High-precision Detection of Facial Landmarks to Estimate Head Motions Based on Vision Models

[1]Xiaohong W. Gao, [2]Sergey Anishenko, [2]Dmitry Shaposhnikov,
[2]Lubov Podlachikova, [2]Stephen Batty, and [3]John Clark
[1]School of Computing Science, Middlesex University, London, NW4 4BT, UK
[2]A.B. Kogan Research Institute for Neurocybernetics, South Federal University, Russia
[3]Wolfson Brain Imaging Centre, University of Cambridge, Cambridge, CB2 2QQ, UK

Abstract: A new approach of determination of head movement is presented from the pictures recorded via digital cameras monitoring the scanning processing of PET. Two human vision models of CIECAMs and BMV are applied to segment the face region via skin colour and to detect local facial landmarks respectively. The developed algorithms are evaluated on the pictures (n=12) monitoring a subject's head while simulating PET scanning captured by two calibrated cameras (located in the front and left side from a subject). It is shown that centers of chosen facial landmarks of eye corners and middle point of nose basement have been detected with very high precision (1 ± 0.64 pixels). Three landmarks on pictures received by the front camera and two by the side camera have been identified. Preliminary results on 2D images with known moving parameters show that movement parameters of rotations and translations along X, Y, and Z directions can be obtained very accurately via the described methods.

Key words: Face Segmentation, Skin Colour, Facial Landmarks Detection, Human Vision Models, Head Motion detection, PET

INTRODUCTION

Positron emission tomography (PET) is a relatively lengthy procedure with a typical scanning time up to one hour. It is therefore difficult for a subject to stay still during the data acquisition period. However even relatively small head motions may significantly degrade the image resolution, and hence the quantitative accuracy of PET brain studies[1-5]. In addition, head motion also causes misalignment between the emission and transmission scan data, leading to erroneous correction for photon attenuation. Consequently, motion tracking and correction is necessary to preserve brain image resolution and to ensure that quantitative data are preserved as accurate as possible.

Physically constraining the head during PET brain imaging, for example strapping the head tightly to a headrest can be uncomfortable especially for a long period of data acquisitions, and does not completely eliminate the head movement [6]. Therefore, post-acquisition methods have been developed to reduce the degrading effects of motion. These methods fall into two categories: image realignment; and raw data reorientation prior to image reconstruction. Optimal correction for motion requires accurate determination of the motion parameters (rigid body motion is assumed) and accurate reorientation of the raw data based on these parameters, which in turn demands determination of motion parameters accurately, mainly three translation and three rotation parameters.

In this study, a new approach to detect the head motions is developed through the study of images recording the scanning situation of a subject's head externally using off-the-shelf digital cameras. Image processing is based on two biologically motivated vision models. One is CIECAM97 [7, 8] for measuring colour appearance simulating human's vision of perceiving colours. The other vision model is BMV model (Behaviour Model of Vision) imitating some mechanisms of the real visual system for viewing shapes [9, 10]. These models are used for segmentation of face regions on the pictures of the subject's head and

---

Corresponding Author: Xiaohong W. Gao, School of Computing Science, Middlesex University, London, NW4 4BT, UK

for the detection of local facial landmarks on the segmented regions respectively, which is built on the previous study[11-13] presenting relatively low performance.

METHODOLOGY

Image collection: A system consisted of two cameras is used to monitor the head movement on a subject performing a PET brain scan. The camera is Canon EOS-1D Mark II, which has 28.7 x 19.1 mm CMOS (Complementary Metal-Oxide Semiconductor) and 8.2 million effective pixels (3504 x 2336 pixels). Before the shooting, cameras and colour monitors are calibrated as detailed in [12].

The pictures (n=12) with known subject head positions (measured by the build-in red laser beam of the PET scanner) and illumination conditions have been collected to test the developed algorithms. All pictures have the same dimension (640x427 pixels). Fig. 1 illustrates the head position during the scanning.
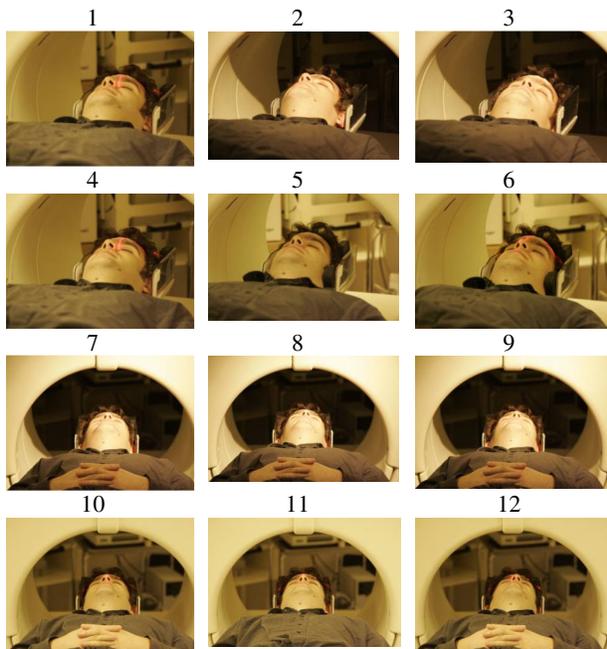

Fig. 1: Examples of PET brain scanning. Viewing parameters: (1-3) left camera, distance 1.175 m, lightness 353 cd/m^2; (4-6) left camera, distance 1.175 m, lightness 32.7 cd/m^2; (7-9) front camera, distance 2.350 m, lightness 353 cd/m^2; (10-12) front camera, distance 2.350 m, lightness 32.7 cd/m^2

Segmentation of Face Region based on skin colour: As seen from Fig. 1, the surrounding colour of the head is very close to the colour of face, which is the main reason why colour appearance model of CIECAM [7,8] is applied to segment head from the rest of scene due to the face that CIECAM model can predict a colour as accurate as an average observer. In addition, the following steps are implemented for the fast processing of images:

1) Images can be quickly classified using the initial estimation of colour attributes (hue, chroma, and lightness) calculated by CIECAM in the squares (20x20 pixels) located in the left and right top corners of the images as described in [13].

2) Colour range of each attribute can be obtained by operating in smaller area (11x11 pixels) on two images only (numbers 2 and 8 in Fig. 1). These colour ranges are then utilised to segment other pictures for the same subject.

Detection of local facial landmarks: After segmentation of face regions, the positions of local facial landmarks on segmented image are estimated. Basic algorithms for feature description of chosen facial landmarks are similar to those for detection of the most informative facial regions (i.e. eyes, nose, and mouth). Feature description of each facial landmark is provided by space-variant input window (IW) as demonstrated in Fig. 2, and is represented by multidimensional vector $\vec{F}$. The vector consists of components with values of primary features detected in the vicinity of each of 49 IW nodes A_i , i=0, 1…48.
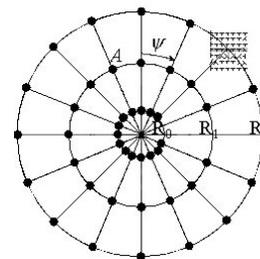

Fig. 2: Structure of space-variant input window with detailed description of context area around each nodes (one of them is shown as rectangular grid)

In comparison with [12, 13], the algorithms for description of features by IW and detection of local facial landmarks used in this study depict the following characteristics:

(1) According to the preliminary tests, eyes corners and middle point of nose basement are chosen as facial landmarks in the consideration that they have a set of relatively constant local features and are visible on all head pictures (Fig. 1).

(2) Each component of feature vectors is in line with the orientation of local "colour" edge detected from colour attributes, which 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 as calculated using equations 1 and 2.

$$G^{-}_{pq\varphi_k} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(p-\sigma\cos\varphi_k)^2 + (q-\sigma\sin\varphi_k)^2}{2\sigma^2}\right) \quad (1)$$

$$G^{+}_{pq\varphi_k} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(p+\sigma\cos\varphi_k)^2 + (q+\sigma\sin\varphi_k)^2}{2\sigma^2}\right) \quad (2)$$

Note σ is a root mean square deviation of the Gaussian distribution, φ_k is the angle of the Gaussian rotation, φ_k = k*22.5°, k = 0-15, $p \in [-3; 3]$, $q \in [-3; 3]$.

$$CE_{ij\varphi_k} = \sum_{p,q} C_{i-p, j-q}\left(G^{-}_{pq\varphi_k} - G^{+}_{pq\varphi_k}\right) \quad (3)$$

Note C_{ij} are the maps of colour attributes of lightness, chroma, and hue. The maximum response of all 16 kernels $CE_{ij\varphi_k}$ defines the contrast magnitude of a local edge at its pixel location, which in turn determines the orientation of a local colour edge.

(3) Initial estimations of each chosen landmark, including eye corners and middle point of nose basement for each subject, is obtained by positioning single IW in the corresponding region centers on one image only and works as template feature vectors. Then all consequent images of the same subject are scanned by IW to search image points with feature vectors similar to the template feature vectors, which can then be compared using equation 4 :

$$K^b = \sum_{i=0}^{i=48} th\left(\frac{1}{16} \sum_{j=0}^{j=15} Sg\left(Or_{ij}^b; Or_{ij}^{rw}\right)\right)$$

$$Sg(x; y) = \begin{cases} 1, & \text{if } x > 0 \text{ and } y > 0; \\ 1, & \text{if } x = 0 \text{ and } y = 0; \\ 0, & \text{in other case}; \end{cases} \quad (4)$$

$$th(x) = \begin{cases} x, & \text{if } x > threshhold; \\ 0, & \text{in other case}; \end{cases}$$

Note Or_{ij} is the density of j-th *colour* orientation in context area for the i-th IW node; whilst index b denotes template vectors, and index *rw* the current vectors; *threshold* in equation 4 is equal to 2, empirically defined.

Algorithms of estimation of head motions: Preliminary algorithm to estimate translational and rotational parameters of head motions (6 parameters in total along X, Y, and Z directions) is via the evaluation of angular and spatial relations between local facial landmarks detected on 2D consecutive pictures captured using each camera (Fig. 4). Dynamic analysis of linear and angular parameters of triangle (in the case of detection of 3 local facial landmarks) during PET scanning can give rise to information on all six motion parameters, whilst in the case of two detected landmarks, motion parameters to be estimated are significantly limited up to one translation and two rotation parameters.

RESULTS

Examples of segmented face regions are given in Fig. 3, showing that CIECAM model provides face segmentation accurately. These regions are then segmented for the detection of local facial landmarks.

The detection of the local facial landmarks on the segmented images is presented in Fig. 4. During preliminary evaluation, the maximum value of K^b (more than 35) calculated via equation 4 is chosen as facial landmarks. The performance of this method in terms of processing time, exactness, and the number of false positive regions, is significantly higher while processing segmented face regions than to the original pictures, given none false detection of landmarks while processing a group of images of twelve.

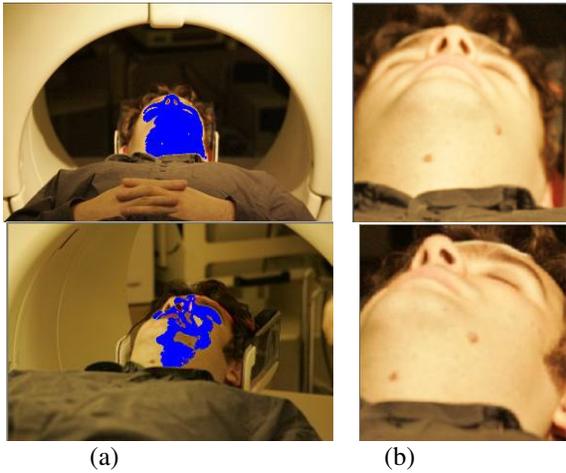(a)   (b)

Fig. 3: Examples of segmented face regions during PET scanning (N 11, N 4 in Fig. 1 correspondingly) captured by the front (upper row) and left side (bottom row) cameras. (a) results of coloured marks of the face; (b) image segmentation of (a) based on coloured marks (172x201 pixels for frontal view, 136x159 pixels for left view)
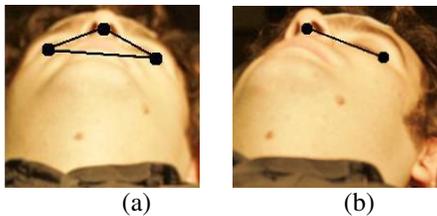

(a)   (b)

Fig. 4: Examples of detection of facial landmarks on the pictures captured by the front (a) and left side (b) cameras. Identified landmarks are marked by black spots

By and large, all the visible landmarks (n=30) have been identified (3 landmarks on pictures captured by the front camera and 2 by the side camera). When compared with pictures with pre-known movement parameters, the identified facial landmarks (i.e. eye corners and middle point of nose basement) give very high precision with deviations of 1 ± 0.64 pixels.

After the detection of facial landmarks, the movement parameters including translation and rotation along each of X, Y, and Z directions can be figured out by comparison of landmarks from two different images (normally two consecutive images) of the same subject doing the same scan. Preliminary estimations on 2D

images with known moving parameters show that movement parameters can be obtained very accurately via the described methods as illustrated in Fig. 5. In current implementation, the processing time (colour segmentation and facial landmark detection) varies from 1 up to 2 seconds per picture on a standard Pentium.IV PC.
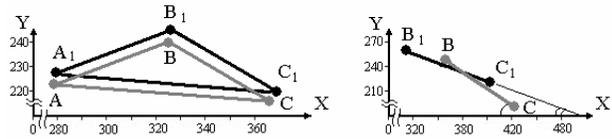

Fig. 5: Spatial relations between facial landmarks detected on processed pictures: (a) pictures captured by frontal camera, (A, B, C and A_1, B_1, C_1 - landmarks identified on pictures N 11 and N 8 in Fig. 1 correspondingly); (b) picture captured by left camera (B, C and B_1, C_1 – landmarks identified on pictures N 1 and N 3 in Fig. 1 correspondingly). The pixels of initial pictures (N 11 in (a) and N 1 in (b)) are presented on X and Y axes

CONCLUSIONS AND DISCUSSION

In this study, a new approach to monitor the head motions based on two biologically motivated vision models, i.e., CIECAM97 [7,8] and BMV [9,10], has been developed. These two models are applied for colour-based segmentation of face regions and for detection of local facial landmarks respectively, which is built on the previous study [12,13] to enhance precision. In particular, centers of chosen facial landmarks, including eye corners and middle point of nose basement, are detected with very high precision with deviation about 1 pixel, in contrast to 3 pixels from the previous study. Preliminary estimations on 2D images with known moving parameters show that movement parameters (i.e. rotation and translation along X, Y, and Z directions) can be obtained very accurately via analysis of spatial and angular relations between identified facial local landmarks. In the future, the detected motion parameters will be utilized in the process of PET brain image reconstruction, which requires list-mode (event-by-event data acquisition) data.

This algorithm of estimation of head motion parameters is sensible to the accuracy of local facial landmarks, which however does not pose any problem

as the detection of landmarks is robust by using the two vision models.

REFERENCES

1. Bloomfield, P., T. Spinks, , J. Reed, L.Schnoor, A. Westrip, L. Livieratos, R. Fulton, and T. Jones, 2002. The Development of a Motion Correction System in Neurological PET Using List Mode Data Acquisition and a Polaris Tracking System, Neuroimage, 16: S15.
2. Hutton B., A Kyme., Y. Lau, D. Skerrett, and R. Fulton, 2002. A Hybrid 3-D Reconstruction/Registration Algorithm for Correction of Head Motion in Emission Tomography, IEEE Trans. Nucl. Sci., 49: 188-194.
3. Ruttiman U., P. Andreassen, and D. Rio, 1995. Head Motion During Positron Emission Tomography: Is it Significant?, Psychiatry Res: Neuroimaging, 65: 41-43.
4. Tellmann L, R. Fulton, U. Pietrzyk, I. Nickel, I. Stangier, O. Winz, and H. Herzog, 2006. Concepts of Registration and Correction of Head Motion in Positron Emission Tomography, Z Med Phys., 16 (1): 67-74.
5. Woods R., S. Cherry, and J. Mazziotta, 1992. Rapid Automated Algorithm for Aligning and Reclining PET Images, J Comput Assist Tomogr., 16 : 620-633.
6. Green M., J. Seidel, S. Stein, T. Tedder, K. Kemper, C. Kertzman, and T. Zeffiro, 1994. Head Movements in Normal Subjects during Simulated PET Brain Imaging with and without Head Constraint, J Nucl Med, 35: 1538-1546.
7. IEC, 1999. International standard, Part 2-1: Colour Management – Default RGB Colour Space – sRGB, *IEC 61966-2-1*, first edition.
8. Luo M., and R. Hunt, 1998. The structure of the CIE 1997 Colour Appearance Model (CIECAM97s), Color Res. Appl., 23: 138-146.
9. Rybak, I., V. Gusakova, A. Golovan, L. Podladchikova, and N. Shevtsova, 2005. Attention-guided Recognition Based on «What»; and «Where» Representations: A Behavioral Mode. In: The First Encyclopedic Volume on Neurobiology of Attention (eds. L. Itti, G. Rees, and J. Tsotsos) pp. 663-670. Elsevier Academic Press.
10. Shaposhnikov D., L. Podladchikova, and X. Gao, 2003. Classification of Images on the Basis of the Properties of Informative Regions, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, 13 : 349-352.
11. Gao X., S. Batty, J. Clark, T. Fryer, and A. Blandford, 2001. Extraction of Sagittal Symmetry Planes from PET Images. Proceedings of International Conference on Visualization, Imaging, and Image Processing, Spain, pp: 428-433.
12. Gao X., S. Batty, L. Podlachikova, D. Shaposhnikov, and J. Clark, 2007. Detection of Head Motions Using a Vision Model. Proceedings International Conference on Telehealth, Canada, pp: 167-171.
13. Gao X., D. Shaposhnikov, S. Anishenko, and S. Batty, 2007. Human Face Determinants to Detect Head Movements during PET. Proceedings of Conference on Neuroinformatics-2007, Russia, pp: 102-109.