HEAD POSE ESTIMATION FOR TASKS OF THE HUMAN-COMPUTER INTERACTION[1]

S. Anishchenko[2], V. Osinov[2], D. Shaposhnikov[2]

[2]A.B. Kogan Research Institute for Neurocybernetics, Southern Federal University, 194/1, Stachki Ave, 344090, Rostov-on-Don, Russia, sergey.anishenko@gmail.com

Visual information may play important role in interaction between people and computer. In this work the system which can drive computer utilizing automatic user face analysis is described. System consists of two biologically motivated models of vision and colour appearance and three algorithms. Face segmentation algorithm based on combination of different colour spaces provides a high rate of facial areas detection (p=0.95). Head movement direction identified by means geometrical relation between facial landmarks (eye corners, nose tip and point under nose). It was obtained that change of angle formed by facial landmarks particular eye corners and nose has similar dynamics as Yaw. Described system can process up to 7 fps with high rate of correct movement estimation (p=0.95).

Introduction

The Human-Computer Interaction (HCI) is the research area which aims to study, model and improve the interaction between users and computers. Existing methods of HCI use only information flow runs in direction from user to computer which passively waits for a user's command given by means of keyboard, mouth, or touch screen. It is obvious that while interacting with computer visual information about the user can be captured, analyzed, and further processed in order to make the interaction more intuitive, natural, and intelligent [4].

There are two main groups of approaches to utilize visual information about the user in HCI.

During interaction people display meaningful signal through many modality such as emotion and movements. The first group of algorithms in line with affective computing makes use unpremeditated movements and facial expression as input information from the user to the computer in order to adjust applications to the user's affective state.

The second type of approaches makes use of visual behavioral cues from the user as means of command in control situations when the user advisedly perform facial expressions, eye and head movements. The example of this approach is typing or pointing using eye tracking [5]. Also, head movements and facial expressions can be used to drive some objects in applications.

It is obviously that for implementation any of described approach a new effective and fast algorithms for processing and interpret visual information about user must be designed. In this work the system for head pose tracking is described. Algorithms are based on biologically inspired models of vision.

Description of head pose estimation system

During interaction with computer visual information about user is captured by digital camera and the processed as described below. On the first frame of video sequence the face feature points (eye corners, nose tip and point under nose) are detected in few steps. Firstly face segmentation is used to detect particular area of face. Secondly three face points are detected inside the face.

[1] Support of the Russian Foundation for Humanities,
grant N 09-06-95218 a/F

In assumption that facial landmarks cannot shift farther than ten pixels in any direction the only area (size 21x21 px) around previously detected points is scanned to detect new landmarks location on the consequent frames.

Starting from second frame the information about feature shifting on two consequent frames is analyzed to infer about head movement direction if it presented. The scheme of system is presented on the Fig 1.

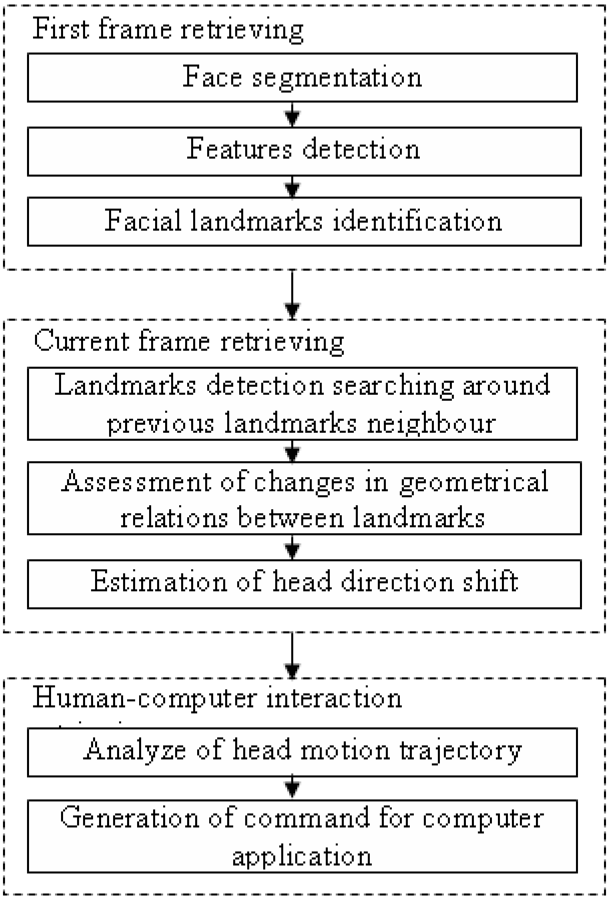
First frame retrieving
Face segmentation
Features detection
Facial landmarks identification
Current frame retrieving
Landmarks detection searching around previous landmarks neighbour
Assessment of changes in geometrical relations between landmarks
Estimation of head direction shift
Human-computer interaction
Analyze of head motion trajectory
Generation of command for computer application

Fig. 1. Scheme of system.

Thus, head motion can be tracked and analyzed to generate command for computer. Below each stage of developed system is considered in detail.

Face segmentation algorithm

Accordingly the above scheme, the first module in our system is used for face segmentation. Described in detail in early works algorithm [1] is based on skin colour. A group of colour elements for modeling of skin pixels was chosen among 26 considered attributes from 7 color spaces and models. For this purpose three training video sequences captured under changing lighting conditions were processed. It was shown that based on the chosen attributes a skin colour pixels can be detected independently on persons and varying (in certain range) lighting condition. The pixel is classified as skin if:

$3 < H < 13$ and
$18 < A < 28$ and
$15 < M\text{-}S < 18$ and
$10 < C\text{-}S < 13$,

where H is Hue (HLS), A is Achromatic response, C is Chroma, M is Colourfulness, S is Saturation (CIECAM02) [6].

After detection, all pixels are grouped into areas of interest using nearest neighbor method. And finally regions are verified based on edges density of brightness (CIECAM02) inside a particular region.

Computer simulation has shown high performance of developed algorithm. Among all images (n=180) used for simulation the number of false positive region is 0, false negative is 0.04±0.02. Image collection was created under lamplight with two persons of different age, gender and race [1]. Resolution of captured video was 640x480 px.

In this work described algorithm of segmentation was tested on the images from public available video database [7, 8] with unknown lighting condition.

This database’s frames have resolution 320x240 pixels. During capturing people moved their heads and movement was measured using "Flock of Birds" magnetic device mounted on head. Distance from subject to camera is about 86 cm [6, 7].

In our study for testing 30 frames of three persons was used. The face area was segmented in 100% cases. The example of face detection is shown in Fig. 2.

[1] Support of the Russian Foundation for Humanities, grant N 09-06-95218 a/F

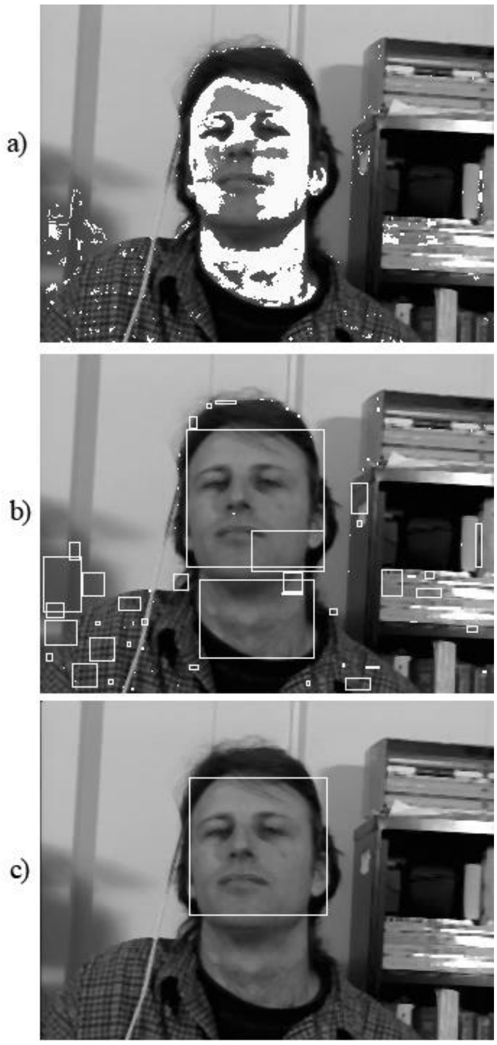

Fig 2. Face image segmentation: a) marked up skin colour pixels, b) pixels, grouped into potential facial areas, c) verified face area.

Facial landmarks detection

Eye corners, nose tip and the point under nose are chosen as facial landmarks in the consideration that they are relatively constant local features from viewpoint of context description and facial expression. In computer implementation, feature description of each facial landmark is formed by space-variant input window [3] and represented by the set of histograms (n=49). Each histogram is calculated based on local edges of Achromatic response (CIECAM02) around vicinity of each of 49 input window nodes A_i, i=0, 1…48. Local colour edges are extracted by using Sobel method [2].

In this work the performance of feature detection method was estimated based on public available video database described in Sec. 2 (n=30) [6, 7]. To estimate precision, the distance between manually marked up and automatically detected points was considered. The average distance was 1.7±0.15 px. The example of detected feature points is presentenced in Fig. 3a,b.

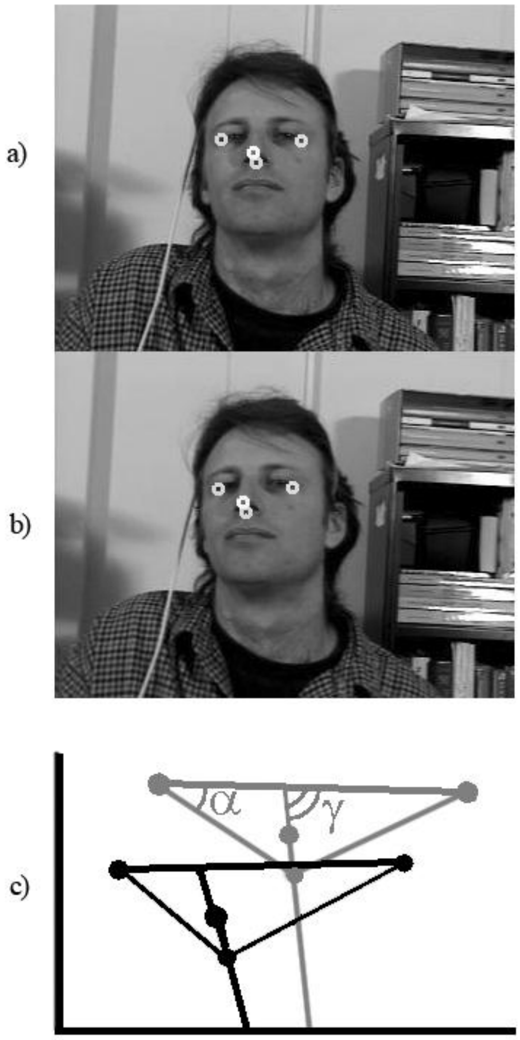

Fig. 3. Detected facial landmarks on first (a) and consequent (b) images; (c) - the scheme of direction estimation.

Head pose estimation

Angular relations between detected facial points on both current and previous video frames are considered to estimate direction of head movements if presented (Fig. 3).

Since, during projection on an image plain the information about depth of scene is lost, it is impossible to estimate the position of head in 3D world coordinate system based on

[1] Support of the Russian Foundation for Humanities, grant N 09-06-95218 a/F

landmarks coordinates on the image plain. However, direction of movement can be detected and estimated based on angular relations between facial feature points on two consequent images (Fig. 3).

To choose parameters of the image landmarks angular relations, which corresponds to head movement parameters, few angular parameters was considered on the facial video database with known movement (see description above). On the Fig. 4 an example of dependence parameters from head movement is shown. It is obvious that dynamic of Yaw corresponds to Alpha better than to Gamma (Fig. 3). (See Fig. 3c). Therefore, considering change of the Alpha angle on a consequent image it can be concluded did user moved his head in Yaw direction. By the same way an angular parameters which corresponds to Pitch and Roll were chosen.

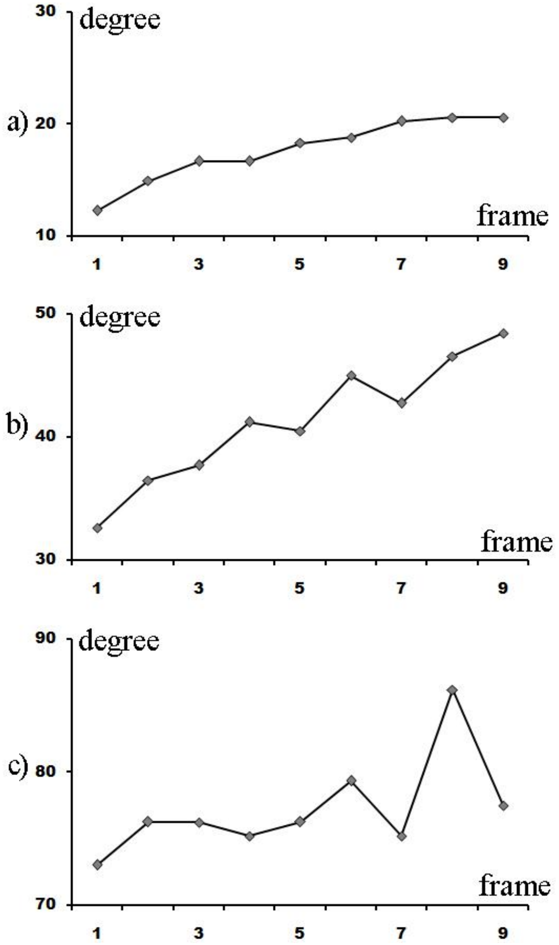

Fig. 4. Dynamics of changes parameter depend on head movement: a) Yaw, b) Alpha (Fig. 3c), c) Gamma (Fig. 3c).

The system testing

Three video sequences with 30 frames with three persons were used for system testing. Computer simulation was performed using Laptop Asus U3S with Intel(R) Core(TM) 2 Duo 2,2 GHz processor and 2,5 Gb RAM. Average processing time of one frame from video database described above is 147±2 msec. Thus the system is able to process up to seven frames per second. Direction of head movement was estimated correctly in 95% cases.

Conclusion

In our study a system for Human-Computer Interaction based on estimation of head movements is presented. Among parameters of angular relations between facial landmarks was chosen ones which can characterize head movement direction. The developed software is optimized to process up to seven frames per second. Future work includes integrating described system into computer game to drive the game using video information about user.

Reference

1. Anishenko S., Shaposhnikov D., Comley R., Gao X. Facial image segmentation based on mixed colour space. // In Proc. XV Int. Conf. on Neurocybernetics, 2009, Rostov-on-Don, Russia. - v.2. - p. 231-234.
2. Bakaut P.A, Kolmogorov G.S, Image segmentation: detection of area’s edges. // Foreign electronics. – 1987. -10. – p.25-47. (in Russian)
3. Gao X.W., Anishenko S., Shaposhnikov D., Podlachikova L., Batty S., Clark J. High-precision detection of facial landmarks to estimate head motions based on vision models. // J. Comp. Sci., 2007. - 3 (7). - p.528-532.
4. Jaimes A., Sebe N. Multimodal human-computer interaction: A survey. // Computer Vision and Image Understanding. - 2007. -108 (1-2). – p. 116-134.
5. Majaranta P., Räihä K.-J. Text entry by gaze: Utilizing eye-tracking. // In I.S. MacKenzie and K.Tanaka-Ishii (eds.), Text Entry Systems: Mobility, Accessibility, Universality, 2007. – p.175-187.
6. Mark D. Fairchild. Color appearance model. - 2004
7. http://www.cs.bu.edu/groups/ivc/HeadTracking
8. Cascia E. L., Sclaroff S., Athitsos V. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. // Pattern Analysis and Machine Intelligence, 22(4), 2000.

[1] Support of the Russian Foundation for Humanities, grant N 09-06-95218 a/F