Road Traffic Control Gesture Recognition using Depth Images

Quoc Khanh Le[1], Chinh Huu Pham[1] and Thanh Ha Le[1]

Abstract – This paper presents a systemused toautomatically recognize the road traffic control gestures of police officers. In this approach,the control gestures of traffic police officers are captured inthe form of depth images.A human skeleton is then constructed using a kinematic model. The feature vector describing a traffic control gesture is built from the relative angles found amongstthe joints of the constructed human skeleton. We utilizeSupport Vector Machines (SVMs) to perform the gesture recognition. Experiments show that our proposed method is robust and efficient and is suitable for real-time application. We also present a testbed system based on the SVMs trained data for real-time traffic gesture recognition.

Keywords: Traffic control gestures, Gesture recognition, Depth images

1. Introduction

Human traffic control is preferred for developing nations because of the relatively few cars, few major intersections, and the low cost of human traffic-controllers [1]. In a human traffic control environment, drivers must follow the directions given from the traffic police officer in forms of human body gestures. To improve the safety of the drivers, our research team is developing a novel method used to automatically recognize traffic control gestures.

There have been a few methods developed for traffic control gesture recognition in the literature. Fan Guo*et al.* [2] recognized police gestures from the corresponding body parts on the color image plane. The detection results of this method were heavily affected by background and outdoor illumination because the traffic police officer in a complex scene is detected by extracting the reflective vest using color thresholding. Yuan Tao *et al.* [3] affixed on-body sensors onto the back of the officer'shand to extract the gesture data. Although this accelerometer-based sensor method may output accurate hand positions, it gives an extra onusto the police and requires a unique communication protocol for the vehicles. Meghna Singh *et al.* [4] used Radon transforms to recognize air marshals' hand gestures for steering aircraft on the runway.However, since a relatively stationary background in thevideo sequence is required, this method is not practical for automotive traffic scenarios.

Human gesture recognition for traffic control can be related tothat used for human-robot interaction. Bauer *et al.* [5] presented an interaction system where a robot asks a human for directions, and then interprets the given directions. This system includes a vision component where the full body pose is inferred from a stereo image pair. However, this fitting process is rather slow and does not work in real time. Waldherr*et al.* [6] presented a template-based hand gesture recognition system for a mobile robot, with gestures for the robot to stop or follow, and rudimentary pointing. Sincethe gesture system is based on a color-based tracker, several limitations are imposed on the types of acceptable clothing,which mustcontrast with the background. In [7], Van den Bergh *et al.* introduced a real-time hand gesture interaction system based on a Time-of-Flight (ToF) camera. Haarlet-based hand gesture classification uses both depth images from the ToF camera and the color images from theRGB camera. Similar ToF-based systems havealso been described in the literature [8-10]. The use of the ToF camera allows for a recognition system robust to all colors of clothing, to background noise, and thepresence of other people. However, ToF cameras are expensive and suffer from a very low resolution and a narrow angle of view. M. V. Bergh *et al.* [11] implemented a pointing hand gesture recognition algorithm based on the Kinect sensor to tell a robot where to go. Although this system can be used for real-time robot control applications, it cannot be applied directly to a traffic control situation because of the limitation of meaningful gestures presented only by the pointing of hands.

The approach of using RGB images or videos for human detection and recognition faces challenging problems, due to variations in pose, clothing, lighting conditions, and backgroundcomplexity. It results in a reduction of the detection and recognition accuracy or in anincrease in the computational cost. Therefore, the approach of using 3D reconstruction information obtained from depth cameras has been a recent focus of study [12-16].

The researchers in [17] proposed another pose classification scheme based on the joint angles found in a human body. By using the joint angles, a variety of human poses can be modeled. This includes not only the poses already

* Corresponding Author: Thanh Ha Le
1 Human Machine Interaction Laboratory, Department of Information Technology, Vietnam National University, Hanoi {khanhlq_53, huupc_53, ltha}@vnu.edu.vn
Received: March 21, 2012; Accepted: June 15, 2012

existing in the dataset, new poses were able to generated directly from the estimated features. The classification method was based on the range of joint angles interpreted in test experiments. Therefore, the need to build a dataset of the ranges of the joint angles is required.The joint angles are computed from 3D models extracted from existing datasets. Hence, this research is considered to be a notable milestone in human pose recognition.

This paper presents a road traffic control gesture recognition system. This approach defines six common types of body gestures used by police officers to control the flow of vehicles at an intersection in Vietnam. In order to recognize the defined gestures, depth images are used instead of RGB images. Depth images have several advantages over 2D intensity images; depth images are robust to changes in color and illumination and are simple representations of 3D information. To make police officer recognition and tracking easier, the depth image is good meansto discernthe gestures of officers. Moreover, a skeleton presentation of police officer body is computed quickly from the depth data of the depth images. As done in [17], feature vectors are created based on the relative angles amongstthe joints of the skeleton model. However, the feature vectors are extracted using a simpler methodin order to reduce the computationcomplexity. In order to perform the gesture recognition, we evaluated the recognition performance bySupport Vector Machines (SVMs) classifiers. The experiment results show the recognition feasibility using SVMs along with an acceptable computation time for real-time applications. We also present a testbed system based on the SVMs trained data for real-time traffic gesture recognition.

The remainder of this paper is organized as follows: In Section 2, we discusshuman parts recognition using depth images. The details of our proposed approach are presented in Section 3. Theexperiment results demonstrating our approach's performance and the testbed system are found in Section 4. Finally, conclusionsaredrawnin Section 5.

2. Human Body Parts Recognition Using Depth Images

For human body part recognition purposes, PrimeSense has created an open source library – Open Natural Interaction (OpenNI) [20] – to promote natural interactions. OpenNI provides several algorithms usedfor PrimeSense's compliant depth cameras in natural interaction fields. Some of these algorithms provide the extraction and tracking of a skeleton model from the user interacting with the device. Fig. 1 illustrates this human detection and recognitionprocess. The kinematic model of the skeleton is a full skeleton model of the body consisting of15 joints, as shown in Fig. 2. The algorithms provide the 3D positions and orientations of thejoints and updates at the rate of 30fps.

Fig. 1. Human detection and recognition in depth images

Other research using depth images for human body part estimation have also been addressed. In [21], J. Charles *et al.* proposed a method for learning and recognizing 2D articulated human pose models from a single depth image obtained usingMicrosoft Kinect™. Although the pose estimation is substantially recognized, the 2D representation of the articulated human pose models makes the human activity recognition process more difficult comparedto the3D representation of OpenNI. In [22], L. M. Adolfo *et al.* presented a method for upper body pose estimation using anonline initialization of the pose and anthropometric profile. A likelihood evaluation is implemented to allow the system to run in real-time. Although the method in [22] has a better performance, whencomparedtoOpenNI in limb self-occlusion cases, only the upper presentation of the body pose is suitable for a small range of recognition applications. Forthese reasons, we chose OpenNI to preprocess the depth images to obtain the human skeleton models.

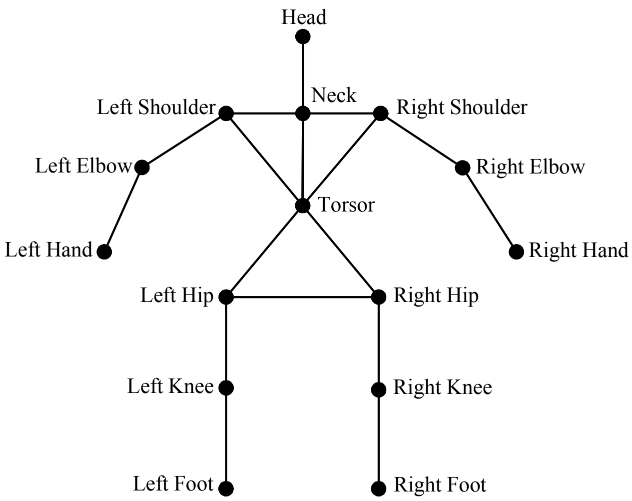

Fig. 2. The Open NI human body kinematic model

3. Road Traffic Control GestureRecognition

3.1 The Road Traffic Control Gestures

In a human traffic control system, a human traffic controller is able to assess the traffic within visual range around the traffic intersection. Based on his observations, he makes intelligent decisions and deliverstraffic signals in the form of his arms' directions and movements to all of the incoming vehicle drivers. In this paper, we only consider the arm directions for classifying the traffic control commands. Based on observations at a real traffic intersection in Vietnam, we categorizedthe control commands into three types, as shown in Table 1.

Table 1. The Three Types of Control Commands

Type	Command	Human Arm Directions
1	Stop all vehicles in every road directions.	Left/right arm raises straight up
2	Stop all vehicles in front of and behind the traffic police officer.	Left/right arm raises to the left/right
3	Stop all vehicles on the right of and behind the traffic police officer.	Left/right arm raises to the front

Six traffic gestures can be constructedfrom these control command types. Each traffic gesture is a combination of the arms' directions.The various gestures arelisted in Table 2.

Table 2. The Six Defined Traffic Gestures

Gesture	Human arm directions	Command type
1	left hand raises straight up	1
2	right hand raises straight up	1
3	left hand raises to the left	2
4	right hand raises to the right	2
5	left hand raises to the front	3
6	right hand raises to the front	3

As stated in the previous section, human physiology,including arm directions, can be represented by a skeleton model consisting of 15 joints, namely, the head, neck, torso, left shoulder, right shoulder, left elbow, right elbow, left hand, right hand, left hip, right hip, leftknee, right knee, left foot, and right foot. Therefore, the recognition of traffic gestures can be done using a skeleton model. Fig. 3 depicts an example of a traffic gesture and its correspondingskeletal joint structure. Sincethe skeleton model visualizes human parts simply usinga set of the relative joints, the skeleton appears to have a significant recognition advantage over the raw depth and color information. Therefore, instead of directly doing human parts recognition using depth and color images, we do skeleton recognition after preprocessing the Kinect's depth images by using the OpenNI library.

Fig. 3. Traffic gestures and skeletal joints

3.2 Feature Selection

For classification, a fixed size feature vector, which is invariant to translation, rotation, and scaling, must be extracted for each skeleton. In our research, we used the relative angle between the joints for the feature vector attributes,sincethis information is invariant in the real space coordinate system.

Research discussed in [23] builds the 3D human model from the depth images and extracts the feature vectors based on the angles regarding the 3D body parts and coordinate axes. The angle computation based on this information, however, is highly complex. Therefore, the 3D body parts are simplified to body joints. Then, the set of joints is used to construct a human skeletal model. By applying OpenNI [20], the 3D human model is transformed into a human skeleton. There are 15 joints which are used to denote 15 corresponding human body parts, as depicted in Fig. 2.

Based on the descriptions in Table 2, the differences between traffic gestures mainly depend on the relative angles between the upper body parts,i.e. the arm, hand, shoulder, and backbone, which is constructed from the neck and torso.

In order to construct a feature vector for a traffic gesture, we denote body part vector V(x,y) as the vector from joint name x to joint name y;the body part angle $\angle(V(x_1, y_1), V(x_2, y_2))$ is the angle between the two body part vectors *V(x1, y1)* and *V(x2, y2)*. The feature vector can then be constructed by the combination of 10 predefined body part angle radians as:

- $\angle$(*V*(left elbow, left shoulder), *V*(left elbow, left hand))
- $\angle$(*V*(right elbow, right shoulder), *V*(right elbow, right hand))
- $\angle$(left shoulder, neck), *V*(left shoulder, left elbow))
- $\angle$ (right shoulder, neck), *V*(right shoulder, right elbow))
- $\angle$ *V*(neck, head), *V*(neck, left shoulder))
- $\angle$(neck, head), *V*(neck, right shoulder))
- $\angle$(left shoulder, left hand), *V*(head, torso))

- ∠(right shoulder, right hand), *V*(head, torso))
- ∠ (left shoulder, left hand), *V*(left shoulder, right shoulder)
- ∠(right shoulder, right hand), *V*(left shoulder, right shoulder)

It is well known that the angle between two vectors is computed by:

$$\cos(\alpha) = \frac{\vec{V}_1 \cdot \vec{V}_2}{\left|\vec{V}_1\right| \times \left|\vec{V}_2\right|} \tag{1}$$

From the above method, it is easy to calculate the feature vectorattributes. The feature vector of the gesture "left hand raises straight up" can be described by this vector (π, π, π/2, π/2, π/2, π/2, 0, 0, π/2, π/2).

3.3 Training and Classification

Support vector machines (SVMs) [24] area set of related supervised learning methods that analyze data and recognize patterns and are used for classification and regression analysis. In our research, SVMs use the training data to generate a SVM model with six labels linked to thesix gestures. In real-time process, data from each frame is transformed to SVMs data. After that, we used the SVMs prediction function to compare this data with the SVMs training model.The result will be the gesture that ismost similar tothis action. All ofthe SVMs processing (training and predicting) have been collected in an open source library – libSVM [25].

4. The Experiment Results

4.1 Training Data Collection

4.1.1 Obtaining the Depth Datausing Microsoft Kinect™

Earlier depth sensors were expensive and difficult to use in human environments because they required lasers. Fortunately, with the emergence of cheap but high quality depth sensors, such as the MS Kinect and Asus Xtion, researchers around the world are now encouraged to include depth information in their work. Recently, Microsoft has launched the Kinect, a peripheral designed as a video game controller for the Microsoft X-Box Console. Nevertheless, despite its initial purpose, the system facilitates research in human detection, tracking, and activity analysis, thanks to the combination of its high capabilities and low cost. The sensor provides a depth resolution similar to ToF cameras, but at a fraction of the cost. To obtain the depth information, the device uses PrimeSense's Light Coding Technology [18], in which Infra-Red (IR) light is projected as a dot pattern in the scene. This projected light pattern creates textures that help to find the correspondence between the pixels, even for shiny or texture-less objects or under harsh lighting conditions. In addition, because the pattern is fixed, there is no time domain variation other than the movements of the objects in the field of view of the camera. This ensures a precision similar to the ToF,however PrimeSense's mounted IR receiver is a standard CMOS sensor, which reduces the price of the device drastically.

Fig. 4 depicts the block diagram of the reference design used by the Kinect sensor [19]. The sensor is composed of one IR emitter, responsible foremitting the light pattern to the scene and a depth sensor responsible forcapturing the emitted pattern. It is also equipped with a standard RGB sensor that records thescene in visible light. Both the depth and RGB sensors have a resolution of 640x480 pixels. The matching calibration process between the depth and the RGB pixels and the 3D reconstruction is handled at the chip level.

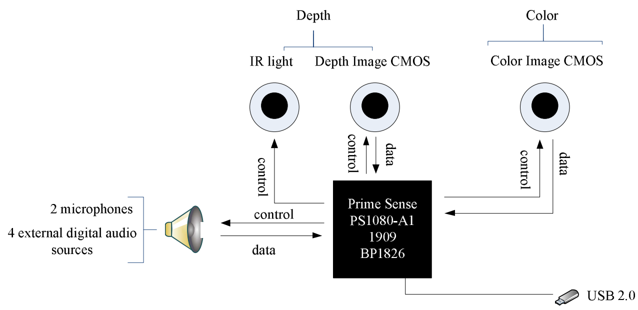

Fig. 4. The block diagram of the Prime Sense reference design [18]

Unfortunately, there is an important issue regardingthe MS Kinect in that the Kinect depth perception is highly unreliable in outdoordaylight. In outdoor scenes, because the sun is a strong source of infra-red, the infra-red structured light emitted by Kinect sensor will be overwhelmed. A solution is to use a different class of infra-red depth sensor. Instead of using spatial light patterns, the sensor sends out temporally modulated IR light and measure the phase shift of the returning light signal. These sensors are known as ToF cameras. However, ToF cameras are expensive and suffer from a very low resolution. Therefore, the Kinect system is still a good candidate for obtaining depth data. In addition, [26] indicated that the Kinect can be made to work properly in outdoor environments.

4.1.2 Data Collection

To collect the training data samples, we captured a traffic gesture database for a group of five persons. Each person performed a traffic gesture at different locations and angles to Kinect sensor. For each traffic gesture, we recorded about 5000 frames of depth images. The coordinates of all 15 skeletal joints fromeach frame werethen calculated and stored in the traffic gesture database. The number of training vectors in our traffic gesture database totaled30509; each vector waslabeled by its gesture index number.

4.2 The Method

The Weka tool [27] wasused to train and test the human pose recognition accuracy employing C-SVMs classifiers with C=1.0 andkernel RBF. The 30509 samples ofdata set waslabeled usingthe six defined gestures. The test mode used10-fold-cross-validation, which means that 1/10 samples are retained as the validation data for testing the model, and the remaining 9/10 samples are used as the training data. Table 3 shows the experiment results (in percentage) forC-SVMs. The True Positive (TP) rate is the proportion of samples which were classified as gesture *x*, among all of the examples truly labeled as gesture *x*. The False Positive (FP) rate expresses the proportion of the samples classified as gesture *x*, but labeled as a different gesture among all of the examples which are not labeled as gesture *x*. Precision indicates the proportion of the samples which is truly labeled as x among all those which were classified as gesture *x*. The experiments were done on a Windows Pentium 4PC with 1GB of RAM.

Table 3. The SVMs Results by Class

Gesture	TP Rate	FP Rate	Precision	Recall	F-Measure
1	99.8	0.0	100.0	99.8	99.9
2	100.0	0.0	100.0	100.0	100.0
3	100.0	0.0	100.0	100.0	100.0
4	100.0	0.0	100.0	100.0	100.0
5	100.0	0.0	99.8	100.0	99.9
6	100.0	0.0	100.0	100.0	100.0

The results indicate that the proposed method achieved high recognition rates. It also can been seen that the misrecognized gesture rate of the SVMs classifier is practicallyzero. The experiments also show that the running time of the SVMs classifier is 2.86 seconds toprocess the whole database. Thisindicates that the SVMsclassifier can be used inreal-time applications. Therefore, from these results, we suggest theuse of SVMs classifier for training and predicting traffic control gestures in real-time application.

4.3 The Testbed System

A real-time testbed system for traffic control gesture recognition wasbuilt; the data flow diagram forthe system is presented in Fig. 5. The system is divided into two parts: training and prediction. In the training part, the entiretraffic gesture dataset, as outlinedin the previous section, wasused to train the SVMs classifier. All of the parameters of the trained classifiers are stored to be usedby theprediction. In the prediction part, the depth images captured from the Kinect sensor at the rate of 30 fps are recognized using the OpenNI library and the skeletal models are determined. Each skeletal model is then predicted to obtain the gesture number. Because of misclassification, especially at the border of two human poses, the gesture number forthe same human pose may vary rapidly. Therefore, we choose the gesture number that occurs the most overa specific period (1 second) to indicate the current human gesture of the traffic police officer and signal it to the vehicle driver. Fig. 6 shows the user interface of this application when recognizing human gestures.

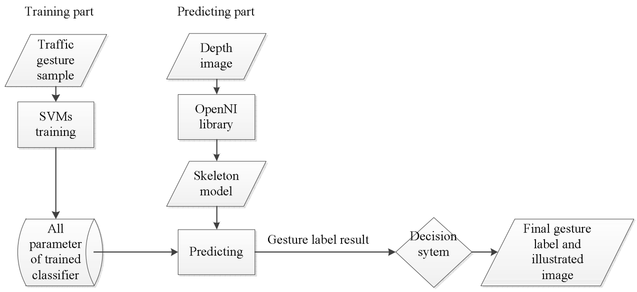

Fig. 5. The information flow in our system

Fig. 6. Three examples of the testbed graphic user interface

5. Conclusion

We have presented an algorithm used torecognize the command gestures of traffic police based on depth images retrieved from a Kinect sensor. The depth images provide 3D information useful in representing a variety of human poses. A key feature of our proposed approach is the use of the geometric relations amongstskeletal joints extracted from the depth images to build feature vectors and use them to train and recognize gestures.

There are several advantages of the proposed approach. First, the method requires no special clothing or markers commonly used in motion-capture applications. Second, the gestures can be recognized even if they are not performed perfectly. Experiments show that the proposed method is robust and efficient. In future work, we intend to enhance the flexibility of this algorithm by using dynamic gestures to enable the systemto handle more traffic policegestures.

References

[1] KhanS, "Automated versus Human Traffic Control for Dhaka and Cities of Developing Nations", ICCT, 2007.Article (CrossRef Link)

[2] Fan Guo, ZixingCai, Jin Tang, “Chinese Traffic Police Gesture Recognition in Complex Scene”,Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE 10th International Confer-

ence, 2011.Article (CrossRef Link)
[3] Yuan Tao, Wang Ben. “Accelerometer-based Chinese traffic police gesture recognition system”, Chinese Journal of Electronics, pp. 270-274, 2010.Article (CrossRef Link)
[4] Meghna Singh, MrinalMandal and AnupBasu.“Visual gesture recognition for ground air traffic control using the Radon transform”, IEEE/RSJ IROS, 2005.Article (CrossRef Link)
[5] A. Bauer, K. Klasing, G. Lidoris, Q. Mhlbauer, F. Rohrmller, S. Sosnowski, T. Xu, K. Khnlenz, D. Wollherr and M. Buss, “The Autonomous City Explorer: Towards Natural Human-Robot Interaction in Urban Environments”, International Journal of Social Robotics, 1 (2009) , no. 2 , 127-140.Article (CrossRef Link)
[6] S. Waldherr, S. Thrun, R. Romero, D. Margaritis, “Template-based recognition of pose and motion gestures on a mobile robot”, Proceeding of AAAI-98, AAAI Press/The MIT Press, 1998.Article (CrossRef Link)
[7] M. Van den Bergh and L. Van Gool, “Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction”, Proc. of the IEEE Workshop on Applications of Computer Vision (WACV2011), January 2011.Article (CrossRef Link)
[8] P. Breuer, C. Eckes, S. Muller, ”Hand gesture recognition with a novel IR Time-of-Flight range camera: A pilot study”, Proc. of the Third International Conference on Computer Vision/Computer Graphics Collaboration Techniques, MIRAGE, 247- 260 (2007).Article (CrossRef Link)
[9] E. Kollorz, J. Penne, J. Hornegger, A. Barke, ”Gesture recognition with a Time-of-Flight camera”, International Journal of Intelligent Systems Technologies and Applications, 5(3/4), 334-343 (2008).Article (CrossRef Link)
[10] S. Soutschek, J. Penne, J. Hornegger, J. Kornhuber, ”3-d gesturebased scene navigation in medical imaging applications using Timeof-Flight cameras”, Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 1-6 (2008).Article (CrossRef Link)
[11] M. V. Bergh, D. Carton, R. D. Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. V. Gool, and M. Buss, “Real-time 3D Hand Gesture Interaction with a Robot for Understanding Directions from Humans”, Proc. of IEEE Conf. on RO-MAN, 2011.Article (CrossRef Link)
[12] V. Ganapathi, C. Plagemann, D. Koller, S. Thrun,“Real time motion capture using a single time-of-flight camera” Proceedings of CVPR 2010. pp.755~762.Article (CrossRef Link)
[13] HP. Jain and A. Subramanian, “Real-time upper-body human pose estimation using a depth camera”, In HP Technical Reports, HPL-2010-190, 2010.Article (CrossRef Link)
[14] J. Rodgers, D. Anguelov, H.-C. Pang, and D. Koller., “Object pose detection in range scan data”, In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2006. Article (CrossRef Link)
[15] Y. Zhu, B. Dariush, and K. Fujimura, ”Controlled human pose estimation from depth image streams”, Proc. CVPRWorkshop on TOF Computer Vision, June 2008. Article (CrossRef Link)
[16] J. Smisek, M. Jancosek and T. Pajdla, “3D with Kinect,” IEEE ICCV Workshops, 2011.Article (CrossRef Link)
[17] V. Maik, D.T. Paik, J. Lim, K. Park and J. Paik, “Hierarchical pose classification based onhuman physiology for behaviour analysis”, Computer Vision, IET, Vol.4, pp.12-24. Article (CorssRef Link)
[18] PRIMESENSE LTD. PrimeSense’s Frequently Asked Questions (FAQ) website, June 2011.Article (CrossRef Link)
[19] PRIMESENSE LTD. PrimeSense’sPrimeSensor Reference Design 1.08, June2011.Article (CrossRef Link)
[20] OpenNI members, “OpenNI web page – openni.org,” June 2011.Article (CrossRef Link)
[21] J. Charles and M. Everingham, “Learning shape models for monocular human pose estimation from the Microsoft Xbox Kinect”, Proc. of IEEE ICCV Workshops, 2011. Article (CrossRef Link)
[22] L. M. Adolfo, M. Alcoverro, P. Montse and J. R. Casas, “Real-Time Upper Body Tracking with Online Initialization using a Range Sensor”, Proc. of IEEE ICCV Workshops, 2011. Article (CrossRef Link)
[23] Seong-Whan Lee, “Automatic Gesture Recognition for Intelligent Human-Robot Interaction”, Proc. of the 7th International Conference on Automatic Face and Gesture Recognition (FGR’06).Article (CrossRef Link)
[24] Aditya Krishna Menon.“Large-scale support vector machines: algorithms and theory”, USCD Research Exam Report, 2009.Article (CrossRef Link)
[25] Chih-Chung Chang, Chih-Jen Lin, LibSVMArticle (CrossRef Link)
[26] A. Robledo, S. Cossell and J. Guivant, “Outdoor ride: Data fusion of a 3D Kinect Camera installed in a bicycle”, Proc. of ACRA, 2011.Article (CrossRef Link)
[27] Machine learning group at University of Waikato, Weka Tools.Article (CrossRef Link)

QuocKhanh Leearned aBachelor of Computer Science at the University of Engineering and Technology (UET), Vietnam National University (VNU), Hanoi, Vietnam. He was a student of the International Standard Program of UET in Computer Science. He is working at the Human Machine Interaction Laboratory, UET as undergraduate assistant researcher. He is involved in two main projects:Vietnamese talking face construction and human gesture recognition using depth data. His research interests include digital image processing, computer vision,and human computer interaction.

ChinhHuu Pham received his B.S. degree in Computer Science from the University of Engineering and Technology (UET), Vietnam National University (VNU), Vietnam, in 2012. Before his graduate study, he worked at the Human Machine Interaction lab of the Faculty of Information Technology, UET as an assistant researcher. In 2011, he was involved in building a Vietnamese Talking Face project. His research interests include human machine interaction, digital processing, and computer vision.

Thanh Ha Lereceived B.S. and M.S. degrees in Information Technology from the College of Technology, Vietnam National University, Hanoi in 2005. He received a Ph.D. at the Department of Electronics Engineering at Korea University. In 2010, he joined the Department of Information Technology, University of Engineering and Technology, Vietnam National University. His research interests are images, videos processing, and coding. He also takes parts in other research topics including satellite image processing, computer vision, robotics.