HEAD MOTION MONITORING BASED ON FOVEAL APPROACH AND LOCAL FACIAL LANDMARK DETECTION

S. Anishenko[1] D. Shaposhnikov[1] L. Podladchikova[1]
R. Comley[2] K. Sukholentsev[2] X. W. Gao[2]

[1]A.B.Kogan Research Institute for Neurocybernetics, Southern Federal University, Rostov-on-Don, 344090, 194/1 Stachka Ave., Russia, dima@nisms.krinc.ru
[2] School of Computing Science, Middlesex University, London, NW4 4BT, UK, X.Gao@mdx.ac.uk

Algorithms and procedures to monitor human head motions based on three vision models (CIECAM'97, RETINA, and BMV) and detection of local facial landmarks are presented. They are evaluated on video image sequences monitoring the head positions of subjects (n=5) with different skin colours. The range of illumination conditions for high performance of the developed algorithms has been determined (more than 30 cd/m^2). At this illumination level, face segmentation and facial landmark detection are correctly performed (p = 1 for segmentation, p = 0.98 for landmark detection).

Introduction

Estimation of human head motions is important for many practical tasks, including medicine applications [3-5, 8, 10, 13-17]. In particular, head motions can significantly degrade the quality of functional brain studies [4, 8, 10, 14, 16]. Hence, motion tracking and correction is necessary to preserve image resolution and to ensure that quantitative tomography data are applied as accurately as possible. The known methods [4, 8, 10, 15, 17] to reduce the degrading effects of motion fall into several categories: image realignment in frame-mode or list-mode acquisition, optical tracking systems, motion restriction devices mounted on a patient's head, and their combinations. Up to now, optical tracking systems, i.e., video camera based systems has proved better in terms of simplicity. However, the accuracy of the detection of motion parameters still post a challenge. One of the prospective approach to solve the similar problem is concerned with the application of biological vision mechanisms, especially the Foveal Systems [2, 3]. The Foveal Systems imitate space-variant visual acuity, changing from the center of the retina (the fovea) to its periphery, and controlling attention mechanisms for human gaze while image viewing. In our study, a new approach to develop a motion correction system for brain tomography based on three biologically motivated models is proposed. One model is CIECAM [6, 9] for measuring colour appearance invariantly to illuminating conditions. The second one is a Behavior Model of Vision (BMV) simulating some mechanisms of the human vision system for perceiving shapes [11]. The third model is a simplified retina-like neural networks model for motion detection [12]. These models are used for colour segmentation of facial area on initial pictures, detection of Local Facial Landmarks, LFL (external eye corners and middle point of nose basement), and motion moment determination respectively. Earlier [1, 7], the basic algorithms and overall system architecture have been described in details. In this paper, some modifications of this system are presented based on the results of computer simulation on processing video image sequences monitoring subjects with different skin colours at various illuminaton conditions.

Basic algorithms and procedures

The proposed system includes following basic modules (Fig. 1). The input module consists of two digital cameras to monitor the head movements for a subject. Before the shooting, cameras are calibrated. The first captured image of the subject are segmented to determine face area and then segmented facial image fragments are processed to detect LFLs (Fig. 2). After that the coordinates of LFLs detected on initial images are fed into the module of head movements detection to locate the receptive fields (RF) of their elements on corresponding image points (Fig. 3, b). These video images are processed by motion detection module in real-time. If a motion is detected on any picture of video sequences, the picture is stamped at this time moment with a signal to determine the new positions of LFL being generated. A set of landmark positions for each motion moment during video session is stored inside the system to calculate motion parameters.

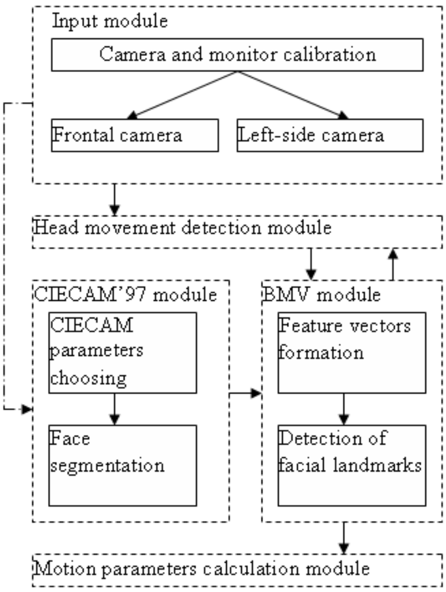

Fig. 1. Overview of the system for head motion detection and analysis.

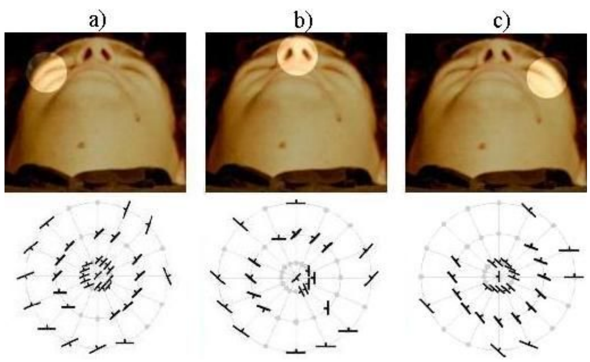

Fig. 2. Feature vectors of facial landmarks for a subject: (a) right eye corner, (b) middle point of nose basement, (c) left eye corner.

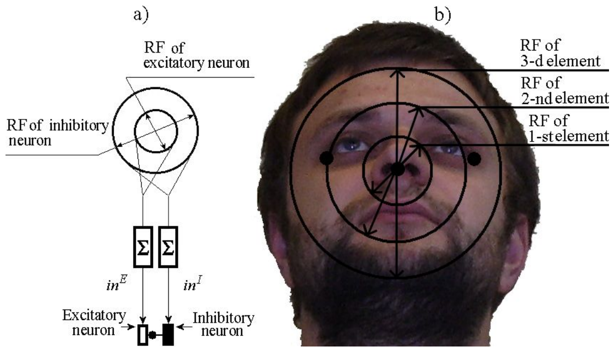

Fig. 3. An element of the retina-like model (a) and receptive field location for three excitatory neurons on a facial landmark (b).

In the computer implementation of the face segmentation algorithms and procedures, the values of colour attributes of hue, chroma, and lightness calculated by CIECAM, which are the property of facial skin colours, can be gained by using the averaged range for the given skin colours or by obtaining the colour range for the first image of a sequence of pictures for a given subject. These ranges are then applied as a threshold to segment the head from the rest of the pictures.

According to the preliminary evaluation, external eye corners and the middle point of a nose basement are chosen as LFL, which is in the consideration that they have a set of relatively constant local features. In the computer implementation, feature description of each LFL (see Fig. 2) is formed by space-variant input window and represented by a multidimensional vector. The vector components are the values of primary features detected in the vicinity of each of 49 input window nodes A_i, i=0, 1…48. Each component of the feature vectors is in line with the orientation of a local "colour" edge detected from colour attributes (lightness, chroma, and hue) that are extracted by convolving a map with a set of 16 kernels. Each kernel is sensitive to one of 16 chosen orientations. The whole set of 16 kernels are determined by differences between two oriented Gaussians with shifted kernels.

Initial feature description of each chosen LFL is obtained by positioning an input window manually in the center of a corresponding region on the first image only of the video sequence for each subject, and works as a template feature vector. Then all the subsequent images of the same subject are processed by the input window to search for image points with feature vectors similar to the template feature vector.

For the module of head movement detection, the basic elements apply a pair of excitatory and inhibitory neurons (Fig. 3,a) that have different sizes of their receptive fields (RF) and time delay similar to [12]. Three such pairs with embedded RFs are located on a center of each LFL.

Computer simulations

Two sets of image data have been employed in the computer simulation. The first image collection (n=12) with a subject lying down in the PET scanner includes the pictures with known head positions (measured by the built-in red laser beam of the PET scanner) and illumination conditions. All pictures monitoring a subject's head while simulating PET scanning have the same resolutions (640x427 pixels) and are recorded by the calibrated cameras. The second collection contains video image sequences of the subjects (n=4) with different skin colours under varying illumination levels. During a video session, the subjects performed voluntary head movements (rotation in plane and tilt up to 30°). Each image in this collection has a resolution of 640x480 pixels. The examples of face segmentation and facial landmark detection are presented in Fig. 4.

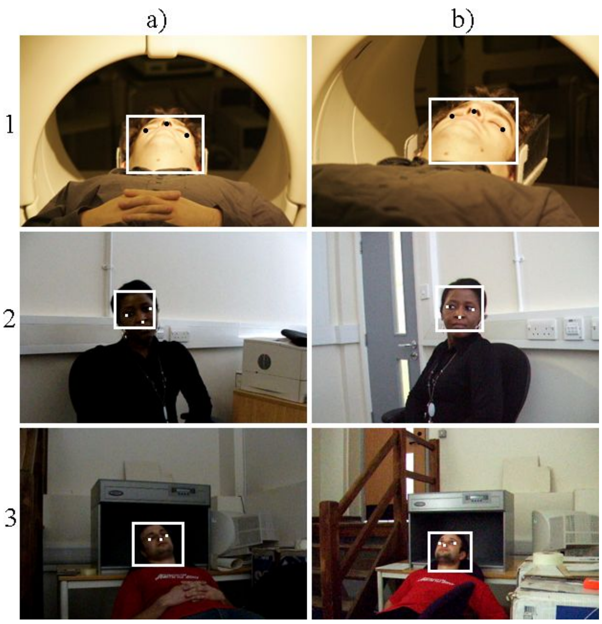

Fig. 4. The examples of face segmentation (face areas are marked by white squares) and facial landmark detection, a) frontal camera, b) left side camera. Illumination levels in (1), (2) and (3) are equal to 353, 78, 6 cd/m^2 correspondingly.

During the computer simulation it was evident that the performance of the algorithms for segmentation and detection depended heavily on the illumination level (Fig. 5). The probability of correct segmentation and landmark detection are equal to 1, 0.98 at high illumination levels compared to the low levels with 0.8, 0.7 respectively.

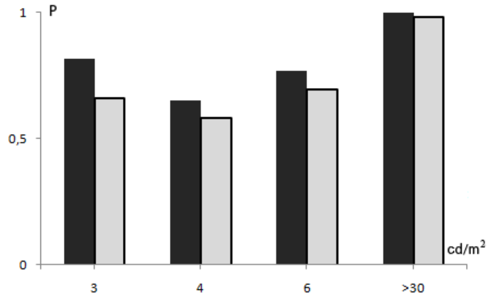

Fig. 5. The dependence of the probability of correct face segmentation (black columns) and landmark detection (gray columns) on illumination level.

The activities for three excitatory neurons of the retina-like model while processing video image sequences in real-time are presented in Fig. 6. It is seen that neuron sensitivity to head motions is dependent on its RF size.

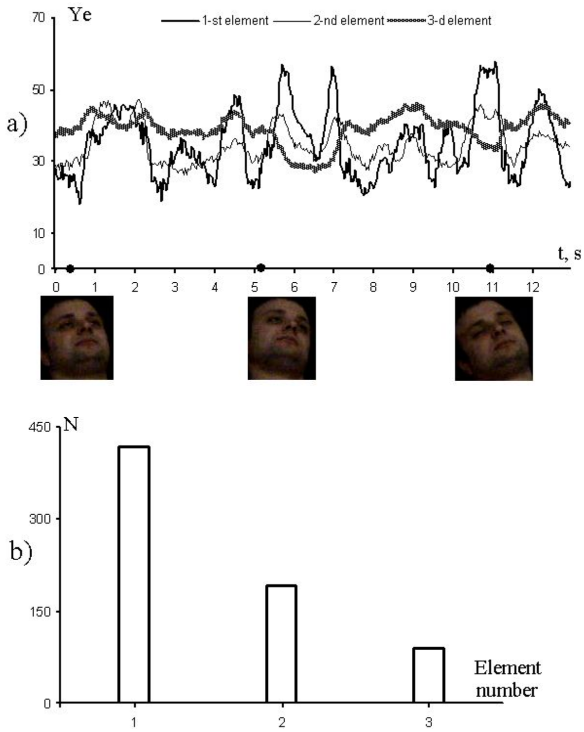

Fig. 6. Examples of activity for three excitatory neurons of the retina-like model while processing video image sequences in real-time: (a) dynamics of the output functions Ye for the neurons with minimal (10 pixels), middle (20 pixels) and maximal (30 pixels) RF radii. The pictures segmented during moments of sharp changes of Ye are shown in the lower row; (b) the number Ye values over the threshold.

Conclusion

Algorithms and procedures to monitor head motions based on three vision models (CIECAM'97, RETINA, and BMV) are presented. The developed algorithms have shown very good performance while processing video image sequences taken with illumination level over 30 cd/m^2. At this illumination level, face segmentation and LFL detection are correctly performed (p=1 for segmentation, p=0.98 for landmark detection, the exactness of landmarks location is equal to 1.45±0.85 pixels). Preliminary estimations indicate that the movement parameters can be obtained very accurately. In particular, rotation angle with less than 5° may be estimated by retina-like model. More accurate evaluation can be reached by analyzing spatial and angular relationship between local facial landmarks in-between pictures of a sequence. In the current implementation, processing time per a picture is equal to 150 ms, 1 s, and 1 s for retina-like, CIECAM`97 and BMV model respectively. It is proposed that some modification and optimizations on those algorithms may improve both system performance and computational cost.

Acknowledgement: This work is funded by the BRIDGE Research Cooperation Award from British Council, UK.

References

1. S. Anishchenko, V. Osinov, D. Shaposhnikov, L. Podlachikova, R. Comley, X.W. Gao. Toward a Robust System to Monitor Head Motions during PET Based on Facial Landmark Detection: a New Approach. // In proc. 21-st IEEE Int. Symp. On Comp.-Based Med. Sys., Finland, 2008. – 2008. – P. 50-52.
2. C. Bandera, P. Scott. Machine Vision for Pursuit. The Foveal Alternative? // J. of Electronic Defense. – 1991. - Vol. 14, iss. 10. – P. 61-66.
3. A. Benoit, A. Caplier. Biological Approach for Head Motion Detection and Analysis. // In Proc. 13-th European Signal Processing Conference EUSIPCO-2005, Antalya, Turkey, 2005.
4. P.M. Bloomfield, T.J. Spinks, J. Reed, L. Schnoor, A.M. Westrip, L. Livieratos, R. Fulton, T. Jones. The Development of a Motion Correction System in Neurological PET Using List Mode Data Acquisition and a Polaris Tracking System. // Neuroimage. – 2002. - No. 16. – P. 15.
5. P. Bühler, U. Just, E. Will, J. Kotzerke, and J. van den Hoff. An Accurate Method for Correction of Head Movement in PET. // IEEE Transaction on Medical Imaging. – 2004. – Vol. 23, iss. 9. - P. 1176-1185.
6. CIE, The CIE 1997 Interim Colour Appearance Model (Simple Version), CIECAM97s. // CIE TC1-34, April, 1998.
7. X.W. Gao, S. Anishenko, D. Shaposhnikov, L. Podladchikova, S. Batty, and J. Clark. High-precision Detection of Facial Landmarks to Estimate Head Motions Based on Vision Models. // J. of Computer Science. – 2007. - Vol. 3, No. 7. – P. 528-532.
8. B.F. Hutton, A.Z. Kyme, Y.H. Lau, D.W. Skerrett, R.R. Fulton. A Hybrid 3-D Reconstruction/registration Algorithm for Correction of Head Motion in Emission Tomography. // IEEE Trans. Nucl. Sci. – 2002. - No. 49. – P. 188-194.
9. M.R. Luo, and R.W.G. Hunt. The Structure of the CIE 1997 Colour Appearance Model (CIECAM97s). // Color Res. Appl. – 1998. -No. 23. – P. 138-146.
10. A. Rahmim. Advanced Motion Correction Methods in PET. Review Article // Iran J. Nucl. Med. - 2005. Vol. 13, No. 241. – P. 1-17.
11. I. Rybak, V. Gusakova, A. Golovan, L. Podladchikova, N. Shevtsova. Attention-guided Recognition Based on «What»; and «Where» Representations: A Behavioral Model. // Neurobiology of Attention, Elsevier, 2005. - P. 663-670.
12. N.A. Shevtsova, A.V. Golovan, L.N. Podladchikova, V.I. Gusakova, D.G. Shaposhnikov, A. Faure. Estimation of Motion Parameters by Retina-Like Neural Network Model. // Neurocomputers for Image and Signal Processing, Springer Verlag GmbH, 2008 (in press).
13. T.R. Steger and E. F. Jackson. Real-time Motion Detection of Functional MRI Data // J. of Applied Clinical Med. Phys. – 2004. – Vol. 5, No 2. – P. 64-70.
14. L. Tellmann, R. Fulton, U. Pietrzyk, I. Nickel, I. Stangier, O. Winz, H. Herzog. Concepts of Registration and Correction of Head Motion in Positron Emission Tomography. // Z Med Phys. -2006. - Vol. 16, No. 1. -P. 67-74.
15. H. Watabe, K. Koshino, P.M. Bloomfield, R.R. Fulton, H. Iida. Development of Motion Correction Technique for PET Study Using Optical Tracking System. // International Congress Series. – 2004. – No.1265. – P. 31 – 38.
16. R.P. Woods, S.R. Cherry, J.C. Mazziotta. Rapid Automated Algorithm for Aligning and Reclining PET Images. // J Comput Assist Tomogr. – 1992. -No. 16. – P. 620-633.
17. H. Zaidi, M.-L. Montandon, and S. Meikleb. Strategies for Attenuation Compensation in Neurological PET Studies. Review. // NeuroImage. – 2007. – No. 34. – P. 518 – 541.