Effective Detection of Human Face with GentleBoost Approach

Lam Thanh Hien†, Do Nang Toan††, Ha Manh Toan†††, Thanh-Lam Nguyen††††
(L.T. Hien – Corresponding author)
† Office of Academic Affairs, Lac Hong University, Vietnam
†† Graduate University of Science and Technology, Vietnam National University, Vietnam
††† Institute of Information Technology, Vietnamese Academy of Science and Technology
†††† Office of Scientific Research, Lac Hong University, Vietnam

Summary
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 in real time or with low detection rate, or large-slanting angles of faces, etc. To overcome such drawback, this study proposes an innovative approach with GentleBoost to effectively detect human faces in images. Our proposed algorithm with two consecutive phases namely "learning phase" and "detecting phase" has shown its effectiveness. Specifically, its performance has been well validated through hundreds of images collected from reliable databases and self-recorded sources. It is found that though the detection rate from our approach is lower than that of traditional Haar-AdaBoost, ours still provides satisfactory results in terms of precision and recall. More importantly, it is about 6 times faster than that of traditional Haar-AdaBoost, promising a great potential to be integrated into practical applications that need to detect human faces in real time.

Key words:
Detect human face, GentleBoost, Haar-AdaBoost, Real time detection, Precision, Recall

1. Introduction

Detecting human face has been a popular research topic in the field of informatics and image processing because it can provide numerous useful results for practical applications; such as human-computer interaction, teleconferencing, virtual reality, 3D audio rendering [1-4], driver's drowsiness [5], eye gaze classification [6]. However, it is also a complicated problem that has well attracted the special attention of academic scholars and experts worldwide. Thus, several detection methods have been proposed and continuously improved over the past few years as reviewed by Setu & Rahman [7], Shuka et al. [8], Malik et al. [9], Naik & Lad [10], Chavan & Bharate [11], Kakade [12], Solanki & Pittalia [13], Sharma & Kaur [14], Kiran et al. [15], Lin et al. [16] and so on.

For example, Murase & Nayar [17] considered Principal Component Analysis (PCA) in a parametric Eigenface model based on to recognize the face and its direction in a certain space. Because each pixel is treated as a random variable, they need a large sample size which is actually a critical shortcoming as this takes considerable time in collecting and analyzing the data. Or, Ballard & Stockman [18], Horprasert et al. [19], and Matsumoto & Zelinsky [20] investigated some specific facial features, such as eyes, nostrils, and mouth, which fail in dealing with multi-ocular analysis of face of head images [21]. Du et al. [22] proposed a model based on the ridge-valley characteristics of human faces while the human face model proposed by Hien et al. [23] 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. [21] claimed that environment lighting conditions, the camera angles, the face orientation towards cameras significantly affect the performance of the models. Literally, several existing methods fail to work with large-slanting angles of faces as shown in Fig. 1.

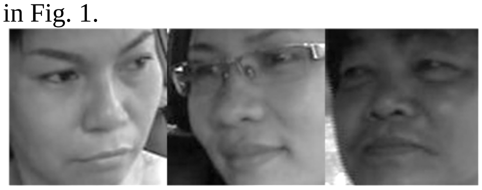

Fig. 1. Examples of slanting faces

Among the numerous approaches, Setu & Rahman [7] claimed that Viola & Jones [24]'s face detector (VJFD) is the first ever face detection framework to effectively work in real time because it contains three main components, including: integral image, classifier learning with Adaboost, and intentional cascade structure [25]. Literally, based on the traditional analyses of the facial features, the region of each facial component like left eye, right eye, nose, etc., can be easily determined; thus, a face can be also detected if those components respectively identified. Specifically, Schneiderman & Kanade [26] used a variable function to extract facial features for their machine learning process based on AdaBoost to detect human face. Whereas, Viola & Jones [24] used AdaBoost algorithm in vertical combination with the Haar-like features to effectively detect human face.

Fundamentally, the key of AdaBoost is to combine weak classifiers into a stronger one; where (1) "A weak

Manuscript received December 5, 2016
Manuscript revised December 20, 2016

classifier” is referred to as a mathematical algorithm that can provide a correct classification rate of more than 50% and a hypothesis resulted from a weak classifier is called “weak hypothesis”, denoted by $h_m(x)$, and (2) “A strong classifier”, denoted by $H(x)$, is obtained by a linear combination of M weak classifiers, i.e. $H_M(x) = sign\sum_{m=1}^{M}\alpha_m h_m(x).$ x is classified based on a function $H(x) = sign[H_M(x)]$, where the value of $|H_M(x)|$ indicates the reliability level. Specifically, consider a problem with 2 layers with a training sample of M classifiers labeled (x_i, y_i) for $(i = \overline{1, M})$ where $y_i = \pm 1$ is called “label” and $x_i \in R_n$ is called “training sample”. Consequently, a linear combination of M weak classifiers $h_m(x)$ results in a strong classifier $H_M(x)$, i.e. $H_M(x) = \sum_{m=1}^{M} h_m(x).$

The construction of $h_m(x)$ via AdaBoost algorithm is developed as the following:

Assume $H_{M-1}(x) = \sum_{m=1}^{M-1} h_m(x)$; $H_M(x)$ is considered as the best classifier if $H_M(x) = H_{M-1}(x) + h_m(x)$ leads to a minimum value of $H_m = \arg\min \sum_{m=1}^{M} e^{-y_i H_M(x_i) h_M(x_i)}$. Then, a function for the minimum value is determined by:

$$h_M(x) = \frac{1}{2}\log\frac{P\left(y = +1|_{x,w^{M-1}}\right)}{P\left(y = -1|_{x,w^{M-1}}\right)}.$$

where w^{M-1} is the weight of the classifier at *M*.

Viola & Jones [24] claimed that their approach provides a fast speed with a correct detection rate of more than 80%. However, it may result in high false detection [7]; therefore, a great number of remedy solutions have been proposed, such as using pre-filtering or post-filtering methods based skin color filter to provide complementary information in color images. Wu & Ai [27] and Tabatabaie et al. [28] claimed that using a skin color as a pre-filtering stage can improve the performance of VJFD in reducing the false detection. Or, Niazi & Jafari [29] pointed that using skin color in post-filtering HSV color space can also significantly reduce false positive detection in the VJFD. To reduce the effects of lighting, Erdem et al. [30] applied an illumination compensation algorithm in the first step before combining VJFD and the skin color detector to detect face. Wang & Abdel-Dayem [31] proposed an algorithm for face detection based on edge information and hue. However, the results were not accurate for all type of images [7]. To overcome such drawbacks, this paper aims at proposing a human face model based on an innovative and simpler approach by deploying GentleBoost method.

To achieve the objective, this paper is organized as the following. Section 2 provides clear details about our proposed approach with specific algorithms. Experimental results are elucidated in Section 3. Some conclusions make up the last section.

2. Our Proposed Approach

Our approach focuses on the binary classification of each image region of interest to detect if there is a face in the region. The consideration for the decision is done with a series of binary classifiers, and a detected image region is validated if all of the classifiers in the series are fully satisfied. The binary classifiers are constructed based on decision trees where each knot is actually a sub binary classifier. Then, the detection of a human face can be done with 2 major phases: (1) Learning phase, creating a typical database of face images for training and learning from a collection of images with or without faces; (2) Detecting phase, matching the target image against the database to decide if any face is detected in the target image.

Particularly, a decision tree can be constructed based on the training database in the following structure:

$$\{(I_s, v_s, w_s) : s = 1, 2, ..., S\}$$

where, v_s is the correct label of image I_s and w_s is its weight accordingly.

Therefore, it is mandatory to classify and label the images with either +1 or -1. In addition, the weight w_s allows us to indicate the importance level of each input sample in the training database. Each knot on a decision tree is constructed based on the selection of best binary classifier for the database, meaning that the following objective function must achieve its minimum value:

$$WMSE = \sum_{(I,v,w)\in C_0} w.(v - \overline{v}_0)^2 + \sum_{(I,v,w)\in C_1} w.(v - \overline{v}_1)^2$$

where: C_0 and C_1 are respectively the training groups based on the binary classification results of 0 and 1; $\overline{v}_0$ and $\overline{v}_1$ are the means of label values in C_0 and C_1, respectively.

Consequently, from the initial database, the construction of decision tree results in two separate categories in each knot during the learning phase. The learning algorithm can be programmed as the following:

Input: $U = \{(I_s, v_s, w_s) : s = 1, 2, ..., S\}$

Output: $T = \{N_0, N_1, ...,\}$

```
begin
        T:= Ø
        Idx0 = {0, 1, 2,..,S-1}
        Stack:= Ø
        push(Stack, {N0, Idx0});
```

```
          while (Stack ≠ Ø)
          {Ni,Idxi}:= pop(Stack);
        if (Ni.level>=MAX_DEPTH)
          Continue;
        else
          min_err:= MAX_VALUE;
          best_bincls:= null;
          for all bincls of BCS
            e:= WMSE(bincls, U, Idxi);
             if (e< minerr)
               best_bincls:= bincls;
               min_err:= e;
             endif
          endfor
    setupNode(Ni, best_bincls, U, Idxi);
          {Idxi*2+1,           Idxi*2+2}:=
SplitDataSet(U, Idxi, best_bincls);
          push(Stack, {Ni*2+1, Idxi*2+1});
          push(Stack, {Ni*2+2, Idxi*2+2});
        endif
      endwhile
    end
```

As such, the total learning time for a tree is the sum of total learning time of each level which is actually the sum of learning time of each knot. At each knot, the learning time is calculated by summing up the learning time of each possible sub classifier. In the optimization process in each knot, as the space for the investigated classifiers is quite large, a sub classifier randomly generated is alternatively used to improve the efficiency. This is done by using the GentleBoost approach.
With the above structure for a decision tree, a sub binary classifier is designed as the following.

2.1 Comparison of Pixel Intensity

A comparison of pixel intensity on image I is defined as:

$$B(I:l_1,l_2)=\begin{cases}0 & if\ I(l_1)<I(l_2)\\ 1 & otherwise\end{cases}$$

where $I(l_i)$ is the intensity value of the image I at position l_i. In this technique, the positions of l_1 and l_2 are determined in a space of $[-1,+1]\times[-1,+1]$ as this will make the collection of positions on the image totally independent from the sample image. To elucidate this comparison approach, let's consider an example as shown in Figure 2.

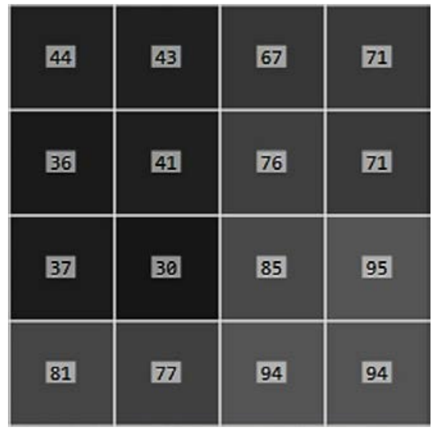

Figure 2. An example of pixel intensity

The image is a grey scale, i.e. the value of any pixel intensity is in [0,255]. The comparison of the pixel intensity of this image is shown in Table 1.

Table 1. Results of pixel intensity comparison

l_1	l_2	$I(l_1)$	$I(l_2)$	$B(I:l_1,l_2)$
(0, 0)	(1, 0)	44	43	1
(0, 0)	(2, 0)	44	67	0
(0, 0)	(0, 1)	44	36	1
(0, 0)	(1, 1)	44	41	1
(0, 0)	(2, 1)	44	76	0
(2, 0)	(3, 0)	67	71	0
(2, 1)	(3, 1)	76	71	1
(1, 2)	(2, 3)	30	94	0
(0, 3)	(3, 1)	81	71	1
(1, 1)	(3, 3)	41	94	0

It could be said that the comparison of pixel intensity is one of the easiest approach for a classifier because it works without predefined parameters. Such comparison is simpler than Haar-like extraction approach because it doesn't need to have integral images. More importantly, our proposed approach can effectively work with different slanting angles of the objects because we only need to integrate a transformation operation on the 2-dimensional space for the 2 positions to be compared.
At each knot, a set of sub comparison operations is investigated. Two random positions on the image are normally generated in a space of $[-1,+1]\times[-1,+1]$. If the decision tree has D levels, and we need to have B comparisons at each knot for the training set of S samples, the total time for training the decision tree is $O(D{\cdot}B{\cdot}S)$.

2.2 Usage of partial means

$$B\left(I : R(x,y,w,h), d\right) = \begin{cases} 1 & if \ \frac{1}{w \times h} \sum_{x \le i < x+w} \sum_{y \le j < y+h} I(x,y) < \delta, \\ 0 & otherwise \end{cases}$$

where *I(i,j)* is the value of pixel intensity of image *I* at position *(i,j)*. The classification results are based on the comparison of means of all pixels in the rectangular region *R(x,y,w,h)* against the threshold δ.

The calculation of partial means actually needs to take all values in the region. However, by using the integral image technique, we can confirm that the total calculation time is *O(1)*. The partial means are compared against a threshold parameter δ which is determined during the learning phase. Basically, the threshold parameters are respectively the marginal positions calculated from the samples. Similarly, if a decision tree has *D* levels and we need *B* comparison operations at each knot for *S* samples, the total training time for the tree is determined by $O(D \cdot B \cdot S^2)$.

3. Experiments and Results

In our experiments, we trained our detection program (learning phase) with (1) 3,500 images in the human face database GENKI-SZSL in the MPLab GENKI owned by California University, San Diego [32]; and (2) 3,019 negative images provided by OpenCV HaarTraining. The whole training was uninterruptedly done on a desktop computer Core i7- 3.6 GHz, RAM 8GB for 263 minutes.

Then, in the detecting phase, we used 450 face images in the database by Markus [33] at California Institute of Technology. These images were taken under different conditions in terms of light, emotional expressions on the faces, and background. Each image has a resolution of 896x592 in JPEG format. In addition, another 398 images taken in different situations, such as drivers, workers in a production line, students in class, players on a recreational area, etc., were also used as self-recorded samples to further evaluate the performance. These self-recorded images were stored under PNG format to avoid the loss of their information.

The detection performance of our proposed approach is compared against that of the Haar-AdaBoost algorithm provided by OpenCV; specifically, we used the sample haarcascade_frontalface_alt_tree with their default parameters. The comparison results under the two investigated sets are shown in Table 2.

Table 2. Detection Effectiveness of Haar-AdaBoost & our proposed approach

Database	***Indicators***	***Haar-AdaBoost***	***Proposed Approach***
Markus [31]	Number of undetectable images	7/450	27/450
	Number of false detection	16	7
	Average process time *(seconds)*	0.109713	0.018982
Self-recorded	Number of undetectable images	6/398	22/398
	Number of false detection	13	6
	Average process time *(seconds)*	0.108244	0.018805

Among the 450 investigated images in the database of Markus [33], our proposed approach can effectively detect 423 images containing human faces. By manually checking the 423 images, we found that there were 7 pieces containing no faces (false detection), meaning that our approach can correctly detect 416 among the 423 images. Consequently, the performance of our proposed approach can be evaluated by:

Precision = 416/423 = 0.9834 (or 98.34%)

Recall = 416/450 = 0.9244 (or 92.44%).

And among the 398 self-recorded images, our proposed approach can effectively detect 376 images containing human faces, as shown in Figure 3. Among the 376 images, there were 6 false detections, meaning that our approach can correctly detect 370 among the 398 images. Consequently, the performance of our proposed approach can be evaluated by:

Precision = 370/376 = 0.9840 (or 98.40%)

Recall = 370/398 = 0.9296 (or 92.96%).

From the two experiments, the high values of the precision and recall indicate that our approach can provide satisfactory detection rate though Haar-AdaBoost actually outperforms ours in having a higher detection rate. Especially, our proposed approach can run about 5.78 times faster than Haar-AdaBoost; thus, it can be used in the development of practical applications that need to detect human faces in real time.

4. Conclusion

This study proposes an innovative approach with GentleBoost to effectively detect human faces in images. Our proposed algorithm consists of two consecutive phases namely “learning phase” and “detecting phase”. Its performance has been well validated through hundreds of images collected from reliable databases and self-recorded sources. Though the detection rate from our approach is lower than that of Haar-AdaBoost, it still provides satisfactory results in terms of precision and recall. More importantly, it is about 6 times faster, which makes the proposed approach greatly potential to be integrated into practical applications that need to detect human faces in real time. Therefore, our proposed approach fulfills the existing gaps in the current literature of detecting human face in real time and real world applications.

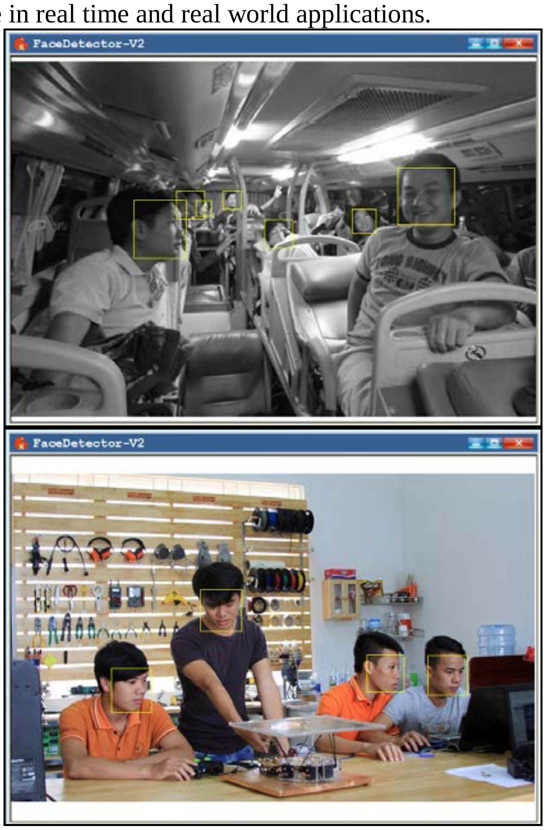

Figure 3. Detection examples with self-recorded samples

Acknowledgements

This research was supported under the project B2015-TN06-01.

References

[1] Y. Sankai, et al., “Cybernics: Fusion of human, machine and information systems,” Japan: Springer, 2014.
[2] L. Zhao, et al., “Real-time head orientation estimation using neural networks,” In 2002 International Conference on Image Processing, vol. 1, pp. I-297 – I-300, 2002.
[3] C. Canton-Ferrer, et al., “Head orientation estimation using particle filtering in multiview scenarios,” Lecture Notes in Computer Science, vol. 4625, pp. 317-327, 2008.
[4] C. Wang, et al., “Robust automatic video-conferencing with multiple cameras and microphones,” IEEE International Conference on Multimedia and Expo, vol. 3, pp. 1585–1588, 2000.
[5] L.T. Hien, et al., “Detection of human head direction based on facial normal algorithm,” International Journal of Electronics Communication and Computer Engineering, vol. 6, pp. 110-114, 2015.
[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, pp. 966-969, August 2014.
[7] T.A. Setu and M.M. Rahman, “Human face detection and segmentation of facial feature region,” Global Journal of Computer Science and Technology: G Interdisciplinary, vol. 16, pp. 1-8, 2016.
[8] S. Shukla, et al., “Review of face recognition technology using feature fusion vector,” International Journal of Advanced Engineering Research and Science, vol. 3, pp. 19-22, 2016.
[9] V. Malik, et al., “A comprehensive study of face detection and recognition,” International Journal of Latest Trends in Engineering and Technology, vol. 6, pp. 41-47, 2016.
[10] R.K. Naik and K.B. Lad, “A review on side-view face recognition methods,” International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, 2984-2991, 2016.
[11] V.D. Chavan and A.A. Bharate, “A review paper on face detection and recognition in video,” International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 4, pp. 97-100, 2016.
[12] S.D. Kakade, “A review paper on face recognition techniques,” International Journal for Research in Engineering Application & Management, vol. 2, pp. 1-4, 2016.
[13] K. Solanki and P. Pittalia, “Review of face recognition techniques,” International Journal of Computer Applications, vol. 133, pp. 20-24, 2016.
[14] N. Sharma and R. Kaur, “Review of face recognition techniques,” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, pp. 29-37, 2016.
[15] P. Kiran, et al., “Human Machine Interface Based on Eye Wink Detection,” International Journal of Informatics and Communication Technology, vol. 2, no. 2, pp.116-123, 2013.
[16] Z. Lin, et al., “Efficient Train Driver Drowsiness Detection on Machine Vision Algorithms,” TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 11, no. 5, pp. 2566-2573, 2013.
[17] H. Murase and S.K. Nayar, “Illumination planning for objects recognition using parametric eigenfaces,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, pp. 1219-1227, 1995.
[18] P. Ballard and G.C. Stockman, “Controlling a computer via facial aspect,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 25, pp. 669–677, 2005.
[19] T. Horprasert, et al., “Computing 3-D head orientation from a monocular image sequence,” IEEE International Conference on Automatic Face and Gesture Recognition, pp. 242–247, 1996.

[20] 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, pp. 499–504, 1996.

[21] C. Canton-Ferrer, et al., “Fusion of multiple viewpoint information towards 3D face robust orientation detection,” IEEE International Conference on Image Processing, vol. 2, pp. 366–369, 2005.

[22] T.L.H. Du, et al., “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.

[23] L.T. Hien, et al., “Modeling the human face and its application for detection of driver drowsiness,” International Journal of Computer Science and Telecommunications, vol. 3, pp. 56-59, 2012.

[24] P. Viola and M.J. Jones, “Robust Real-Time Face Detection,” International Journal of Computer Vision, vol. 57, pp. 137–154, 2004.

[25] C. Zhang and Z. Zhang, “Boosting-based face detection and adaptation,” Synthesis Lectures on Computer Vision, vol. 2, pp. 1-140, 2010.

[26] H. Schneiderman and T. Kanade, “Probabilistic modeling of local appearance and spatial relationships for object detection,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 40-50, 1998.

[27] Y.W. Wu and X.Y. Ai, “Face detection in color images using Adaboost algorithm based on skin color information,” Proceedings of the First International Workshop on Knowledge Discovery and Data Mining, IEEE Computer Society Washington, DC, USA, pp. 339-342, 2008.

[28] Z.S. Tabatabaie, et al., “A hybrid face detection system using combination of appearance-based and feature-based methods,” IJCSNS International Journal of Computer Science and Network Security, vol. 9, pp. 181-185, 2009.

[29] M. Niazi and S. Jafari, “Hybrid face detection in color images,” IJCSI International Journal of Computer Sciences Issues, vol. 7, pp. 367–373, 2010.

[30] C.E. Erdem, et al., “Combining haar feature and skin color based classifiers for face detection,” IEEE 36th International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011), 2011.

[31] S. Wang and A. Abdel-Dayem, “Improved Viola-Jones Face Detector,” Department of Mathematics and Computer Science, Laurentian University Sudbury, Canada, 2012.

[32] MPLab-Machine Perception Laboratory, The MPLab GENKI database, GENKISZSL subset, 2009. URL http://mplab.ucsd.edu/~nick/GENKI-R2009a.tgz.

[33] W. Markus, Frontal face dataset, 1999. URL http://www.vision.caltech.edu/html-files/archive.html

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. 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.

Thanh-Lam Nguyen is currently a Deputy Head of Scientific Research Office of Lac Hong University. He has published several publications as listed on http://orcid.org/0000-0002-8268-9854. His main research interests are Statistics, Data Analysis, Fuzzy Control, Quality Management