Mobile Device-based Three-Dimension Coordinate Estimation Using Face Detection

Byung-Hun Oh[1], Kwang-Woo Chung[2] and Kwang-Seok Hong[3]

[1,3] *School of Electronic Electrical Engineering, Sungkyunkwan University, Suwon-Si, 440-746, Korea*
[2] *School of Information and Communication Engineering, Korea National University of Transportation, Uiwang-Si, Korea*
[1] *sincelife@skku.edu,* [2]*ckw1201@ut.ac.kr,* [3] *kshong@skku.ac.kr*

Abstract

In this paper, we propose a 3D coordinate estimation for 3D space using a mono camera in mobile devices. Generally, a marker is used for the detection of 3D movement in 2D images. However, this marker-based system has obvious defects since markers always need to be included in the image. To overcome this problem, we replaced the markers with a planar face by estimating the face detection. The experimental result indicates the feasibility of the proposed algorithm for face movement-based mobile user interfaces.

Keywords: *Mobile Device, 3D Coordinate Estimation, Face Detection.*

1. Introduction

The rapid growth of mobile devices including smartphones and tablet PCs has introduced multimedia content and web services into our daily lives. Users want to increase the efficiency of their interaction with mobile devices and applications. To satisfy this demand, user interfaces are increasingly gaining in importance [1]. Mobile user interfaces(UI) allow for interaction between a user and a mobile device. A user’s ability to efficiently control the device is strongly related to the UI.

Research has been carried out on new types of input interface systems. Accordingly, the latest mobile devices can support a variety of UIs ranging from keypads to multi-touch screens and sensors. However, as the number of UI sensors increases, the sensors will become difficult to integrate into existing small form-factor mobile devices at hardware level [2, 3]. In addition its function cannot be represented exactly in a simplified two-dimensional form. Therefore, vision-based 3D mobile UIs can serve as an important way to use camera-equipped mobile devices since no new hardware is necessary.

To introduce the concept of the three-dimensional, Researchers discuss the interaction techniques for generic 3D tasks on the basis of the use of traditional 2D images. Markers are the key element for deriving three-dimensional structural information from images in a way that recognizes the existing methods [4, 5]. However, this marker-based system has obvious defects since markers always need to be included in the image or additional equipment is required for controlling objects, which results in reduced immersion.

We present a 3D coordinate estimation based on face detection using the Adaboost algorithm [6, 7, 8]. The proposed systems for mobile UIs require face interaction and were developed for a front-facing camera. They also estimate the face movement and gestures and operate 3D mobile applications using face images conveyed through the front-facing camera.

2. System Architecture

Figure 1 shows the implementation details of the proposed architecture that can accurately detect a face from a single image captured by a mobile phone camera. The detection module is composed of two pre-processing procedures which input image and face area estimation as sequential steps using the Adaboost algorithm. The two pre-processing procedures used are based on gradient information and color information. An AND image between the morphological gradient combination image and the color-base pre-processed image is finally used to input the Adaboots algorithm in order to detect the face area. The estimated coordinates and the area of the face are then further refined and fed into the 3D position estimation, which determines the 3D coordinate of mobile applications.

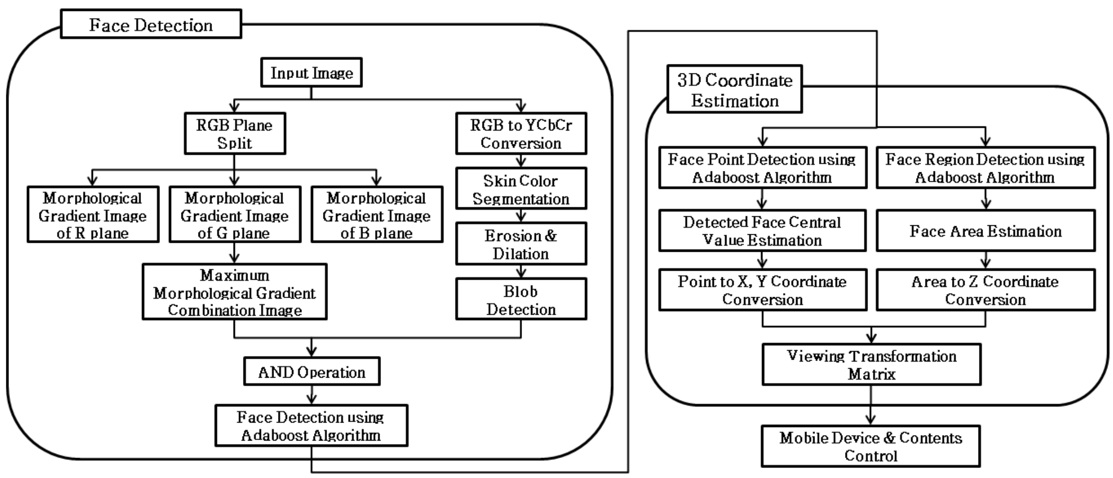

Figure 1. Proposed System Architecture

3. Proposed Method

3.1. Face Detection

For the skin color segmentation, each pixel is classified as being either skin or non-skin and is then converted into a new binary image with the threshold value analysis defined as follows:

$$SkinColor(x,y) = \begin{cases} 1 & if(77 \leq C_b \leq 127) \cap (133 \leq C_r \leq 178) \\ 0 & otherwise \end{cases} \quad (1)$$

To reduce the effects of small background objects in the binary image, two basic morphological operations are performed; non-skin color objects that are larger than the 3 by 3 structuring element still exit. We labelled each blob in order to remove large objects except for the face area.

An obvious gradient value cannot be acquired by using only a gray image because of the equalizing effects of red, green, blue(RGB) to gray conversion. Thus, we devised the maximum morphological gradient values in the split R, G, and B planes and combined them into a single image. This allows for clearer gradient values than those of a gray image. The Maximum Morphological Gradient Combination(MMGC) [9] image is defined in the following equation:

$$MMGC = \sum_{j}^{height} \sum_{i}^{width} \max\left(MG_R(i,j),\ MG_G(i,j),\ MG_B(i,j)\right) \tag{2}$$

Finally, we obtained an AND image between MMGC and the resulting image of the skin color segmentation and blob detection. The face detection from the AND image, which includes both the clear gradient information and the non-skin color subtraction, has a higher performance than that derived from the original image.

3.2. Face Detection Based on Adaboost Algorithm

For our face detection system, we collected the half-circle area images of the face from the pre-processing results as positive samples. The samples were collected under various lighting conditions. We collected 2240 positive samples, and 4500 images from the results of pre-processing were collected as negative samples for the training process. The face cascade classifier is a 13-stage cascade that is 20 × 10 in size.

3.3. 3D Coordinate Estimation using Face Detection

3D position estimation modules use the estimated coordinates and the area of the face. The X and Y coordinate are generated by changing of the central point value of the face detection. The Z coordinate is generated by changing of the area of the face detection.

4. Experiment Result and Application

4.1. Face detection Evaluation

We used a light meter to evaluate the performance of face tracking under varying lighting conditions. Five lighting conditions were set as shown in Table 1 including average lux values.

Table 1. Average lux values in the five different lighting conditions

	Places	Intensity of Illumination (lux)
Condition 1	Public areas with dark surroundings	20
Condition 2	Office	400
Condition 3	Corridor	1,000
Condition 4	Outdoor in shade	2,000
Condition 5	Outdoor in sunny side	60,000

To evaluate the performance of the proposed face detection, approximately 1,000 images including 100 images of each condition (five lighting conditions with simple backgrounds and five lighting conditions with complex backgrounds) were used in the real-time face detection program. The trained cascade classifier for face detection is a 13-stage cascade classifier with the required false alarm rate set at 1×10-6 and its size is 20×10 pixels. For the detection results, we measured the number of hits, misses, false pictures, and the detection time and detection rate.

Table 2. The performance of face detection classifier

		Hit	Missed	False	Detection Rate	Detection Time
Simple Background	Cond.1	98	2	1	97%	97 ms
	Cond.2	100	0	1	99%	112 ms
	Cond.3	99	1	0	99%	104 ms
	Cond.4	98	2	1	97%	99 ms
	Cond.5	100	0	2	98%	101 ms
Complex Background	Cond.1	96	4	1	95%	109 ms
	Cond.2	98	2	0	98%	102 ms
	Cond.3	97	3	2	95%	99 ms
	Cond.4	99	1	2	97%	109 ms
	Cond.5	95	5	2	93%	104 ms
Average		98	2	1.2	97%	103 ms

By analyzing the detection results, we found that the detection rate is 97% and the detection time is 0.103 seconds. The proposed face detection algorithm showed good robustness against lighting variance and complex background including skin-like color. As a result, this indicates the feasibility of the proposed algorithm for face movement-based mobile user interface.

4.2. Application

The proposed system can accurately estimated the 3D position. As shown in Figure 2, the lines are drawn according to the face 3D positions. We implemented a few test applications to demonstrate the proposed system's strengths and limitations. The proposed system can accurately estimate the 3D position and it is applicable to 3D user interfaces as shown in figure 3.

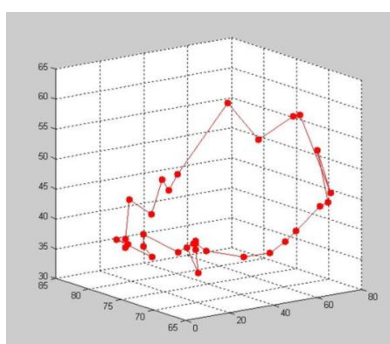

Figure 2. 3D trajectory for the face tracking test

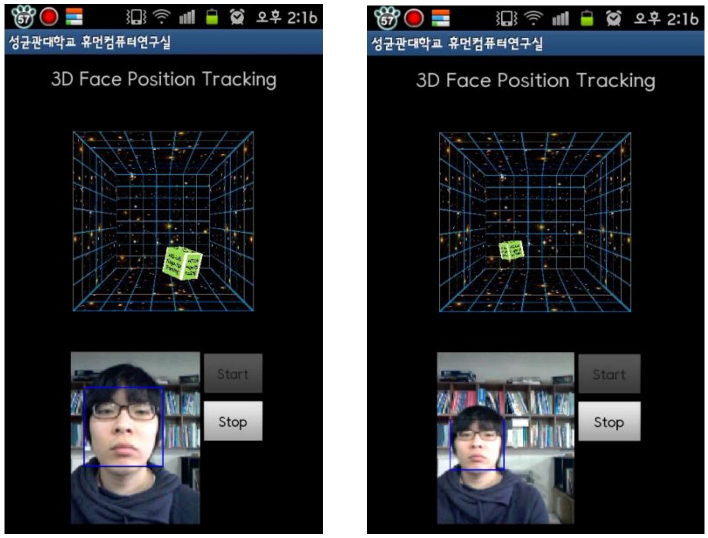

Figure 3. 3D face tracking application

5. Conclusion

We propose a 3D coordinate estimation for 3D space using a mono camera in mobile devices. The proposed system for mobile UIs requires face interaction and was developed for a front-facing camera. The proposed system further estimates the face movement and operates 3D mobile applications using the face gesture images through the front-facing camera. As a result, this indicates the feasibility of the proposed algorithm for face movement-based mobile user interfaces.

Acknowledgements

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2011-0018397) and by MKE, Korea under ITRC NIPA-2012-(H0301-12-3001).

References

[1] A. Mulderr, “Hand gestures for HCI.”, Technical Report 96-1, vol. Simon Fraster University **(1996)**.

[2] E. Koh, Jo. Won, C. Bae, “On-premesie skin color modeling method for vision-base hand tracking”, Consumer Electronics, ISCE IEEE 13 International Symposium**(2009)**.

[3] N. Kim., “Interactive Face Warping Software for Smartphone”, The Institute of Webcasting, Internet and Telecommunication, VOL. 11 No. 5, pp.65-71**(2011)**.

[4] H. Zhou, B. Duh and Billinghurst M., “Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR”, IEEE/ACM International Symposium on Mixed and Augmented Reality, pp.193-202**(2008)**.

[5] M. Fiala, “ARTag, a fiducial marker system using digital techniques”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05), Vol.2, pp.590-596**(2005)**.

[6] V. Paul and J. Michael, “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, pp. 511-518**(2001)**.

[7] Y. Freund and E. Robert, “A decision theoretic generalization of on-line learning and an application to boosting”, Computational Learning Theory: Eurocolt 95, pp.23-37**(1997)**.

[8] S.-H. Lee, M.-J. Lim, “Implementation of U-Healthcare Environment for Patient Recognition Applied Algorithms of Extracting Face Feature Points”, The Institute of Webcasting, Internet and Telecommunication, Vol. 9 No. 4, pp.53-57**(2009)**.

[9] J.-H. An, B.-H. Oh and K.-S. Hong, "A Robust Fingertip Detection Algorithm for Complex Backgrounds Using AdaBoost Algorithm", ICCCIT 2011, Dec**(2011)**.

Authors

Byung-Hun Oh

received the B. S. degrees in electronic engineering from the Eulji University, in 2011, respectively. He is presently a M.S candidate at the department of Information and Communication Engineering Sungkyunkwan University. His current research focuses on digital image processing and pattern recognition.

Kwang-Woo Chung

received the B. S., M. S., and Ph.D. in electronic engineering from the Sungkyunkwan University, Seoul, Korea, in 1989, 1991, and 1995, respectively. He has been a professor at Korea National University of Transportation, Uiwang-Si, Korea. His current research focuses on digital signal processing and speech recognition.

Kwang-Seok Hong

received his B.S., M.S., and Ph.D. in Electronic Engineering from Sungkyunkwan University, Seoul, Korea in 1985, 1988, and 1992, respectively. Since March 1995, he has been a professor at Sungkyunkwan University, Suwon, Korea. His current research focuses on human-computer interaction, five-sense recognition, interaction, and representation.