Proceedings of the Third IASTED International Conference
TELEHEALTH
May 31–June 1, 2007, Montreal, Quebec, Canada
ISBN Hardcopy: 978-0-88986-667-6/ CD: 978-0-88986-668-3

DETECTION OF HEAD MOTIONS USING A VISION MODEL

Xiaohong W. Gao, Stephen Batty, Lubov Podlachikova[1], Dmitry Shaposhnikov[1], John Clark[2]

School of Computing Science, Middlesex University, London, NW4 4BT, UK
[1]A.B. Kogan Research Institute for Neurocybernetics, Rostov State University, Russia
[2] Wolfson Brain Imaging Centre, University of Cambridge, Cambridge, CB2 2QQ, UK.
x.gao@mdx.ac.uk

ABSTRACT

Head movement during a brain scan is one of the major reasons causing the blurriness of images. In order to improve the quality of scanned brain images, the head movement parameters should be known and embedded into the image reconstruction algorithms. In this study, a system consisted of two cameras is studied to monitor the head movements under the scanning conditions. The camera is calibrated before the shooting. Images are acquired by the video camera and transferred to a colour monitor that has been calibrated. The range of face skin colours are then obtained based on these images captured under varying lighting conditions to perform face segmentation. The vision model, Foveal System for Face Images (FOSFI) is then applied to find the position of eyes and nose, arriving at the findings of parameters of rotation and translation of head motion. Preliminary results on 2D images with known moving parameters show that movement parameters can be obtained very accurately via the described methods.

KEY WORDS
Motion detection, modelling, PET imaging

1. Introduction

Positron emission tomography (PET) is an imaging modality that can quantitatively measure *in vivo* physiology, biochemistry, and pharmacology. However, to acquire a set of PET data it usually takes about 1 hour, which sometimes is difficult for a subject to stay still that long, even if the subject is physically constrained. This is particularly true for patients who are imaged whilst suffering some discomfort from their medical conditions, such as acute stroke and hydrocephalus. Consequently, head motion significantly degrades the quality of PET studies of the brain [1, 2]. Even relatively small motions may significantly compromise the resolution, and hence the quantitative accuracy of the image data. In addition, head motion also causes misalignment between the emission and transmission scan data, leading to erroneous correction for photon attenuation. Hence, motion tracking and correction is necessary to preserve image resolution and to ensure that quantitative data corrections are applied as precisely as possible.

Physically constraining the head during PET brain imaging, for example strapping the head tightly to a headrest can be uncomfortable especially for long data acquisitions, and does not completely eliminate head movement [1]. 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. Therefore, knowing motion parameters correctly is very important.

In this study, a system consisted of two cameras is proposed to monitor the head movement with a subject lying down in the PET scanner with known head movements.

2. Methodology

Figure 1 illustrates the setup of scanning situation. The red laser beam is applied at the beginning of the scan to make the head position in the middle of the scanner.

In order to find out the movement of the head, the location of face is segmented based on the face skin colour. This method requires the accuracy of the colour reproduction. Camera calibration and colour monitor calibration are therefore conducted at first.

Figure 1. The setup of PET scanning position.

2.1 Calibration of Camera Canon EOS-1D Mark II

Canon EOS-1D Mark II is the digital camera investigated in this study, which has 28.7 x 19.1 mm CMOS (Complementary Metal-Oxide Semiconductor), and 8.2 million effective pixels (=3504 x 2336 pixels). Figure 2 illustrates the steps of camera calibration.

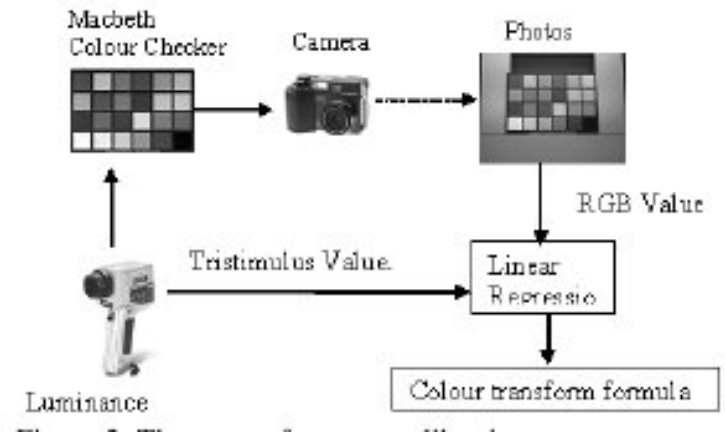

Figure 2. The steps of camera calibration.

Firstly, the camera is applied to take pictures of 24 colour patches from Macbeth Colour Checker. Then these images are transferred to a colour monitor that has been calibrated in advance under the same illuminant. For each colour pattern, the middle part of the image with 200 x 200 pixels is used to work out the mean values of sRGB that is the colour space setup for the camera. Image processing software MATLAB is utilized doing the processing. sRGB is a colour space and is device dependent [4]. Therefore it is possible to transfer RGB values (in sRGB space) into CIE tristimulus values, X. Y, and Z. Only the middle part of each image is chosen which is in the consideration that some cameras have images with four blurring corners. The 24 colour patches are then measured using colour meter CS-100A to obtain their CIE XYZ values. The two sets of XYZ values are then used to obtain camera calibration model by the least square method.

2.2 Monitor Calibration

In order to make sure that the same colour appears the same all the time, colour monitor needs to be calibrated. This is done by using optical software with Spyder sensor [5]. Once applied, the sensor is attached to the center of the monitor to send back the detected colour. The software then adjusts the colour attributes based on these measurement until predefined viewing conditions is satisfied, such as illuminant D65.

2.3 Skin Colour Segmentation

A subject is scanned under four light conditions simulating to the real scanning situation shown in Figure 3.

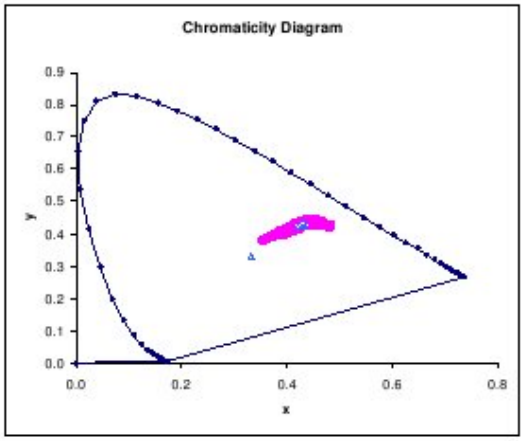

Figure 3. The chromaticity diagram showing skin colour with the dots situated together and the four lighting conditions with open triangles. The triangle in the centre is the equal energy point with (x, y)=(0.333, 0.333).

Based on this diagram, the range of skin colour is calculated using mean ± standard deviation. Any pixels with (x, y) values falling inside this range will be classified as skin colour, leading to the face being segmented. After segmentation, only face colours are reserved. The face is more of less symmetrical with reference to the line passing through the nose. Figure 4 shows the co-ordinate system for a head, which is used in this study to find out the three rotation angles, i.e., yaw, roll, and pitch angles.

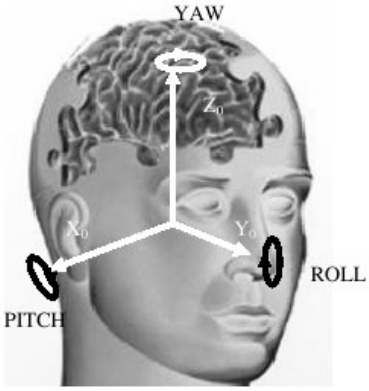

Figure 4. The co-ordinate system for detection of head movement.

After the segmentation of face colour, the background is given a neutral colour. Then the symmetrical line of the face is found out by means of correlation method, which is similar to that in [6] using Eq. (1).

$$CC(c_k, \theta_j) = \max\{\max\{XCorr(S_i, rot(2\theta_j, ref_v(S_i, c_k))\}\} \quad (1)$$

where CC is the cross-correlation value, θ_j is the yaw angle, and c_k the symmetrical line on image S_i. The *ref* (S_i, c_i) is the reflection image of S_i.

2.4 Determination of Eye, Nose, and Mouth Regions

The positions for informative facial regions, i.e. eye, nose, and mouth, are then determined by the application of vision model of Foveal System for Face Images (FOSFI) after face segmentation. Basic algorithms are similar to those developed earlier [7, 8]. Feature description in each facial area is provided by space-variant input window (IW) and is represented by multidimensional vector $\vec{F}$ of oriented segments extracted in the vicinity of each of 49 nodes of IW (Figure 5). Orientation and contrast of local segments are determined by means of calculation of the difference between two oriented Gaussians with spatially shifted centers. Space-variant representation is emulated by Gaussian convolution with different kernels depending on distance from the IW center (5×5 pixels for the central part of the IW, 7×7 - for the immediate, and 9×9 - for the peripheral part). On the other hand, oriented segments density in the context area of IW nodes is estimated (see Figure 5).

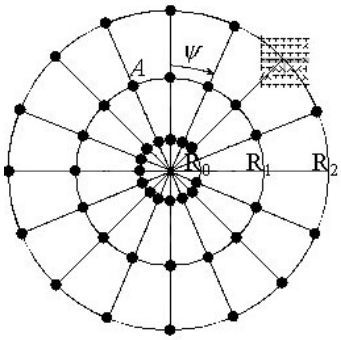

Figure 5. Structure of space-variant input window with detailed description of context area around each nodes (one of them is shown as rectangular grid).

For example, feature vector $\vec{F}$ is based on edge orientation α in the vicinity of each of 49 sensor nodes A_i , i=0, 1…48. Let $x_0 = X_0$, $y_0 = Y_0$ be co-ordinates of the central sensor node, then co-ordinates (x_i, y_i) of peripheral sensor node A_i , i=1, 2…48 can be determined as follows:

$$x_i = X_0 + R_l \cos\psi_k, \qquad y_i = Y_0 + R_l \sin\psi_k \quad (2)$$

where R_l , l=0,1,2 is the radius of l-th concentric circle of the sensor (R_0 = 3 pixels, R_1 = 9, R_2 = 15) and ψ_k = k·22.5°, k= 0,1…15 is the angle of the radiating line corresponding to the i-th sensor node.

Contrary to the cascade method for detection of informative facial regions developed earlier [7] specific prototype description for each region (eye, nose, and mouth) is received by single positioning IW in region centers for initial image like that presented in Figure 1. Then all consequent images are scanned by IW with the location of image points similar to prototype feature vectors being determined. Feature vectors are compared by Eq. 3:

$$K^b = \sum_{i=0}^{i<49} \left[\mathrm{sgn}\left(Or_i^b - Or_i^{rw} \right) \cdot \left(1 - abs\left(\rho_i^b - \rho_i^{rw} \right) \right) \right] \text{ where } \mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x = 0; \\ 0, & \text{in other case} \end{cases} \quad (3)$$

where Or_i is dominating orientation of contrast segment in context area for the given IW node (orientation is determined by step 22.5^o and denoted as 0, 1, 2, 3…., 15); whilst index *b* denotes prototype vectors, and index *rw* the current vectors. ρ is the density dominating contrast oriented segment in the context area of each IW node.

3. Results

Preliminary research has been done based on one subject under 4 lighting conditions to study horizontal movement, which includes translation horizontally and rotation in 2D images captured using one camera. Figure 6 shows an example of segmentation where the green colour around of the face represents the face boundary.

Figure 6. The face boundary with green colour segmented based on skin colour.

Figure 6 demonstrates that a face can be segmented using the skin colour. If the range setting is right, more face colour can be included even for the shadow parts. Figure 7 shows the segmented results of face with symmetrical line. By comparison with two symmetrical lines of two images, the rotation angles of raw and roll can be calculated. To find pitch angle as illustrated in Figure 4, the images taken from another camera should be analyzed together with the images from the first camera, which is still under investigation at the moment.

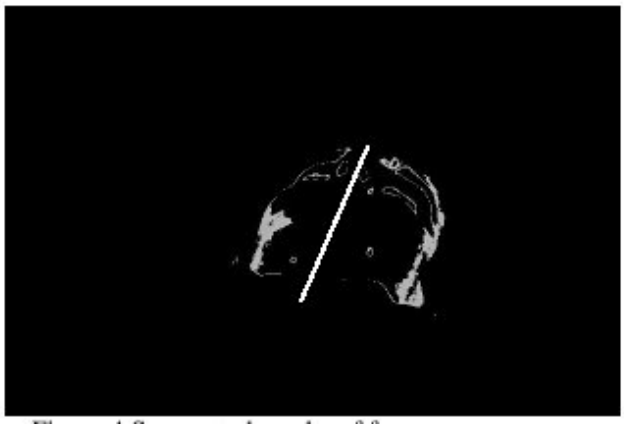

Figure 7. Segmented results of face with symmetrical line.

The K^b in Eq. (3) is varying during preliminary processing of all images. Its value resulted in maximal exactness to detect the informative facial regions is chosen (more than 35). The results of automated detection of informative facial regions are presented in Figure 8. For all used images, the variations for location of informative regions centers were no more than ± 3 pixels.

Figure 8. Detection of eye, nose, and mouth regions after face segmentation.

4. Conclusion

In this study, we investigate the feasibility to find head moving parameters using face skin colours and facial informative regions. Preliminary results show that this approach works alright in terms of finding rotation angles and translation distances.. However, the study is only based on one subject, which does not represent all the skin colours. In the future, we will include more subjects with different range of skin colours. Also we will investigate the 3D images captured using two cameras. It is expected that combination of approaches with colour segmentation and facial informative regions can improve face contour detection and determination of head movements with demand exactness.

Further work also includes incorporating parameters of motions into image reconstruction algorithms using list-mode data, which will be provided by GE Company.

Acknowledgements

This work is part of TIME (Tele-imaging in Medicine) project funded by European Commission under Asia-IT&C program (CN/ASIA-IT&C/009(91450)), and is also supported in part by Russian Foundation for Basic Research, Grant No. 05-01-00689, in Russia.

References

[1]. Green MV, Seidel J, Stein SD, Tedder TE, Kemper KM, Kertzman C, Zeffiro TA. Head movements in normal subjects during simulated PET brain imaging with and without head constraint. *J Nucl Med*, **35**: 1538-46, 1994.

[2]. Ruttiman UE, Andreassen PJ, Rio D. Head motion during positron emission tomography: Is it significant? *Psychiatry Res: Neuroimaging*, **65**: 41-3, 1995.

[3]. Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reclining PET images. *J Comput Assist Tomogr*, **16**: 620-633, 1992.

[4]. IEC, International standard, Part 2-1: Colour Management – Default RGB colour space – sRGB, *IEC 61966-2-1*, first edition, 1999.

[5]. http://www.colorcal.com/

[6]. Gao, X.W., Batty, S., Clark, J., Fryer, T., Blandford, A., Extraction of Sagittal Symmetry Planes from PET Images, *Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'2001)*, pp 428-433, ACTA Press, 2001.

[7]. Shaposhnikov D. G., Podladchikova L. N., Gao X. 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, 2003.

[8]. Rybak, I., Gusakova, V., Golovan, A., Podladchikova, L., Shevtsova, N. Attention-guided recognition based on «What»; and «Where» representations: A behavioral model. *The first encyclopedic volume on Neurobiology of Attention*. Eds. L. Itti, G. Rees, and J. Tsotsos. Elsevier, Academic Press, 2005, 663-670.