07 - 10 September 2009

PROCEEDINGS

54. IWK

Internationales Wissenschaftliches Kolloquium

International Scientific Colloquium

# Information Technology and Electrical Engineering - Devices and Systems, Materials and Technologies for the Future

Faculty of Electrical Engineering and Information Technology



TECHNISCHE UNIVERSITÄT
ILMENAU

**Impressum**


Herausgeber: Der Rektor der Technischen Universität Ilmenau
Univ.-Prof. Dr. rer. nat. habil. Dr. h. c. Prof. h. c.
Peter Scharff

Redaktion: Referat Marketing
Andrea Schneider

Fakultät für Elektrotechnik und Informationstechnik
Univ.-Prof. Dr.-Ing. Frank Berger

Redaktionsschluss: 17. August 2009

Technische Realisierung (USB-Flash-Ausgabe):
Institut für Medientechnik an der TU Ilmenau
Dipl.-Ing. Christian Weigel
Dipl.-Ing. Helge Drumm

Technische Realisierung (Online-Ausgabe):
Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau

Verlag: isle
Verlag ISLE, Betriebsstätte des ISLE e.V.
Werner-von-Siemens-Str. 16
98693 Ilmenau





ISBN (USB-Flash-Ausgabe): 978-3-938843-45-1
ISBN (Druckausgabe der Kurzfassungen): 978-3-938843-44-4

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=14089

# ON COMBINING FACE AND GESTURE FEATURES FOR CREATING A MULTI-USER GESTURE RECOGNITION SYSTEM

*Valiantsin Hardzeyeu, Ronny Paduschek*

Fraunhofer IDMT, Ehrenbergstr. 31, 98693, Ilmenau, Germany
{hrd, pdk}@idmt.fraunhofer.de

## ABSTRACT

The goal of modern human-computer interaction (HCI) systems is to fully immerse the end-user to the virtual environment using non-invasive techniques for tracking human pose and gestures. In this paper we introduce a novel method for video-based head and hand gesture mapping based on the natural geometrical scene characteristics. The method involves an adaptive mechanism for hand/head regions-of-interest extraction, dynamic estimation of the relationships between head and hands regions, followed by analysis and real-time recognition of the dynamic gestures. Furthermore, a proposed model has been implemented for the multi-user interaction schema, allowing real-time multi-user human – machine interaction with a single webcam.

***Index Terms*** **–** HCI, face detection, gesture recognition.

## 1. INTRODUCTION

Different object recognition methods have been used for head/face detection and for hand gesture recognition separately. In our studies we concentrate on combining these two technologies and adapt them for a multi-user needs. For the current project two different systems, described in [1] and [2], have been taken as a reference. In [1] the authors discuss the problematic of a holistic approach to facial expression recognition and its circumference with the gesture localization. For their research they have used a tuned HMM method for dynamic gestures classification. However, the proposed model seems to be suitable only for symmetrical hand movements (e.g. hand clapping). Thus, there is a need for an extended asymmetrical model, which may provide a user with more capability to communicate with machines via natural gesticulation using independent movement of both hands. In [2] the analytical method for personalized gesture control of the mobile robot is described. The authors use a non-independent face-gesture model. This issue of not taking in account the relationship between user's face and hand positions reduces the gesture recognition possibility only to the static hand gestures. Furthermore, the development of a multi-user interface, where two or more users might "talk" to the machine at the same time, will enhance the classical field of HCI.

## 2. PROPOSED MODEL

Thus, we introduce an analytical model for gesture mapping in multi-user applications. The approach involves the asymmetrical dynamic gesture recognition technique combined with an efficient face localization algorithm. Fig. 1 shows the proposed technology architecture.

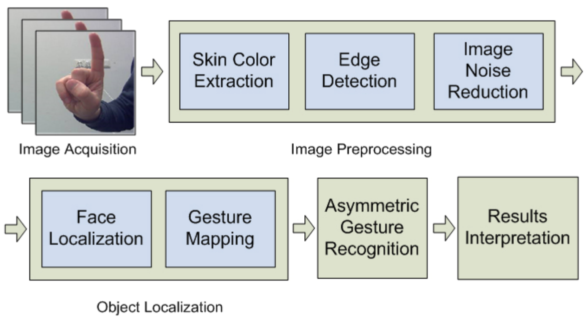


*Figure 1. Algorithm architecture.*

The proposed method is divided into three main parts. The first part deals with skin color extraction, edge detection and noise reduction. In the beginning several approaches for skin color extraction [4, 5, 8, 9] have been tested. The method from [4] showed better results and fully fit the requirements for the development of the sufficient gesture recognition system. The edge detection is performed with help of a modified Canny algorithm from [3]. Image noise reduction algorithm has been implemented as in [6].

The second part comprises the object localization algorithm, where the face and gesture regions are detected. The algorithm for this part comprises the approaches which are described in [7, 10, 11].

The main section in our model is the asymmetric gesture recognition, which introduces a triangular model and serves as a core for the proposed algorithm. This model is shown in Fig. 2a.

The first image in Fig. 2a shows the picture, captured by the camera. As it might be seen on the second image, the corresponding face and hand regions have been extracted and then marked through $P_f$, $P_{lh}$ and $P_{rh}$ on the third image.

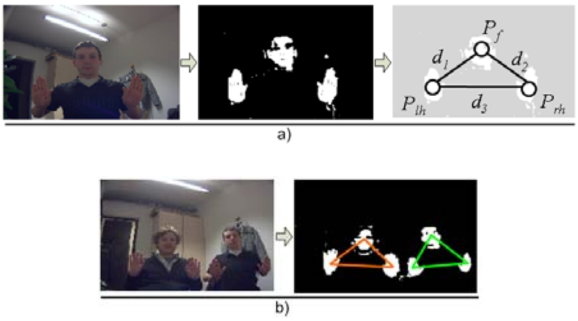


*Figure 2. Triangular model and its representation for single and multi-user application.*

The relationship between the hands and the face has been build up in triangular form. The correspondence between $P_f$, $P_{lh}$ and $P_{rh}$ is described by its Euclidean distance $d_i$.

The algorithm starts with initialization of the triangular model, which will represent a particular user. First, the face will be detected and if the detection succeeds, the image area below the face will be scanned for the corresponding hand regions. Here the hands have to be hold for a pre-defined duration, for instance for several seconds, and for the initialization they should be located below the head region.

The model exists as long as all components $P_f$, $P_{lh}$ and $P_{rh}$ appear on the image. To guarantee the triangular relationship to be hold continually, a cancelling criterion has been defined. If the triangle disappears from the image for less than one second and then shows up again with the same coordinates $P_f$, $P_{lh}$ and $P_{rh}$, the tracking process goes further. If the triangle $\Delta$ does not show up again during a pre-estimated period $\tau$, the model starts the initialization phase once again (Eq. 1).

$$\Delta_i = \begin{cases} 1 & , if\ \tau_k \leq 1s \\ 0 & , if\ \tau_k > 1s \end{cases} \qquad (1)$$

As soon as the triangular relationship has been set up and as long the triangular is hold successfully, the user is allowed to perform the gestures.

In order to realize the multi-user modus (Fig. 2b) for simultaneous control of different applications, the synchronization is needed. This means, that the hand regions should be synchronized with the corresponding face region. For this purpose the hands detection in multi-user modus will be done using a region-of-interest principle from the previous image.

## 3. EVALUATION

Evaluation aimed to prove the concept on the real world images. First, the face and hand regions test have been performed. Fig. 3 shows the real world images with the corresponding regions highlighted.

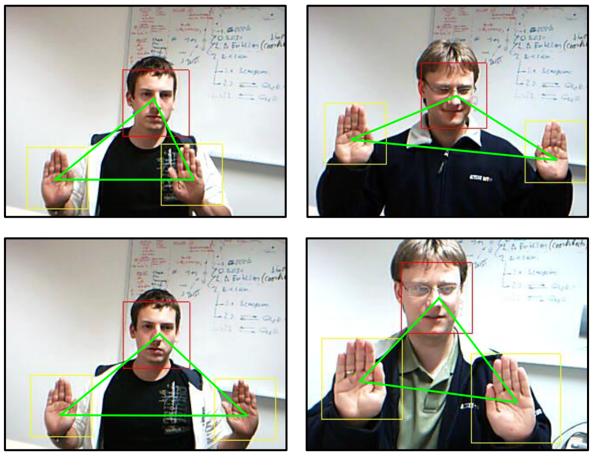

*Figure 3. Images with face (red) and hand (yellow) regions detected.*

At the same time the triangular model has been built between the centers of hand and face regions. Thin green bars in Fig. 3 show the corresponding Euclidean distances. As it was described above, for the initialization phase the hand regions should be below the head region.

On the next stage we tested, how the system will react on the horizontal dynamic gestures. The subject has been asked to hold his left hand steady and do the movement only with the right hand towards the left one. As it can be seen in Fig. 4, the face and hands regions as well as the Euclidean distances between them have been detected, although right hand wasn't hold always correctly.

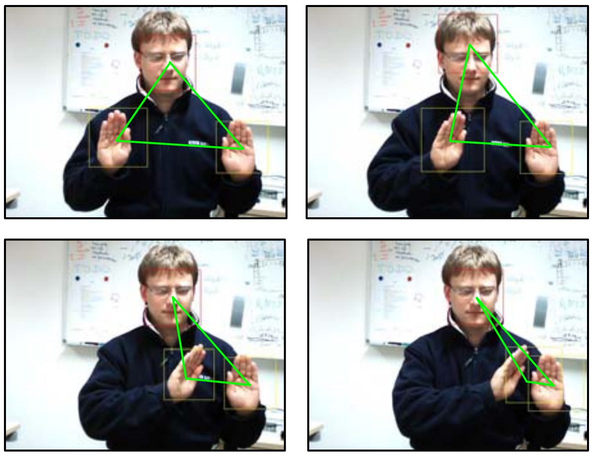

*Figure 4. A dynamic gesture (right to left).*

The next test phase was performed with a multi-user modus and aims to verify how the system will react on more than one user on the image. Fig. 5 shows the test sequence. At the beginning, the one hand of the first subject and both hand of the second subject were hidden. In this case, the system doesn't build any Euclidean distances.

*Figure 5. The multi-user modus.*

On the second image, the first subject shows both hands. Thus, the system builds up the triangle from the Euclidean distances and allowing to make the control gestures. The third image shows the situation, when the hands of subjects lay in the same hand region. Since the system possesses the information from the previous stages and the left subject had his hand in this region before, the right subject is not allowed for gesture control. The last image displays the two-user modus.

## 4. CONCLUSION

In this paper a novel model for personalized dynamic gesture has been discussed. It is based on detection of the hand/head regions-of-interest, dynamic estimation of the relationships between head and hands regions, and the real-time analysis of the user's gestures. The proposed method showed good performance in single user as well as in multi-user modus. In the future this system will be extended with static gestures recognition functionality. This will give more freedom to the end-user to interact with the computer in a natural way.

## 5. REFERENCES


[1] Karpouzis, K. et al. Facial expression and gesture analysis for emotionally-rich man-machine interaction. 3D Modeling and Animation: Synthesis and Analysis Techniques. 2004. Idea Group Publishers.

[2] C. Hu, X. Wang, M. Mandal, M. Meng and D. Li. "Efficient Face and Gesture Recognition Techniques for Robot Control", in Proc. of IEEE CCECE 2003.

[3] Chang, S.H. etal. Small retinal vessel extraction using modified Canny edge detection. In Proc of ICALIP 2008, pp.1255-1259.

[4] Shwu-Huey Yen, Chin-Wei Tsai, Tai-Kuang Li. „Efficient Face/Pose Detection Based on Machine Learning", from http://140.134.132.124.

[5] Harville M. et al."Consistent Image-Based Measurement and. Classification of Skin Color", Proc. of the International Conference on Image Processing, 2001.

[6] Al Bovik et al. Handbook on Image and Video Processing. Academic Press 2000. p. 891.

[7] Peter Dunker, Melanie Keller. Illumination Normalization for Face Recognition - A Comparative Study of Conventional vs. Perception-inspired Algorithms. In Proc. of Biosignals 2008. pp. 237-243.

[8] Emi Takami, Naruaki Toma. "Color system for skin color extraction". Artificial Life and Robotics, Springer, 2008.

[9] Jong Bae Kim, Chae-Hyun Moon, Hang-Joon Kim. Efficient Face Extraction Using Skin-Color Model and a neural Network. Lectures notes in Computer Science. Springer, 2002.

[10] Peter Dunker, Melanie Keller. Diffusion Filtering for Illumination Invariant Face. In proc. of VISAPP 2008. pp. 243 – 247.

[11] Baback Moghaddam, Tony Jebara, Alex Pentland. Bayesian Face Recognition. Int. J. of Pattern Recognition. Vol. 33, N. 11, pp. 1771-1782, 2000.