8th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS
S u c e a v a, R o m a n i a, M a y 25 – 27, 2 0 0 6

VISUAL RECOGNITION OF HAND POSTURES FOR INTERACTING WITH VIRTUAL ENVIRONMENT

Radu Daniel VATAVU[1], Ştefan-Gheorghe PENTIUC[2]
Christophe CHAILLOU[3], Laurent GRISONI[4], Samuel DEGRANDE[5]
"Stefan cel Mare" University of Suceava
str.Universitatii nr.13, RO-720229 Suceava
[1]raduvro@yahoo.com, [2]pentiuc@eed.usv.ro

Laboratoire d'Informatique Fondamentale de Lille
Universite des Sciences et Technologies de Lille
[3]Christophe.Chaillou@lifl.fr, [4]Laurent.Grisoni@lifl.fr, [5]Samuel.Degrande@lifl.fr

Abstract. *The paper addresses the problem of visual recognition of several hand postures corresponding to a few operations commonly performed in virtual environments, such as: object selection, translation, rotation and resizing. Processing is performed in a top-view scenario with a top-mounted camera that monitors the user's hands on the working desktop. By careful choosing and controlling of the scene and lighting conditions, hands segmentation is fast and robust which increases the performances of the hand posture classifier. The chosen classifier was a multilayered perceptron with three layers. By keeping all the processing at a low level of complexity and by considering an appropriate control of the environment, we obtain a real time 25 fps functional system with high detection and recognition accuracy results.*
Keywords: *computer vision, hands detection, human computer interaction, multilayered perceptron.*

Introduction

Human gestures are perceived as a natural mean for interacting and conveying information [1] and gesture based interfaces are looked upon as ideal with respect to the human computer interaction techniques [2, 3]. Even more, video based gesture recognition has the main attraction of not being intrusive and of not requiring the user to wear additional equipments or devices, giving in the end a comfortable feeling of naturalness.

For the special case of virtual environments, appropriate human computer interfaces are in order. VR appears as an impoverished version of the physical world with incomplete sensory cues and simplified and inconsistent world models. The virtual experience is influenced by experiential, cognitive, perceptual and motor differences between users. Hence, the interaction technology should be appropriate so that the overall user experience in the virtual environment should not be diminished.

The paper discusses the visual recognition of a set of hand postures that have been selected in accordance to several commonly commands that may be performed for interacting with virtual objects inside VR environments. Hand postures have been identified for common operations such as: selection, translation, rotation and resize of virtual objects. Posture recognition is carried out using a Multilayered Perceptron with a three layers structure.

Previous work

Basic research has been conducted under the general term of gesture recognition, most of which is centered on hand recognition. Gesture recognition can be grouped in two major categories: gesture acquisition (using techniques specific to video and image processing) and actual gesture recognition (techniques that are specific to pattern recognition). Gesture acquisition considers the detection and tracking of an object of interest (for example the hand with the fingers). Detection techniques include video segmentation function of several characteristics: color, motion or mixed (for example edges as the result of an edge detection process). Tracking [4] was successfully employed for continuously determining a series of characteristics of the object of interest (such as position, orientation, etc.). Several approaches have been focusing on hands detection using skin color models [5, 6, 7].

Markov models have been used by Starner & Pentland [8] for the recognition of ASL (American Sign Language). The system uses a color camera in order to detect hands wearing colored gloves. Hong et al. [9] describe a 2D gesture recognition technique in which each gesture is modeled using a finite state machine in the spatio-temporal space.

Kumar et al. [10] propose a method of gesture classification based on temporal motion templates and wavelet transforms. They use a temporal representation of gesture based on differences between consecutive images by building motion templates. For Davis [11], the starting point is motivated by the fact that a human observer can instantaneously recognize gestures without great effort in low level resolution images. A Binary Motion Region, BMR image is computed to act as an index in the gesture library. BMR describes the spatial distribution of motion for a given angle and for a give gesture.

In what concerns the interaction with virtual environments (selection, manipulation and travel), several guidelines and suggestions on selecting appropriate gestures have been proposed [1, 2, 12, 13, 14, 15]. It was taken into account also the solution for tracking the paths as in [16].

Working scenario parameters

The working scenario is as presented in Figure 1, including a top mounted camera that monitors the working area and the user's hands. A snapshot of the camera viewing angle is presented in Figure 2.

Video capture is carried out at a resolution of 320x240 and 25 fps. The working desk assures a homogenous background (see Figure 2, working desk is of blue colour) that allows for a fast and accurate segmentation of the user's hands.

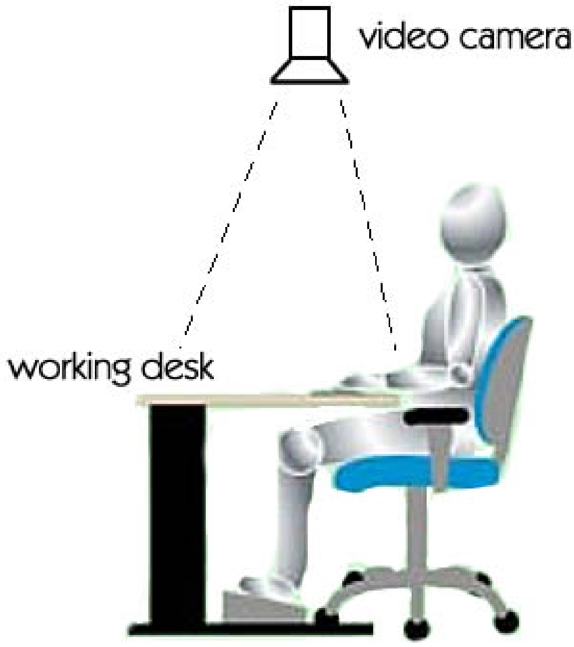


Figure 1. The working scenario (top mounted camera monitoring the working desk and the user's hands).

Figure 2. Camera view of the working area (top view).

Lighting is controlled in order to assure for a good contrast between the user's hands and the working desk. The video camera auto controls the brightness and exposure settings.

Hands detection

Hands segmentation is achieved using a simple low cost skin filtering in the HSV colour space on the hue and saturation components.

$$p \text{ is skin} \Leftrightarrow$$

$$hue(p) \in \lfloor h_{low}, h_{high} \rfloor \wedge saturation(p) \in \lfloor s_{low}, s_{high} \rfloor$$

where p is the current pixel submitted to classification and $\lfloor h_{low}, h_{high} \rfloor$ *and* $\lfloor s_{low}, s_{high} \rfloor$ are the low and high thresholds for the hue and saturation components.

The technique is very fast (the complexity order is O(n) where n is the dimension of the processed video frame) and assures for accurate hands segmentation under the previously mentioned working conditions. Segmentation results are given in Figure 3.

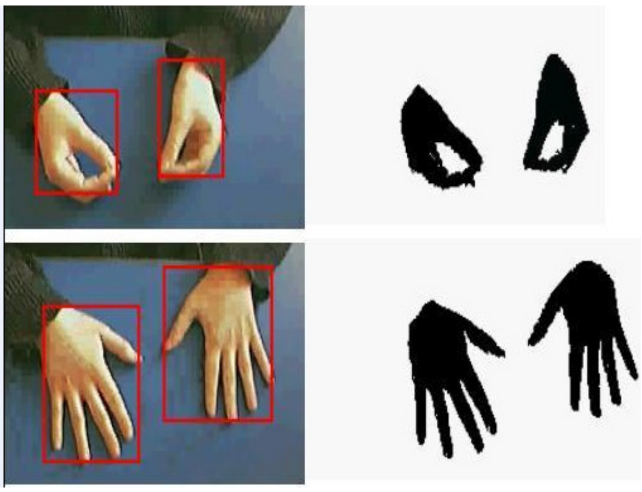


Figure 3. Segmentation results (segmentation is performed in the HSV colour space by filtering on hue / saturation).

The values for the hue / saturation thresholds were chosen experimentally as follows:
$h_{low} = 180$, $h_{high} = 240$ and $s_{low} = 20$, $s_{high} = 150$

where hue varies from 0 to 359 and saturation from 0 to 255.

Hands postures recognition

We have selected four hand postures by considering a few common operations encountered when interacting with virtual objects (see Figure 6) such as: selection, translation, rotation and resize. Two of these operations (selection and translation) are performed with one hand only, the other two (rotation and resize) are two-hand operations. Finally, we only have 3 distinct hand postures as presented in Figure 4.

Figure 4. Selected hand postures.

Recognition is performed using a multi-layered Perceptron, organized using a 3-layer structure of 39 neurons (20-16-3), as follows:

- the 1st layer consists of 20 input neurons coding each hand blob using 5 x 4 = 20 values, normalized in the interval [0..1] (see Figure 5)
- the 2nd layer uses 16 hidden neurons. Experiments showed that 16 neurons in the hidden layer offer the best performance on the testing set
- the 3rd layer with 3 neurons, each outputting a real value in the [0, 1] interval representing the probability of recognition for each of the 3 hand postures.

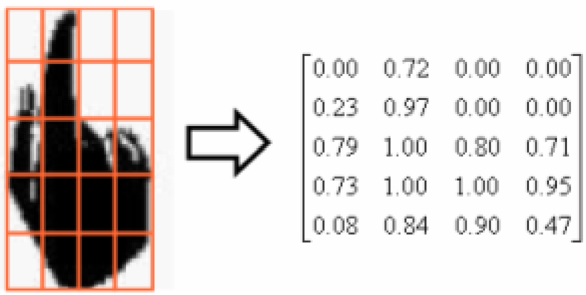
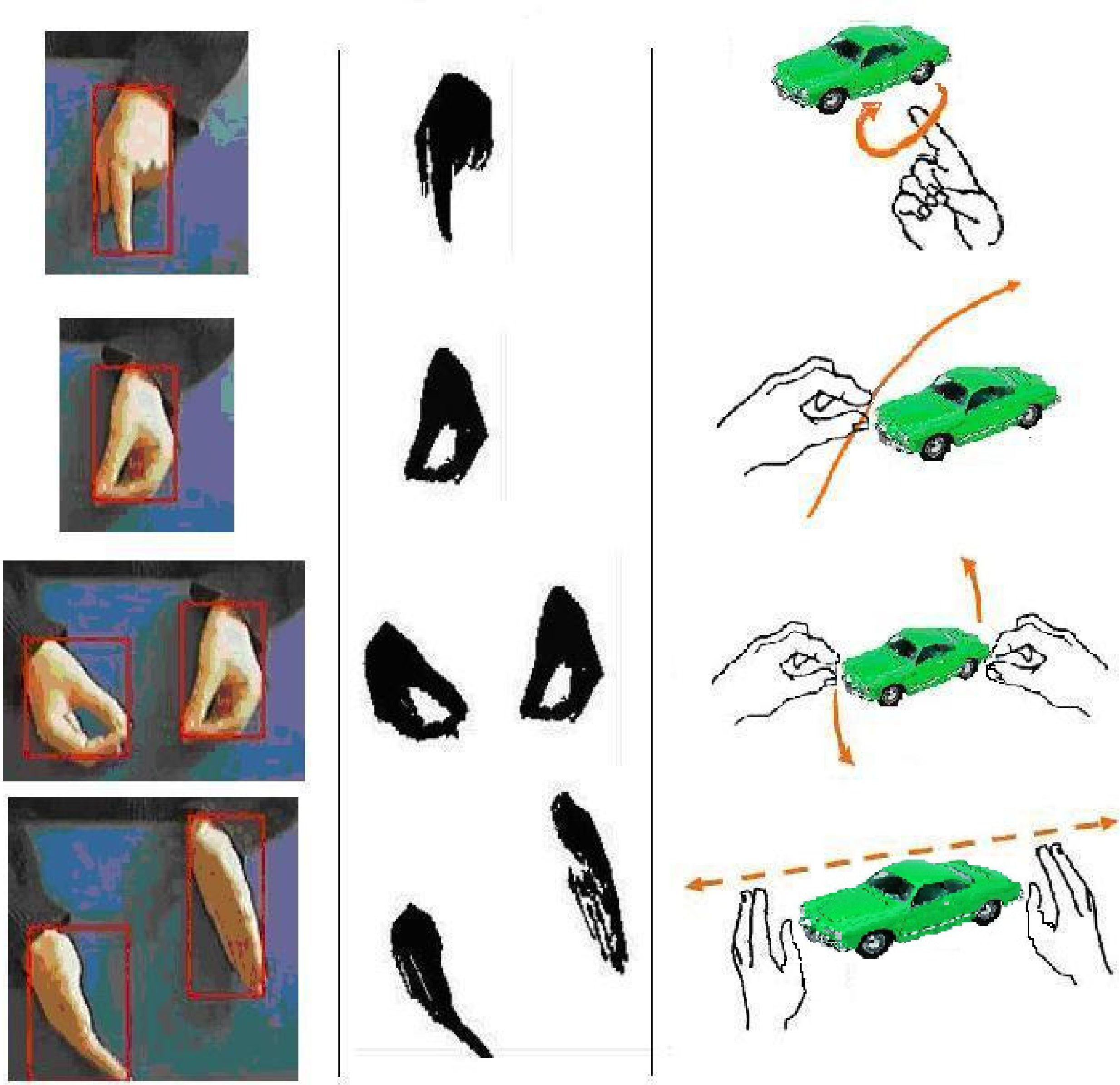


Figure 5. Hand blob coding using a 5x4 matrix structure.



a) video snapshot b) hands segmentation c) corresponding action in the virtual environment

Figure 6. Set of hand postures and associated commands for interacting with virtual objects.

The results obtained on a test set consisting of 67 images show a level of accuracy of 92%. Details with regards to the multi-layered Perceptron are given in table 1.

Table 1. Neural network details

Network structure	39 neurons distributed in 3 layers: 20-16-1
Training set	152 images
Testing set	67 images
Accuracy on the testing set	92% (61 of 67 images correctly classified)

Prior processing includes blob rotation so that the blob's longest axis should be parallel to the vertical axis, see Figure 7. This is done by computing the two axis of the ellipse of inertia having the same area and centre of mass as the hand blob.

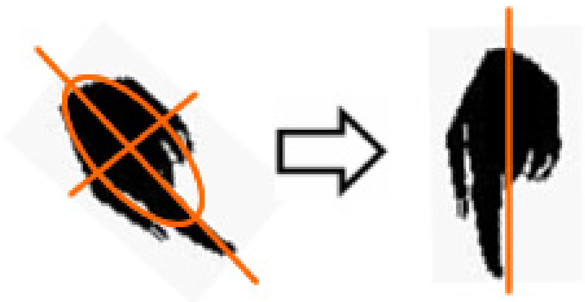


Figure 7. Blob rotation based on the axis of the ellipse of inertia.

The performances obtained overcomes those presented in [17]. A neuro-fuzzy approach, as in [18], can improve the accuracy of the recognition process.

Conclusions

Visual recognition of a set of hand postures is achieved with a real time gesture based interacting system with virtual reality. The hand postures have been selected in correspondence with a few operations commonly performed in virtual environments, such as: object selection, translation, rotation and resizing of virtual objects.
The working scenario includes a top-mounted camera that monitors the user's hands on the working desktop. Scene parameters (such as lighting conditions, working area complexity, etc.) have been carefully chosen which leads to fast and robust hands detection that increases further the performances of the postures classifier. We managed to keep all the processing at a low level of complexity and ended up with a real time 25 fps functional system with high detection and recognition accuracy results.

References

[1] Melinda M. Cerney, Judy M. Vance (2005) *Gesture Recognition in Virtual Environments*: A Review and Framework for Future Development, Iowa State University Human Computer Interaction Technical Report ISU-HCI-2005-01, 28 March

[2] Matthew Turk, *Gesture Recognition*, chapter 10, http://ilab.cs.ucsb.edu/projects/turk/TurkVEChapter.pdf

[3] Axel Mulder (1996) *Hand Gestures for HCI*, Hand Centered Studies of Human Movement Project, Technical Report 96-1, School of Kinesiology, Simon Fraser University, February

[4] Chen, F.S, Fu, C.M., Huang, C.L. (2003) *Hand gesture recognition using a real-time tracking method and hidden Markov models*, IVC(21), No. 8, August, pp. 745-758

[5] T. S. Caetano, D. A. C. Barone (2001) *A probabilistic model for human skin color*, IAP Conf., pp. 279-283

[6] Michael J. Jones, James M. Rehg (1998) *Statistical color models with application to skin detection*, Cambridge Research Laboratory, TR 98/11

[7] Tiberio S. Caetano, Silvia D. Olabarriaga, Dante A. C. Barone (2002) *Performance Evaluation of Single and Multiple-Gaussian Models for Skin Color Modelling*, Proc. 15th Brazilian Symposium on Computer Graphics and Image Processing, pp. 275-282

[8] Starner T., Pentland A. (1995) *Real time american sign language recognition from video using hidden Markov model*, TR. 375, MIT Media Laboratory

[9] P.Hong, M.Turk, T.S.Huang (2000), *Constructing Finite State Machines for Fast Gestures Recognition*, Proc. ICPR, Barcelona, Spain, Septemer 2000 (http://www.cs.ucsb.edu / ~mturk/Papers/ICPR2000.pdf)

[10] Sanjay Kumar, Dinesh Kant Kumar, Arun Sharma, Neil McLachlan (2004), *Visual Hand Gestures Classification using Temporal Motion Templates and Wavelet Transforms*, Proceedings of the 10th International Multimedia Modelling Conference (MMM'04), IEEE Computer Society

[11] James William Davis (1996) *Appearance-Based Motion Recognition of Human Actions*, MIT Media Lab, TR 387

[12] Jose Marque Soares (2004) *Contribution a la Communication Gestuelle dans les Environments Virtuels Colaboratifs*, PhD dissertation sustained 9th July 2004 at the Institut National des Telecomunications and Universite de Technologie de Troyes

[13] Alan Wexelblat (1995) *An Approach to Natural Gesture in Virtual Environments*, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 2 , Issue 3 (September 1995)

[14] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum (2003) *A procedure for developing intuitive and ergonomic gesture interfaces for HCI*, Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, Genova, Italy, April 15-17

[15] Rochelle O'Hagan, Alexander Zelinsky, Sebastien Rougeaux (2002), *Visual Gesture Interfaces for Virtual Environments*, Interacting with Computers, Special Issue

[16] Leonid Timchenko, Yuriy Kutaev, Volodymyr Kozhemyako, Alexander Gertsiy, Andriy Yarovyy, Nataly Kokryatskaya - *Method for Processing of Extended Laser Paths Images*, Advances in Electrical and Computer Engineering, ISSN 1582-7445 - No 2, pp.66-78

[17] Sorin Vlad, Nicolae Morariu (2005) - *Using Neuroshell Neural Networks Simulator for Prediction Problems Solving* , Advances in Electrical and Computer Engineering, ISSN 1582-7445 - No 1

[18] Alexandru Bârleanu - (2005) *Fault Diagnosis of Wall Hang Boliers Using a Neuro-Fuzzy Model,* Advances in Electrical and Computer Engineering ISSN 1582-7445 - No 1