_j^* is accompanied with a set of $r_j = \langle \mu_j, \delta_j, \Gamma_j \rangle$, where μ_j is its expectation; δ_j is its variance; and, Γ_j is the relationship between the angles and their related nodes in the expected net

frame graphs. A set $R=\{r_1,r_2,...,r_m\}$ is exactly our expected data set with a series of models of typical faces.

The above procedure is programmed with the following algorithm.

Input: $I=\{i_1,i_2,...,i_n\}$ *// Collection of input images*

Output: $R=\{r_1,r_2,...,r_m\}$

```
Begin
    G={}, K={};
    For all x ∈ I
        rv = CalculateRidgeValley(x);
        tree = BuildTree(rv);
        G = G ∪ BuildNetFrameGraph(tree);
    EndFor
    Cluster(G, K);
    For all y ∈ K
        μ = CalculateMean(y);
        δ = CalculateVariance(y);
        Γ = Build RBF Relation(μ, y, δ);
        R = R ∪ <μ, δ, Γ>;
    EndFor
End
```

3.2.2 Detecting Human Face

With the established data set with a series of models of typical faces, we now propose another algorithm to detect human face. This task can be briefly summarized in the following steps.

- From the ith input image, we extract its ridge/valley characteristic at different levels (the number of levels depends on the size of the examined area and searching steps).
- From the obtained characteristics, we construct a structure of ordinal relationship T such that each node t on the T satisfies specified height in setting up net frame graphs of corresponding g.
- Search in the data set for a model that has a minimum distance between its expectation and g. Three possible cases are: (1) if the distance is less than a threshold $\alpha*\delta$, we conclude that it is a face; (2) if the distance is greater than a threshold $\beta*\delta$, we conclude that it is not a face; (3) otherwise, we need to interpolate the angle of net frame graphs and check if the angle corresponds to the critical angle of the model; if it does, we conclude that it is a face; if it fails to correspond, we conclude that it is not a face.

Our proposed algorithm to calculate face area corresponding to the nodes in the structure which is supposed to be a face is programmed as the following.

Input: $R=\{r_1,r_2,...,r_m\}$*, I, msize, step*

Output: $Rc=\{rc_0,rc_1,...,rc_m\}$

```
Begin
    IsFace = {};
    rv = CalculateRidgeValley(x);
    tree = BuildTree(rv)
    For all node ∈ tree
    If (HeightNode(node)>=HeightOfElementOn(R)
        g = BuildNetFrameGraph(node );
        m = FindNearestModel(g, R);
        If (Distance(g, m. μ)< α*m.δ)
            IsFace = IsFace ∪ g;
            Elseif(Distance(g, m. μ)< β *m.δ)
        Beign
            r = CalculateRotation(g, m);
            If (r > m.MinAngle and r<m.MaxAngle)
            IsFace = IsFace ∪ g;
            EndIf
        End
        Endif
    EndIf
    Endfor
    For all face ∈ IsFace
    Rc = Rc ∪ CalculateRect(face);
    EndFor
End
```

4. Empirical Experiments

To test the performance of our proposed algorithm in the learning phase, we employ a set of face images collected by Hancock [16] and another set collected by ourselves during this study. Totally, we have more than five thousand images of different people at different slant angles and different emotion statuses. We also establish a set of non-face images and different landscapes collected from online sources. Fig. 10 shows some typical faces used in the learning phase.

Fig. 10. Some typical faces used in the learning phase

To test the performance of our proposed algorithm in the detecting phase, we employ the set of faces collected by Weber [17] with some typical faces shown in Fig. 11 and another set provided by the Vision and Autonomous Systems Center of Carnegie Mellon University (CMU/VASC) [18] with some typical examples shown in

Fig. 12 as well as our collection with large slant angles as typically shown in Fig. 13. The set contains more than one thousand images taken in different conditions.

a) Taken indoors *b) Taken outdoors*
c) Lack of front light *d) Photo of emotional face*
e) Photo of drawing *f) Photo of shaded face*

Fig. 11. Some typical images taken from Weber's collection [17]

Fig. 12. Some typical images in CMU/VASC [18]

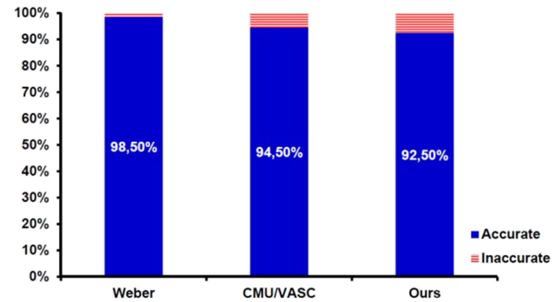
Fig. 13. Some typical images in our collection

The above data of images are tested in the detecting phase whose results are presented in Table 2 and visually plotted in Fig. 14.

Table 2: Experimental results

Collection	*No. of images*	*Correct*	*Incorrect*	*Failed*	*Accuracy*
Weber	450	446	1	3	98.5%
CMU/VASC	180	173	2	5	94.5%
Ours	500	467	4	17	92.5%

Table 2 and Fig. 14 indicate that our proposed algorithm provides very high detection power in the standard sets of Weber and CMU/VASC. Moreover, though our collection has many faces taken from large slant angles, the detection accuracy of 92.5% is considered satisfactory. These numerical figures demonstrate the good performance of



our proposed algorithm in detecting human face with Ridge-Valley-Normal Model.

Fig. 14. Accuracy levels of proposed algorithm in different collections

5. Conclusions

This study further develops the application of the ridge & valley characteristics on human face and incorporates them with the well-known facial normal model to effectively detect human face. Our proposed algorithm consists of two consecutive phases namely "learning phase" and "detecting phase". It was tested with a data set with numerous images collected from different sources. Its experimental results obviously validate its effective performance. Moreover, the searching procedure not only results in a new set of faces but also displays an examined face corresponding to which face model, from which we can promptly calculate the actual direction of human face detected.

REFERENCES

[1] Y. Sankai, K. Suzuki, and Y. Hasegawa, "Cybernics: Fusion of human, machine and information systems", Japan: Springer, 2014.

[2]L. Zhao, G. Pingali, and I. Carlbom, "Real-time head orientation estimation using neural networks", in 2002 International Conference on Image Processing, 2002, Vol. 1, pp. I-297 – I-300.

[3]C. Canton-Ferrer, J. R. Casas, and M. Pardas, "Head Orientation estimation using particle filtering in multiview scenarios", Lecture Notes in Computer Science, Vol. 4625, 2008, pp. 317-327.

[4]C. Wang, S. Griebel, and M. Brandstein, "Robust automatic video-conferencing with multiple cameras and microphones", IEEE International Conference on Multimedia and Expo, 2000, Vol. 3, pp. 1585–1588.

[5]L. T. Hien, D. N. Toan, and T. V. Lang, "Detection of Human Head Direction Based on Facial Normal Algorithm", International Journal of Electronics Communication and Computer Engineering, Vol. 6, No. 1, 2015, pp. 110-114.
[6]A. Al-Rahayfeh, and M. Faezipour, "Application of head flexion detection for enhancing eye gaze direction classification", Conference Proceeding of IEEE Engineering in Medicine & Biology Society, August 2014, pp.966-969.
[7]H. Murase, and S. K. Nayar, "Illumination planning for objects recognition using parametric eigenfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 12, 1995, pp. 1219-1227.
[8]P. Ballard, and G.C. Stockman, "Controlling a computer via facial aspect". IEEE Transactions on Systems, Man, and Cybernetics, Vol. 25, No. 4, 2005, pp. 669–677.
[9]T. Horprasert, Y. Yacoob, L. S. Davis, "Computing 3-D head orientation from a monocular image sequence", IEEE International Conference on Automatic Face and Gesture Recognition, 1996, pp. 242–247.
[10]Y. Matsumoto, and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement", IEEE International Conference on Automatic Face and Gesture Recognition, 1996, pp. 499–504.
[11]C. Canton-Ferrer, J. R. Casas, and M. Pardàs, "Fusion of multiple viewpoint information towards 3D face robust orientation detection", IEEE International Conference on Image Processing, 2005, Vol. 2, pp. 366–369.
[12]T. L. H. Du, D. A. Duc, and D. N. Vu, "Ridge and Valley based Face Detection", IEEE International Conference on Computer Sciences Research, Innovation, Vision for the Future (RIVF'06) - Ho-Chi-Minh city, Vietnam, 2006.
[13]L. T. Hien, T. V. Lang, H. M. Toan, and D. N. Toan, "Modeling the human face and its application for detection of driver drowsiness", International Journal of Computer Science and Telecommunications, Vol. 3, No. 11, 2012, pp. 56-59.
[14]A. Gee, and R. Cipolla, "Determining the gaze of faces in images", Image and Vision Computing, Vol. 12, No. 10, 1994, pp. 639-647.
[15]X. Wan, "Geodesic Distance-Based Realistic Facial Animation Using RBF Interpolation", Journal Computing in Science & Engineering, Vol. 14, No. 5, 2012, pp. 49-55.
[16]http://pics.stir.ac.uk/
[17]http:/home.eps.hw.ac.uk/~sm217/html/datasets.hml
[18]http://vasc.ri.cmu.edu/idb/

AUTHOR PROFILES:

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

Do Nang Toan is an Associate professor in Computer Science of VNU (Vietnam National University). 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 Associate Professor in Computer Science at a research institute of VNU 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.

Ha Manh Toan received the BSc. Degree in Applied Mathematics and Informatics in 2009 from College of Science, Vietnam National University, Hanoi. He is currently working as a researcher at Institute of Information Technology, Vietnamese Academy of Science and Technology. His main research interests are Image Processing, Computer Vision.