Facial Recognition techniques applied to the automated registration of patients in the emergency treatment of head injuries

M.Gooroochurn[a1], D.Kerr[a] , K.Bouazza-Marouf[a], M.Ovinis[a]

[a] *Loughborough University, Loughborough, UK*

Abstract

This paper describes the development of a registration framework for image-guided solutions to the automation of certain routine neurosurgical procedures. The registration process aligns the pose of the patient in the preoperative space to that of the intra-operative space. CT images are used in the pre-operative (planning) stage, whilst white light (TV camera) images are used to capture the intra-operative pose. Craniofacial landmarks, rather than artificial markers, are used as the registration basis for the alignment. To further synergy between the user and the image-guided system, automated methods for extraction of these landmarks have been developed. The results obtained from the application of a Polynomial Neural Network (PNN) classifier based on Gabor features for the detection and localisation of the selected craniofacial landmarks, namely the ear tragus and eye corners in the white light modality are presented. The robustness of the classifier to variations in intensity and noise is analysed. The results show that such a classifier gives good performance for the extraction of craniofacial landmarks.

Keywords – Preoperative to Intraoperative Registration, Registration Basis, Automated Landmark Extraction, Emergency Neurosurgery, Machine Vision, Neural Networks

INTRODUCTION

The research work presented in this paper concerns the development of a registration framework to enable image-guided solutions for certain routine neurosurgical procedures used in the emergency treatment of head injuries. The main motivation was to design the registration process so that it could be used quickly and

[1] Corresponding Author: M.Gooroochurn@lboro.ac.uk

conveniently in most typical Accident and Emergency (A&E) units, thus avoiding the need for costly and complex equipment like laser range scanners and optical tracking devices. In particular, this paper presents results for the fine localisation of the craniofacial landmarks selected for the registration basis of the alignment procedure.

The time frame in the aftermath of a traumatic head injury is critical since the promptness of intervention has a direct impact on the chances of recovery of the patient as well as on the long-term prognosis. Two common neurosurgical procedures employed in the event of a traumatic brain injury are Intracranial Pressure Monitoring (ICP) and External Ventricular Drainage (EVD). ICP is used to measure the pressure inside the cranium and EVD is used to drain off excess Cerebrospinal Fluid (CSF) from the brain ventricles when the pressure is deemed excessive. The need for the proposed registration framework stems from the complications inherent in the present management protocol for head injuries. To alleviate some of these shortcomings, image-guided solutions are often sought, which in turn require the setting up of a registration link between the patient on the operating table and any preoperative planning process.

The efficient and timely treatment of severe head injuries is often precluded by the lack of neurosurgical expertise at A&E units where most cases are first presented. Following stabilisation and consultation, patients may have to be transferred to a neurosurgical centre for further treatment. Whilst being inherently expensive, such transfers are risky for the patient, take up valuable time and reduce the chances of a satisfactory treatment outcome. To tackle this problem, an image-guided, mechatronic system is envisaged, which can be operated by the general surgical team available at most A&E units. This system is designed to “de-skill” the required neurosurgical procedures, so that the local surgical team can deliver rapid and effective treatment with the help of a remotely based neurosurgical specialist. The system would follow a preoperative plan determined by the remote neurosurgeon, using CT images taken by the local radiographer and sent via the internet.

As part of such an image-guided system, registration of the intra-operative patient space to the preoperative CT space is critical. Several methods exist for performing preoperative to intra-operative registration, including point-based [1], surface

matching [2, 3] and intensity-based methods [4, 5]. For most of these procedures, the conventional application of artificial markers to the patient's head puts considerable constraint on the diagnosis and treatment phases. On one hand, failure to attach markers at the first diagnosis stage means that the CT scan needs to be re-taken, should neurosurgical intervention be considered essential. On the other hand implanting markers by default as a prospective method may cause unnecessary extra time and stress for the patient should neurosurgical procedures be deemed unnecessary. Therefore, a retrospective and marker-less registration basis, using anatomical landmarks, offers optimality in the management of head injuries in terms of quick access to essential neurosurgical treatment.

The same anatomical landmarks need to be extracted from the CT images and the intra-operative camera images. In the white light modality, each of the landmarks is reconstructed from stereo camera views using photogrammetry. In this work, the required landmarks are obtained from frontal and profile views as well as from views midway between these at an azimuth angle of 45 degrees. Although extraction of eye features has been widely reported in the literature, ear features have often been neglected. The motives for using the ear as a biometric feature are its invariance to subject emotional state, rigid shape and size constancy over time. Ear features have been found to be of high discriminative value in biometric identification [6]. Research has been carried out in the use of the ear as a recognition biometric feature both in 2D [7] and 3D [8] applications.

Methods for localising craniofacial landmarks within an image have already been developed as part of a more general approach to face recognition. Such an approach typically consists of face localisation, facial feature extraction and face recognition [9]. Face detection and recognition can be treated as pattern classification problems, where statistical methods [10-12] and neural network solutions [13, 14] have tended to take priority over methods based on features extracted by grey-level processing. The reason for this shift in the design of face processing solutions is the ability to use large training sets for machine learning, which have given very good performance, whereas methods relying on grey-level processing have been found to be effective only when operating conditions such as illumination, scale and rotation are closely controlled.

The details of the proposed registration framework were described in [15], where the implications of using the eye corners and ear tragi as the registration basis were assessed via simulation studies and the analytical expression proposed by Fitzpatrick *et al*. [16]. Here, we present a brief overview of the registration framework, before concentrating specifically on how the craniofacial landmarks are imaged and then on the method and results obtained from the automated fine localisation of the selected landmarks. Although these results are as yet limited to simulation studies, data obtained from an artificial skull are used to predict the limitations that could be expected in practice and how the estimates obtained from the simulation study support a user-validated registration framework is discussed.

1 REGISTRATION FRAMEWORK

1.1 Camera Arrangement

The intra-operative component of the registration framework concerns the extraction and localisation in 3D space of the selected craniofacial landmarks. In [17], Ansari *et al.* used a camera set-up with two views (frontal and profile) to reconstruct the 3D coordinates of facial features. The features on the hidden side of the face were reconstructed based on symmetry. The reconstruction accuracy then hinges on how 'true' the frontal view is of the subject. This is because reconstruction by symmetry relies on defining an origin midway between two corresponding eye corners and computing distances with respect to this origin. Thus, not having an accurately aligned face in the frontal and profile view violates the symmetry criterion. The alternative is to use more camera views, so that each of the landmarks appears in at least two views to enable stereo reconstruction. The benefit of using more cameras is that the system does not need to be set up with high accuracy as long as the craniofacial features appear in the camera views and are within the designated calibrated space.

A major difference between the work in [17] and the 3D facial feature reconstruction task considered here lies in the use of the ear tragus as a feature. Thus an additional view is required midway between the frontal and profile views. A three camera system enables on this basis the reconstruction of the two outer and inner eye corners

and one ear tragus, provided that the system is accurately placed with respect to the patient. A schematic of this system is shown in Figure 1(a).

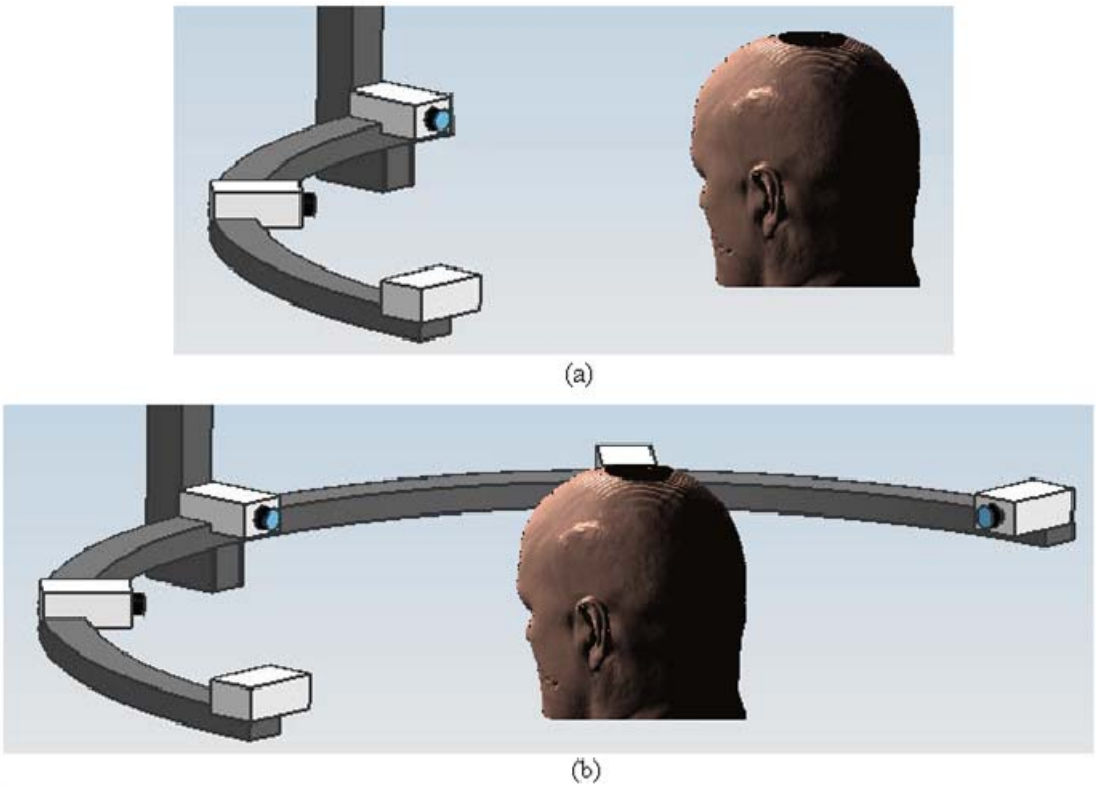

Figure 1: Schematic of Camera System (a) 3-camera system (b) 5-camera system

Alternatively, a four camera system, using two profile and two intermediate positions but excluding the frontal one, would permit the reconstruction of the two outer eye corners and both ear tragi. However as shown in the next section, a frontal view is still important for initial system placement, as aligning specific landmarks of the patient's face with datum lines in the frontal view sets the system working distance as well as largely constraining the roll and yaw of the head. Furthermore, implementation of a facial feature extraction methodology similar to [17] on the two sides of the head without using symmetry can also help to reconstruct the inner eye corners. Such a five camera arrangement shown in Figure 1(b) has been employed in the simulation studies described in [15], results of which are summarised later.

1.2 System Placement in the Operating Room

Registration must be performed such that the object(s) to be reconstructed lies within the calibrated volume as extrapolation outside that space can lead to large errors [18]. The use of an appropriately-sized object for calibrating the cameras sets a given field of view in each of frontal, profile and intermediate camera image planes for a given

working distance and optical parameters of the cameras. Thereafter using the calibrated cameras to make measurements in the scene requires that any image captured be delimited to fall within the field of view in each camera corresponding to that occupied during calibration. The placement of the camera system so as to adequately fill the calibrated space can be planned in the frontal and profile views.

The placement strategy has been described in [19]. It involves the use of cursor lines in the frontal and profile views, which are aligned to specific landmarks on the patient's face in order to ascertain that his/her head lies in the calibrated space. This is shown in Figure 2. The vertical line in the profile view is made to coincide with the nose tip whereas the horizontal line is made to coincide with the eye corner. In the frontal view, the vertical line is made to coincide with the centre of the nose and the horizontal line is made to pass through the two outer eye corners.

Figure 2: Datum cursors for the initial camera set-up. Left, frontal view and right, profile view.

1.3 Registration of the Preoperative to Intraoperative Spaces

Rigid registration of the data set obtained from the CT space and the corresponding data set computed in the intraoperative physical patient space results in a rotation and translation matrix. This solution in its closed form needs at least 3 corresponding 3D points between the two data sets. The transformation can be represented by the following equation:

$$^{OR}X = {}^{OR}R_{CT} \cdot {}^{CT}X + {}^{OR}P_{CTorg} = G({}^{CT}X) \quad (1)$$

^{CT}X is the data set containing the coordinates of landmarks found from the CT space and ^{OR}X is the data set corresponding to the landmarks determined in the Operating

Room (OR) set-up in white light modality. ${}^{OR}R_{CT}$ is the rotation matrix consisting of three unit vectors r_x, r_y and r_z, representing the rotation of the frame of reference of the CT space with respect to the frame of reference used intraoperatively, ${}^{OR}P_{CTorg}$ is the translation vector $[t_x, t_y, t_z]^T$, representing the position of the origin of the frame of reference of the CT space with respect to the frame of reference used intraoperatively and G represents the overall transformation mapping the data set from CT to physical patient space. The solution for the six parameters t_x, t_y, t_z, r_x, r_y and r_z is generally termed as the Orthogonal Procrustes problem for which closed form solutions can be obtained using Singular Value Decomposition (SVD). For N points paired between the two data sets, the rigid body transformation G is found by minimising the objective function:

$$d(G) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\| {}^{OR}X - G({}^{CT}X)\right\|^2} \tag{2}$$

2 CRANIOFACIAL LANDMARK EXTRACTION METHODOLOGY

This section describes the method employed for automated localisation of the desired features in the white light images taken of the patient during the intra-operative registration process, using Gabor filters and a PNN neural network.

2.1 Feature Extraction Using Gabor Filters

The Gabor filter was first proposed by Dennis Gabor in 1946 [23]. The 2D-Gabor filter can be represented in normalised form in the (x,y) image plane as [24]:

$$\psi(x, y, f, \theta) = \frac{f^2}{\pi\gamma\eta} * e^{-\left(\frac{f^2}{\gamma^2}x_r^2 + \frac{f^2}{\eta^2}y_r^2\right)} * e^{j2\pi f x_r} \tag{3}$$

where $x_r = x\cos(\theta) + y\sin(\theta)$

and $y_r = -x\sin(\theta) + y\cos(\theta)$

The Gabor filter is a complex sinusoidal plane wave modulated by a Gaussian envelope, the frequency of which can be varied by the parameter f. Values of γ and η set the standard deviations of the Gaussian envelope along the two spatial dimensions. Angle θ controls the orientation of the filter. The 2-D Gabor representation in the frequency domain can be expressed as:

$$\Psi(u, v, f, \theta) = e^{\frac{-\pi^2}{f^2}(\gamma^2 * (u_r - f)^2 + \eta^2 * v_r^2)} \tag{4}$$

$$where\ u_r = u\cos\theta + v\sin\theta$$
$$and\ v_r = -u\sin\theta + v\cos\theta$$

For a given mask size, it is desirable to choose the filter parameters so that the Gaussian envelope adequately covers the spatial frequency spectrum. Up to two standard deviations in each dimension can be used to ensure adequate coverage. Any amplitude clipping of the Gaussian envelope gives rise to 'ringing' in the frequency domain and must be avoided. A final consideration in setting the filter parameters is to use a spatial frequency less than or equal to the Nyquist frequency of 0.5 cycles per pixel. At the scales of the images used, a 31x31 grid size for representing the different landmarks was found to consistently contain adequate information in the neighbourhood of the central location. A 15x15 element kernel was used to define the Gabor filters, with the Gaussian envelope parameters chosen so that it was truncated to approximately two standard deviations at the periphery of the15x15 kernel.

Setting the Gabor filter parameters as described above on a 15x15 kernel produced a maximum frequency of 0.36 cycles per pixel. As this was close to the Nyquist frequency limit, the parameters were modified slightly to obtain a maximum frequency of 0.25 cycles per pixel instead, with $\gamma=\eta=1$. Filter orientations of 0, 45, 90 and 135 degrees were used, giving a total of eight filters in the bank. With these modified parameters, the coverage of the Gaussian at the lower frequency was not optimal. The result of the Gaussian envelope not reaching two standard deviations from the mean is known to cause 'ringing' in the frequency domain. However, since the features obtained from the filters are first analysed by PCA (section 3.2) to reduce their dimensionality, the adverse effect caused by 'ringing' is likely to be suppressed.

Illumination invariance can be achieved by dividing the filter responses by the root mean square value of the response magnitudes over all the scales and orientations used. If $G_{k,m}$ represents the response at scale k and orientation m at a given location (x, y), then the normalisation step for illumination correction can be expressed as follows [25]:

$$G'_{k,m} = \frac{G_{k,m}}{\sqrt{\sum_{k,m}\left|G_{k,m}\right|^2}} \qquad (5)$$

Gabor kernels were generated for the optimised parameters and convolved with the 31x31 eye corner windows from which the central 15x15 region was selected to form the Gabor feature vector. Eight such Gabor response matrices were obtained over the two scales and four orientations used. The Gabor response, being complex, offers the possibility of using both the phase and magnitude to form the feature vector. However, only the magnitude was used since the feature vector based on both the phase and magnitude led to high dimensionalities, making it difficult subsequently to train the neural networks.

2.2 Collecting the Data Sets for Eye Corner and Ear Tragus Extraction

The eye corners and ear tragus are shown in Figure 3. A 31x31 window was used to generate the feature set for each of these landmarks. At the scale of the images used, a 31x31 window was deemed to contain sufficient information in its neighbourhood to represent both the eye corner and the ear tragus. A given sample window was selected around a landmark such that the latter is located at the centre of the window. These samples were used to generate the feature sets for the true positive training set of the neural network for which expected outputs of +1 were set. Additionally, samples where the craniofacial landmark was not in the central position of the window were collected. Outputs of -1 were set for these samples.

Figure 3: (a) Outer and Inner Eye Corners (b) General outside ear anatomy showing the tragus and the anti-tragus

2.3 Dimensionality Reduction of Data Set and Learning the Classifier

A Polynomial Neural Network (PNN) was used as the classifier for the feature detection and localisation stage. The neural network had one neuron in the output layer and the number of neurons in the input layer was related to the dimensionality of the feature vector. The output, Q of the network can be expressed as:

$$Q = T(\mathbf{W} * (\mathbf{z}, \mathbf{z}.\mathbf{z}^T) + b) \quad (6)$$

where T is the activation function, W is the input weight matrix, b is the bias, z is the input vector and from which the product combinations $z.z^T$ is derived. The dimensional expansion in finding product combinations of the input vector makes it computationally intensive to use all N=1800 elements (15x15x8) of the Gabor feature vector as inputs, since it would increase the number of dimensions to $N^2 + N$. A large dimensionality for an input feature set generally means that a neural network must be trained with a larger data set, in order to cover sufficiently the variation of the input vector. This is a necessary condition for neural network solutions to work robustly [9, 26].

Instead, the feature set obtained for the positive samples was first mapped to a lower dimensional basis using Principle Component Analysis (PCA). The minimum number of dimensions to retain was determined by the method used by Sung and Poggio [28] based on an eigenvalue spectrum. The actual number of dimensions retained (greater than this minimum value) was set based on the ease of training the PNNs as elaborated later in the results section. Additionally, a further input was computed based on the mapped (**z**) and original (**X**) datasets by finding the following distance measure [13]:

$$D = \sum (\mathbf{X} - \overline{\mathbf{X}})^2 - \sum \mathbf{z}^2 \tag{7}$$

This distance measure was added to the input feature vector. The resultant modified feature set was then used to train the PNN. This procedure was followed for extracting all the different craniofacial landmarks.

2.4 Network Tolerance to Illumination Changes and Noise

Given that the image-guided registration process is to take place in an uncontrolled environment (OR), it was considered necessary to assess the robustness of the proposed methodology to variations in illumination and random noise. For testing the performance of the network to illumination changes, gamma manipulation was used to alter the intensity and contrast of a given eye window. Gamma manipulation operates on a normalised input image **r** to produce a normalised output image **c**, using the relation: $c = r^{\gamma}$ where γ is known as the gamma factor. When γ is greater than 1, the input image is darkened while a γ value of less than 1 brightens the input image. The trained PNN was then tested on the images obtained.

Additionally, the image acquisition process, being a basically discrete process, involves noise. The combination of common sources of noise in image acquisition devices can be modelled according to the Central Limit Theorem by a Gaussian probability distribution function. Figure 4 shows examples of images obtained for the eye windows when zero mean Gaussian noise was added at different variances. The variance given for each of the images was normalised in the range [0 1] and the Gaussian noise was added to a normalised grey-scale intensity image. These images were then employed for localising the outer eye corner.

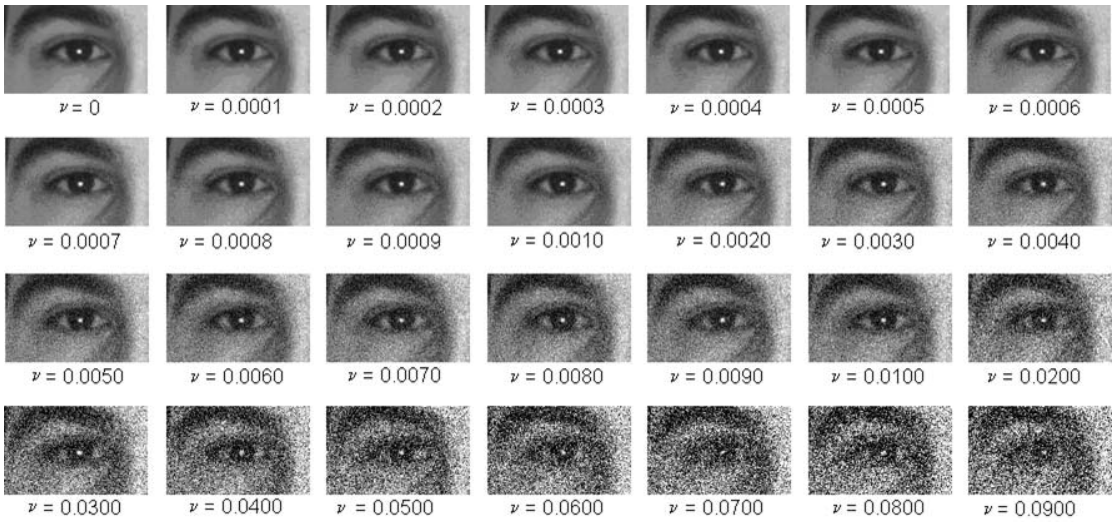

Figure 4: Eye windows corrupted with zero mean Gaussian noise of different variances [27]

3 EXPERIMENTAL WORK AND SIMULATIONS

3.1 Simulation using CT data

As this project has not yet reached the stage of clinical trials, a common set of CT and white light images of real subjects is not available. Instead, a preliminary investigation of the registration process was carried out using fifteen sets of CT scans and projected views generated from surface-rendered models of these scans. Details of the simulation study can be found in [15]. The projected views were calibrated using the simple Direct Linear Transformation (DLT) [20, 21] method following which the craniofacial landmarks were paired and reconstructed from the respective calibrated stereo views. The maximum RMS photogrammetric error obtained was 1.04 mm for the left ear tragus. These reconstructed landmarks were then used to perform rigid-body registration between the CT and simulated white light spaces. From the resultant transformation, typical entry and target points (see Figure 5) were mapped between the two spaces. A maximum RMS registration error of 0.64 mm was

obtained for mapping the entry points for the 15 models and the corresponding maximum registration error for the mapping of target points was 0.58 mm.

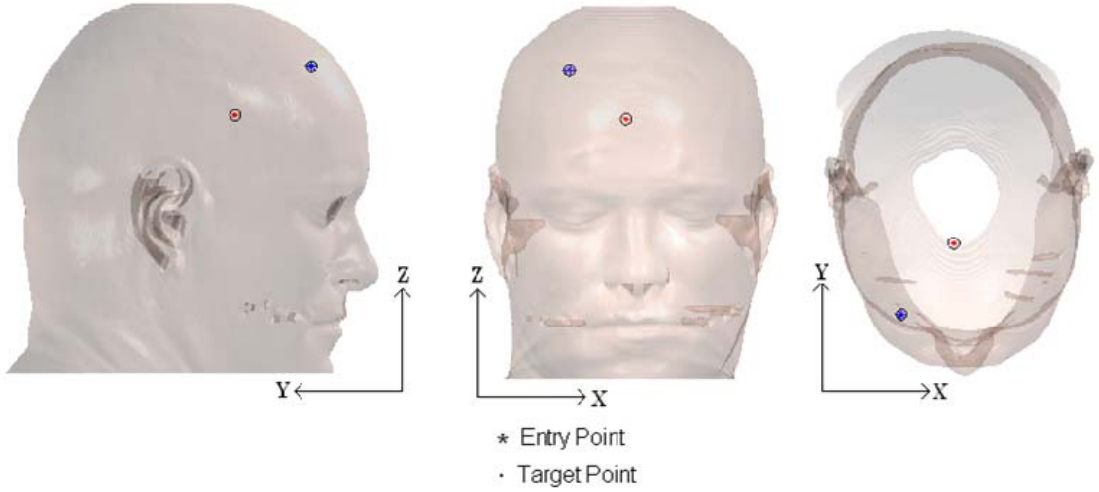

Figure 5: Simulated surgical entry and target points for one of the 15 CT models investigated 1[2]

Moreover, the locus of error estimates plotted on a CT surface-rendered head model, described and presented in detail in [15], computed according to the expression proposed by Fitzpatrick *et al*. [16] showed that within the volume of head, the targeted accuracy of 5 mm can be achieved for a landmark localisation error of 3 mm.

3.2 Photogrammetry Results with Artificial Skull

Experimental work was carried out on an artificial skull to provide error estimates for comparison with the results obtained by simulation as described in [19]. Points were marked over the skull and measured on a CMM machine with respect to a given frame of reference. These reference values were used to compute errors in reconstruction by the DLT method. RMS errors at locations close to the eye corners and ear tragi were 1.00 mm and 0.52 mm for the right side of the skull and 1.31 mm and 0.82 mm for the left side of the skull respectively. These reconstructed points were used as the registration basis to align the skull and the space characterised by the reconstructed skull points. Using the resultant transformation, the mapping of a point on top of the skull led to the following registration error along the three principal dimensions: [-0.61mm, -1.40mm, 0.74mm] giving an RMS error of 0.98 mm. This magnitude of error is considered acceptable since manual location of an entry point by a skilled neurosurgeon, after visual assessment of CT images, is not expected to be better than ±5mm.

[2] http://www.pcir.org

3.3 Detection Rates of PNN for Craniofacial Landmarks Extraction

3.3.1 Eye Corner Extraction

The results presented for the detection rates of craniofacial landmarks extraction are an extension of those provided in [22]. Face images from the AR database [27] were used for testing the algorithm to extract eye corners in the frontal view. The AR database offers four images for a given subject with variations in lighting (normal lighting, bright illumination from the left and right side of the face as well as over the whole face). One hundred sample regions of size 31x31, taken from four images of twenty-five subjects were selected for the normal and high illumination scenarios.

The first test performed was for the outer eye corner for normal and high illumination images. The criterion for successful localisation was set within a 3 pixel margin of the actual eye corner position, based on inspection by an experienced human operator. Having determined the minimum number of dimensions to retain from the eigenvalue spectrum based on similar work by Sung and Poggio [28], the actual number of dimensions retained was based on observation with the training of PNNs. Specifically, it was observed that neural networks trained from product combinations calculated from feature vectors of 20 dimensions were fast to train (<5000 epochs) and yielded very low mean squared errors (MSE) (10^{-7} to 10^{-9}). Those trained on feature vectors with 25 dimensions were also relatively easy to train (<10000 epochs) and gave low mean square errors (10^{-7}), but neural networks trained on feature vectors with more than 30 dimensions were more difficult to train (requiring more than 10000 epochs to converge and yielded MSE of the order of 10^{-2}). In subsequent selections of the number of dimensions, 20 and 25 have therefore been employed wherever appropriate. For the extraction of the outer eye corner, 7 was the minimum number of dimension required, 20 dimensions were thus retained.

The dimensionally reduced dataset obtained from the 100 sample images was used to train the PNN, which converged in less than 5000 epochs giving an MSE of the order of 10^{-9}. For subsequent testing, the trained neural network was applied to 300 sample images of subjects not included in the training set, giving correct detection in 94% of cases. The method of illumination correction used in this case was subtraction of a best-fit intensity plane [13, 14]. It was evident during testing that the bright images

gave inferior performance compared to normally lit ones, despite being part of the training set. Thus the next tests performed for the inner eye corner were based only on face samples taken under normal lighting conditions. These amounted to two per subject in the AR database [27] giving fifty samples for training for the same 25 subjects used previously.

The minimum number of dimensions needed was 9, based on which 20 dimensions were retained. The number of epochs needed for convergence during training and the resultant MSE obtained were similar to that given above. For the tests with the inner eye corner, illumination invariance was achieved by applying Equation (5). When tested over 150 subjects not used as part of the training set, a detection rate of 99% was recorded.

Since the AR database contains only frontal images, the CAS-PEAL [29] face database was used next to test the network performance for the other views. Tests could be performed on the intermediate 45 degrees views only as the profile views of the original database are not included in the current reduced version (CAS-PEAL-R1). 150 images containing the outer eye corner were used to generate the feature set. 15 was the minimum number of dimensions required as per the eigenvalue spectrum method; 20 dimensions were thus retained. The MSE obtained in less than 5000 epochs was of the order of 10^{-8}. The trained network was tested on 208 face images not used in the training set, for which a correct detection rate of 92% was obtained. The detection rate improved to 94% when the false negatives were added to the training set and the trained PNN was tested on 218 new images.

3.3.2 Ear Tragus Extraction

The CAS-PEAL database was used for ear tragus extraction since it consists of images taken at 45 degrees azimuth angle in which the subjects' ears are visible. Normalisation was achieved using Equation (5). 150 images were used to collect ear tragus sample images. The minimum number of dimensions needed was 17, based on which 20 dimensions were retained. The training of the PNN converged in less than 5000 epochs and gave an MSE of the order of 10^{-9}. The trained network was tested on 156 face images not used in the training set, for which a correct detection rate of 99% was obtained.

3.4 PNN Tolerance to Illumination Variations and Noise

The range of intensity obtained by gamma manipulation of the sample eye image was from 58 to 205, while the standard deviation (contrast) ranged from around 21 to 44. The trained PNN was tested over these different images. Using a minimum network response of 0.7 as the threshold for accepting a location as a landmark, an intensity range of approximately 60 to 160 was obtained for successful landmark localisation. Since the network thus appears to tolerate changes in illumination over a fairly wide range, in a well-illuminated environment it should operate properly.

Due to the random nature of the additive noise, the behaviour of the trained network cannot be gauged on a single degraded image. While it is expected to degrade the network performance in some proportion to the amount of Gaussian noise, the behaviour of the network did not show a monotonic degradation in performance. A better understanding of the effect of Gaussian noise can be obtained by testing over a larger number of images corrupted by noise of a given variance. Ten images were used at each noise variance level. These tests were carried out on progressively increasing variances until the network was found to give zero values consistently over most of the ten test images. A common practice to reduce the detrimental effect of Gaussian noise in an image is to perform low pass filtering. So the network performance was also assessed on the images pre-processed with a 3x3 equally weighted averaging filter.

Assuming a minimum network response of 0.7 is needed for landmark localisation, for the case without averaging, the last variance value at which all the images gave correct localisation was 0.0007. On the other hand, with averaging, the last variance at which the responses for all the ten images are correct is 0.0012. Since averaging is found to give a definite improvement in the network performance, it can be effectively applied as a pre-processing step. Furthermore, based on a visual inspection of the images shown in Figure 4, the levels of Gaussian noise at which the performance of the network experiences significant degradation can be avoided by using a camera of reasonable quality, so that the performance of the network in the presence of modest levels of Gaussian noise can be said to be satisfactory.

4 DISCUSSION AND CONCLUSIONS

This paper has described the design and operation of a patient registration framework for use with a mechatronic system to de-skill certain emergency neurosurgical procedures for the management of head injuries. The registration framework is critical to the implementation of image-guided solutions that allow specialised procedures to be carried out by non-specialist medical personnel. The design of the registration framework seeks to avoid costly equipment and computationally intensive algorithms which would render it ineffective for a practical application. Furthermore, the technique of surgical implantation of markers, which has been used as the gold standard for such registration paradigms, has deliberately not been adopted. The use of anatomical landmarks is instead preferred as the organisation of the surgery is thus considerably less constrained and should lead to a quicker time to treatment. The application of such simple design paradigms is possible due to the relatively low accuracy requirements (±5 mm).

The adequacy of a registration framework with these elements (anatomical landmarks, rigid-body registration and use of simple DLT method with no optical error correction) has been tested by simulation studies using surface rendered CT models and experimental work on an artificial skull. Based on the 5 mm accuracy requirement, the error estimates obtained from the simulation study and the experimental work on the skull were comparable. An important factor to be gleaned from these results is the importance of user validation in achieving the desired accuracy with the proposed registration method. With the uncertainty involved in Machine Vision algorithms and the criticality of the registration link, it is imperative to have user validation in the loop.

In the simulation study and the experimental work, all the landmarks were selected manually, so that even if automated methods are employed to realise the landmark extraction, ultimate user validation or verification would ensure the effectiveness of the algorithm. Being similar in principle to the actual validation of the extraction and correspondence of the landmark extraction in the two modalities, the results obtained from the simulation study as well as the experimental work with the artificial skull lend support for a user-validated registration approach.

Automated solutions for extracting the required landmarks in the CT and white light modality are the subject of current research, with a view to furthering synergy between the user and the system. The goal in this respect is to propose a registration solution to the user for validation, rather than to give instructions to the user to perform a series of actions. The preliminary work in fine landmark extraction in the white light modality has been presented as a technique in this automated registration paradigm. A Polynomial Neural Network (PNN) classifier with Gabor filter features as inputs has been proposed for the extraction of the ear tragi and eye corners. The tests performed in the presence of illumination variations showed the importance of having images occupying a dynamic range with no significant number of pixels having intensity values close to saturation or black level clipping. Allowing for a modest variation in illumination conditions provides for the option of using simple image enhancement techniques to correct for variations in contrast and brightness to optimise the network performance.

Based on the results obtained with the proposed PNN classifier for the ear tragus and eye corner extraction in different views, it has been demonstrated to be a useful feature detector. The detection rates achieved parallels the results obtained for face detection using Gabor features over one scale and four orientations and PNN as classifier by Huang et al. [13] where a detection rate of 100% was obtained over 270 images containing a single face with a simple background (collected from various sources). For more complex images consisting of clutter and multiple faces (obtained from the CMU face database), Huang et al. obtained a detection rate of 86%. Earlier work by Huang et al. [30] using intensity and gradient values as feature vectors and the one based on Gabor features [13] showed an improvement in using the latter. It was the basis for performing the craniofacial landmark extraction using Gabor features. Furthermore, the preference given to PNN over a MultiLayer Perceptron (MLP) by Huang et al. was the reason for choosing a PNN as classifier.

The results presented in this paper clearly show that extraction of the outer eye corner is harder than the inner eye corner, most probably due to occlusion of the former by the eyelashes. Further tests can be performed with larger sizes of the feature window e.g. 20x20, 25x25 in combination with different sized Gabor filter masks to assess the effect on the performance of the classifier. The results obtained for the eye corner

extraction compare favourably with other feature detectors, while the results obtained for the ear tragus extraction represent, to our knowledge, the first of their type. The sequel to the work presented would be the further development of the landmark extraction functionality so that the gross feature localisation is done automatically as well. Moreover, the registration protocol would be tested on real data in the two modalities (CT and White Light) gathered from Phantom studies and by clinical trials.

REFERENCES

1. **Fitzpatrick, J., Hill, D.** and **Maurer, C.** Registration. Medical Image Processing. In Volume II of the Handbook of Medical Imaging. **Sonka M.** and **Fitzpatrick J.M.**, ed, 2000.

2. **Henderson, J.M.** and **Bucholz, R.D.** Accurate and ergonomic method of registration for image-guided neurosurgery. *Proceedings of SPIE* 2132, 1994, p 67.

3. **Besl, P. J.** and **McKay, H.D.** A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992, **14**, pp. 239-256 1992.

4. **Clarkson, M.J., Rueckert, D., Hill, D.L.** and **Hawkes, D.J.** Registration of multiple video images to preoperative CT for image-guided surgery. *Proceedings of SPIE*, 1999, **3661,** pp. 14-23.

5. **Clarkson, M.J., Rueckert, D., Hill, D.L.G.** and **Hawkes, D.J.** Using Photo-Consistency to Register 2D Optical Images of the Human Face to a 3D Surface Model. *IEEE transactions on pattern analysis and machine intelligence*, 2001, **23**, pp. 1266-1280.

6. **Pun, K.H.** and **Moon, Y.S.** Recent advances in ear biometrics. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 164-169.

7. **Moreno, B., Sanchez, A.** and **Velez, J.F.** On the use of outer ear images for personal identification insecurity applications. *Proceedings of the IEEE 33rd Annual International Carnahan Conference on Security Technology*,1999, pp. 469-476.

8. **Yan, P.** and **Bowyer, K.W.** An automatic 3D ear recognition system. *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission* IEEE Computer Society Washington, DC, USA, 2006, pp. 326-333.

9. **Zhao, W., Chellappa, R., Phillips, P.J.** and **Rosenfeld, A.** Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 2003, **35**, pp. 399-458.

10. **Gu, L., Li, S.Z.** and **Zhang, H.J.** Learning probabilistic distribution model for multi-view face detection. *IEEE Computer Society Conference On Computer Vision and Pattern Recognition,* 2001, **2**, p 116.

11. **Schneiderman, H.** and **Kanade, T.** Probabilistic modeling of local appearance and spatial relationships for object recognition. *IEEE Computer Society Conference On Computer Vision and Pattern Recognition,* 1998, pp. 45-51.

12. **Cootes, T.F., Taylor, C.J., Cooper, D.H.** and **Graham, J.** Active shape models-their training and application. *Comput. Vision Image Understanding,* 1995, **61**, pp. 38-59.

13. **Huang, L., Shimizu, A.** and **Kobatake, H.** Robust face detection using Gabor filter features. *Pattern Recog. Lett.,* 2005, **26**, pp. 1641-1649.

14. **Huang, L.L., Shimizu, A., Hagihara, Y.** and **Kobatake, H.** Face detection from cluttered images using a polynomial neural network. *Proceedings of the International Conference on Image Processing,* 2001, **2** pp. 669-672.

15. **Gooroochurn, M., Ovinis, M., Kerr, D., Bouazza-Marouf, K.** and **Vloeberghs**, M. A Registration Framework for Preoperative CT to Intraoperative White Light Images. *Proceedings of the Thirteenth Annual Conference on Medical Image Understanding and Analysis* Kingston University London, 14-15 July 2009, **57**, pp. 184-188.

16. **Fitzpatrick, J.M., West, J.B.** and **R., M.C. Jr.** Predicting error in rigid-body point-based registration. *IEEE transactions on medical imaging.,* 1998, **17**, pp. 694-702.

17. **Ansari, A.** and **Abdel-Mottaleb, M.** Automatic facial feature extraction and 3D face modeling using two orthogonal views with application to 3D face recognition. *Pattern Recognit.,* 2005, **38**, pp. 2549-2563.

18. **Chen, L., Armstrong, C.** and **Raftopoulos, D.** An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique. *J. Biomech.,* 1994, **27**, pp. 493-500.

19. **Gooroochurn, M., Ovinis, M., Kerr, D., Bouazza-Marouf, K.** and **Vloeberghs, M.** Preoperative to Intraoperative Space Registration for Management of Head Injuries. *International Conference on Signal and Image Processing* Amsterdam, 2009, **57**, pp. 11-18.

20. **Marzan, G.T.** and **Karara, H.M.** A computer program for direct linear transformation solution of the collinearity condition, and some applications of it.

Proceedings of the Symposium on Close-Range Photogrammetric Systems, 1975, pp. 420-476.

21. **Hatze, H.** High-precision three-dimensional photogrammetric calibration and object space reconstruction using a modified DLT-approach. *J. Biomech.*, 1988, **21**, pp. 533-538.

22. **Gooroochurn, M., Kerr, D., Bouazza-Marouf, K.** and **Vloeberghs**, **M.** Extraction of Craniofacial Landmarks for Preoperative to Intraoperative Registration. *International Conference on Signal and Image Processing*, 2009, **57**, pp. 1-10.

23. **Gabor, D.** Theory of Communication. *J. Inst. Electr. Engrs.*, 1946, **93**, pp. 429-457.

24. **Kyrki, V.** Local and global feature extraction for invariant object recognition. *Ph.D. thesis. Lappeenranta University of Technology.*, 2002.

25. **Kyrki, V., Kamarainen, J.** and **Kälviäinen, H.** Simple Gabor feature space for invariant object recognition. *Pattern Recog. Lett.*, 2004, **25**, pp. 311-318.

26. **Mohamed, A.S.S., Ying Weng, Ipson, S.S.** and **Jianmin Jiang** Face detection based on skin color in image by neural networks. *International Conference on Intelligent and Advanced Systems*, 2007, pp. 779-783.

27. **Martinez, A. M.** and **Benavente, R.** The AR face database, *CVC Technical report*, 1998.

28. **Sung, K. K.** and **Poggio, T.** Example-based learning for view-based human face detection", *IEEE Trans. Pattern Anal. Mach. Intell.* 1998, **20**, pp. 39-51.

29. **Gao, W. et al.** The CAS-PEAL large-scale Chinese face database and evaluation protocols, Technique Report, No.JDL-TR_04_FR_001, Beijing: Joint Research & Development Laboratory, the Chinese Academy of Sciences, 2004.

30. **Huang, L. L. et al.** "Gradient feature extraction for classification-based face detection", *Pattern Recognit* 2003, **36**, pp. 2501-2511.