World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:3, No:9, 2009

Extraction of Craniofacial Landmarks for Preoperative to Intraoperative Registration

M. Gooroochurn, D. Kerr, K. Bouazza-Marouf, and M. Vloeberghs

Open Science Index, Biomedical and Biological Engineering Vol:3, No:9, 2009 waset.org/Publication/13963

***Abstract*—**This paper presents the automated methods employed for extracting craniofacial landmarks in white light images as part of a registration framework designed to support three neurosurgical procedures. The intraoperative space is characterised by white light stereo imaging while the preoperative plan is performed on CT scans. The registration aims at aligning these two modalities to provide a calibrated environment to enable image-guided solutions. The neurosurgical procedures can then be carried out by mapping the entry and target points from CT space onto the patient's space. The registration basis adopted consists of natural landmarks (eye corner and ear tragus). A 5mm accuracy is deemed sufficient for these three procedures and the validity of the selected registration basis in achieving this accuracy has been assessed by simulation studies. The registration protocol is briefly described, followed by a presentation of the automated techniques developed for the extraction of the craniofacial features and results obtained from tests on the AR and FERET databases. Since the three targeted neurosurgical procedures are routinely used for head injury management, the effect of bruised/swollen faces on the automated algorithms is assessed. A user-interactive method is proposed to deal with such unpredictable circumstances.

***Keywords*—**Face Processing, Craniofacial Feature Extraction, Preoperative to Intraoperative Registration, Registration Basis.

I. Introduction

REGISTRATION is a general term used to describe the alignment of two datasets. A registration basis is chosen in both datasets based on which an expression is formulated and optimised to obtain the transformation to bring about the alignment. An image-to-patient registration basis can be broadly classified as either prospective or retrospective [1]. The majority of frameless systems adopts a point-based registration basis [2]. The use of surgically-implanted fiducials and other less invasive techniques such as skin markers are not always practical because of their prospective nature. A retrospective basis goes a long way to easing the organisation of the diagnosis and surgery phases as the patient does not have to carry any artificial fiducials while awaiting operation. Anatomical features can be used as a retrospective registration basis [3]. Commonly used anatomical features of the head include the tragus, medial canthus, lateral canthus and nasion [3,4]. Intraoperatively, the required features may be found by means of relatively inexpensive stereo white light imaging as long as the accuracy requirements are satisfied. Related literature for the registration of CT and white light modalities are: Colchester et al. [5,6,7] and Grimson et al. [8]

The main contribution of this paper is the automated extraction of craniofacial landmarks in the white light modality. Feature extraction is an important task in Face Processing and the vast amount of literature available in this field provides a whole range of tools that can be applied to achieve the desired automated craniofacial feature extraction. The details of the registration protocol along with results obtained from two simulation studies for assessing the validity of the selected craniofacial landmarks (eye corner and ear tragus) are presented in [9]. The registration framework proposed in [9] includes the placement of the system with respect to the patient, which allows us to make certain assumptions which greatly simplifies the Face Localisation step and the setting of feature fields.

As for the craniofacial feature extraction, methods described in the literature can be broadly classified as grey-level image processing and statistical processing. The former relates to the use of image features such as edges and corners and known geometry about the features to be extracted to set decision rules for locating the position of the features. The statistical approach attempts to locate the features by constructing a model of the feature from known samples, which are then applied on unknown images. Template matching and neural network solutions fall under this category. Grey-level techniques have been found to be highly dependent on illumination, which is exacerbated by the need to set thresholds automatically for unsupervised processing.

Template matching [10] and neural network solutions [11,12] offer the possibility of using large training sets. Different templates may even be used for different instances of the same feature, e.g. under rotation conditions. The location of the feature is then found by cross-correlation of the template(s) with the unknown image and a high response above a given threshold is chosen as the feature location. Construction of the templates/filters has been based on normalised image intensities themselves as well as on the use of Gabor filter response over a given number of scales and orientations [11].

The first three authors are from the Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK.

M. Gooroochurn is a research student (e-mail: M.Gooroochurn@lboro.ac.uk).

D. Kerr and K. Bouazza-Marouf are senior lecturers (d.kerr@lboro.ac.uk and k.bouazza-marouf@lboro.ac.uk respectively).

M. Vloeberghs is a consultant neurosurgeon at the Queen's Medical Centre, Nottingham University, NG7 2UH, Nottingham, UK (e-mail: Michael.Vloeberghs@Nottingham.ac.uk).

For the automated extraction of the craniofacial features, a similar approach has been adopted with neural network solutions, in which the network is trained to output a value of 1 when a feature set corresponding to the desired feature is fed to the network, otherwise a value of -1 is the output. Feature vectors can similarly be constructed based on image intensities and Gabor responses. Algorithms based on Gabor masks have been shown to perform better than their intensity/gradient counterparts [13].

Based on the robustness of statistical methods, the different craniofacial feature extraction functionalities have been developed using Neural Networks, with separate nets trained for eye corner extraction in the different views. The same applies to the extraction of the ear tragus. Grey-level methods have been reported widely in the literature for the extraction of eye features, but ear feature extraction has been much less popular, possibly due to their occlusion by hair and absence in frontal view images. The variations in their shape, size and reflectance properties from person to person cast doubts over the robustness of using grey-level operations for ear detection and ear feature extraction. Based on the proven saliency of Gabor features, a neural network approach using Gabor filter response has been adopted in this work. Comparison between the detection rate in using intensity and Gabor features as input is presented in Section III.

The type of neural network used is a Polynomial Neural Network (PNN). The discriminatory power of PNN for classification problems has been shown by Huang et al. [13] for Face Localisation, where it was described to have a better performance than a MultiLayer Perceptron (MLP). PNN combines the input vector by finding product combinations between the input vector elements and thus expands the input feature set. The concomitant expansion in dimensionality is compensated by Principal Component Analysis (PCA) whereby the dataset is mapped to a lower dimensional space.

Section III describes the setting of the landmarks fields based on statistical measures of the craniofacial landmarks, generation of Gabor feature vectors for a given neighbourhood, their reduction into a lower dimensionality feature set using PCA, which is then used for training a PNN. The trained neural networks are ultimately applied for feature detection and localisation.

A. Research Context

The three neurosurgical procedures for which this registration technique is developed pertain to emergency medicine. These neurosurgical procedures are: Intracranial Pressure (ICP) monitoring, External Ventricular Drainage (EVD) and Chronic Subdural Haematoma (CSDH). ICP monitoring is used to measure the pressure of Cerebrospinal Fluid (CSF) inside the cranial vault, based on which the next course of treatment is decided. EVD is the procedure adopted to drain off excess CSF when the level of pressure measured is higher than a given threshold. CSDH is the occurrence of a blood clot between the Dura Mater and the brain surface which exerts undue pressure over the latter. Evacuation of the haematoma is the aim of this procedure, achieved by channelling a catheter into the haematoma capsule and either allowing the blood clot to drain off on its own or irrigating it with saline solution

To this end, this framework has been set up using machine vision tools to provide treatment as fast as possible within the accuracy limits allowed. The design methodology adopted has been to use proven Machine Vision tools of low complexity coupled with more advanced methods when improvements are needed. This design paradigm lends itself to more robust solutions, especially for practical systems. For example, the simple Direct Linear Transformation (DLT) method without error correction has been used for calibration and reconstruction purposes as it yields a linear solution; error correction can be added at a later stage to improve accuracy.

The preoperative space is characterised by a 3D CT surface rendered model of the patient's head, constructed from CT scans. On the other hand, the intraoperative pose is reconstructed from stereo camera views taken from frontal and profile positions with respect to the patient as well as from a third position intermediate between the frontal and profile positions. Pairing of the craniofacial landmarks in the stereo views allows their 3D reconstruction in the white light modality and ultimately the correspondence of the landmarks in the two modalities is used to align the two spaces. The entry and target points specified by a neurosurgeon can then be mapped onto the patient's head inside the Operation Room based on the resulting transformation.

Tests to carry out the proposed registration technique have not been possible to date due to the lengthiness of clinical trials and the associated high costs. [9] presents simulation studies to assess the validity of employing the selected craniofacial features as registration basis for achieving registration errors less than 5mm. The next section describes the placement of the camera system with respect to the patient, following which the automated extraction of the craniofacial landmarks is presented.

II. Registration Protocol: System Placement

To ensure accurate reconstruction of scene points by Photogrammetry, the object being measured should lie within the calibrated space [14]. Due to the spatial constraints laid by the position of the calibrated space onto the image planes in the different views, the extents of this calibrated space are utilised to position the camera system with respect to the patient. The extents of the calibrated space depend on the size of the calibration object employed, which in turn is chosen so that the object to be measured lies completely within that space. A possible method to locate the patient's head within the calibrated space is to use datum lines over the frontal and profile views. These datum lines are permanently marked on the frontal and profile displays and get overlaid over the patient's head in those views. The datum lines can then be used by the operator to position the camera system with respect to the patient.

The extreme position for the boundary of the calibrated space as observed in the profile view is made to match with the nose tip. This ensures that the patient is properly placed with respect to the frontal view in terms of the camera working distance. Additionally, having a central vertical line in the frontal view, which is aligned with the patient's nose centre and the middle of the two eyes, locates the patient correctly in the profile view. An additional horizontal line in the frontal view which is aligned to the eye corners sets the patient's head location with respect to the vertical axis of the camera image planes. Fig. 1 illustrates these datum lines in the frontal and profile views. Graduations on the horizontal line in the frontal view can be used to constrain the yaw movement of the head, whereas making the same horizontal line pass through the eye corners as much as possible constrain the roll of head.

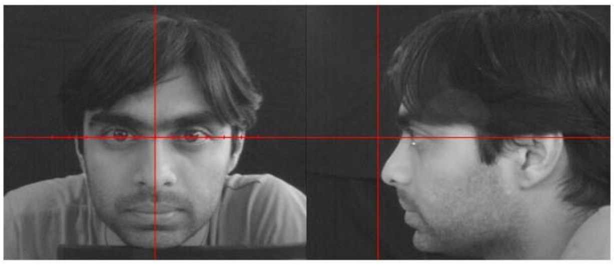

Fig. 1 Datum Lines for Initial Camera Set-up

The next section describes the Machine Vision tools inherited from Face Processing literature to aid in the automated extraction of the selected craniofacial features. Implementing any automated tool, especially in a critical application like robotic surgery, necessitates robustness and consistency of operation. Section IV introduces concepts that can be used to cope with failures of the automated extraction.

III. Craniofacial Landmark Extraction

In a general framework for Face Processing, the following sequences of tasks are normally followed: Face Localisation, Facial Feature Extraction and Face Recognition [15]. The third task is not applicable in the context of this registration framework. However, Face Localisation and Feature Extraction are steps needed to derive the feature set for the white light modality. The system placement method described above resolves most of the Face Localisation problem by using the expertise of the operator in localising a human face and placing the camera system so that the datum lines marked over the display overlays onto images of the patient in a pre-defined manner. The extraction of the eye corners in the frontal view can thus be performed along the horizontal line, starting from the intersection of the horizontal and vertical lines. Setting of the feature fields for the eye corners and the ear tragus in the different views has been done based on Face Statistics, as detailed next.

A. *Face Statistics for Setting Feature Fields*

With the placement strategy discussed in Section II, important inferences can be made about the location of the face and for setting the fields for the eye corners and ear tragus. This section presents simple statistics that can be employed to set these field windows grossly. These statistics were derived from test images of local volunteer subjects (Fig. 2).

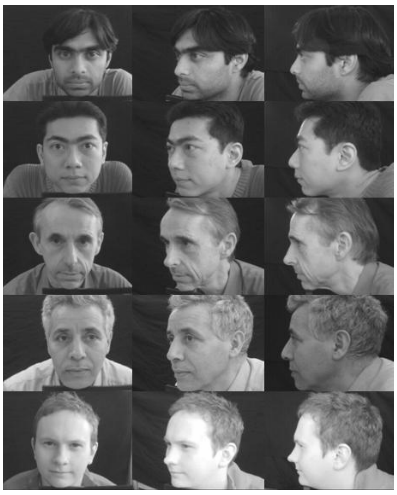

Fig. 2 Sample Images taken at University

The images were taken with cameras placed at frontal, profile and intermediate positions (see Fig. 5). Statistical parameters for specific distance ratios between features in the different views are calculated next. This is made possible by the negligible variation of the scale of the images due to the common protocol used for the camera system placement for all the subjects. The distances shown in Fig. 3 have been used to compute the ratios.

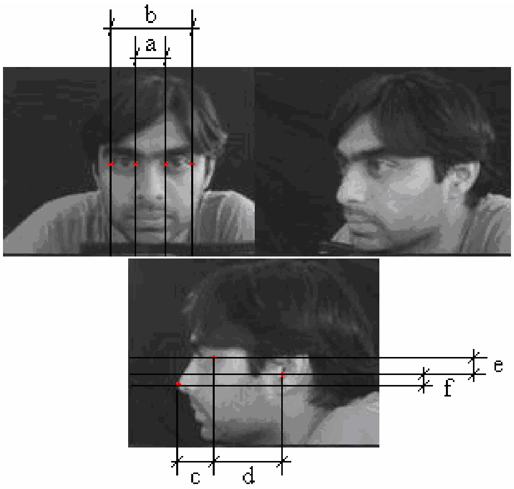

Fig. 3 Distance measures used to compute ratios

Distance a has been used as the reference in the different ratios since it can be quickly inferred from the datum lines in the frontal view. These ratios were computed from 30 image sets similar to those shown in Fig. 2.

Table I summarises the results obtained for the 30 datasets by giving the means and standard deviations for the different ratios.

TABLE I
STATISTICS OF RATIOS

RATIOS	MEAN	STANDARD DEVIATION
b/a	2.52	0.17
c/a	1.11	0.28
(c+d)/a	3.41	0.41
e/a	0.65	0.28
(e+f)/a	0.89	0.19

Since the horizontal datum lines in the frontal and profile images are made to pass more or less over the eye corners, the approximate vertical position of the eye corners are defined in all the views. The horizontal positions of the outer eye corners in the frontal view are set based on the ratio b/a. Ratio c/a can be used to set the horizontal position of the outer eye corner in the profile view. The ear tragus in the profile view is undefined in the horizontal and vertical dimensions, so ratios (c+d)/a and e/a are used in a similar way to set the area of interest (AOI) for the ear tragus with respect to the vertical nose line and the horizontal datum line. The AOI for these fields can be set by taking three standard deviations on either side of the mean (to encompass more than 99% of the expected values).

As described later, a 31x31 window size is used to scan the AOI for detecting the features. In collecting the training set, the feature to be detected was placed at the centre of the 31x31 window. Since the ratios derived above give the spatial distribution of the features, half of the 31x31 window size is added around the AOI obtained from the ratios to make sure the features are not missed. AOIs obtained by applying these ratios and compensating for the size of the scanning window for the outer eye corners in the frontal view and ear tragus in the profile view are shown in Fig. 4 for two sample images.

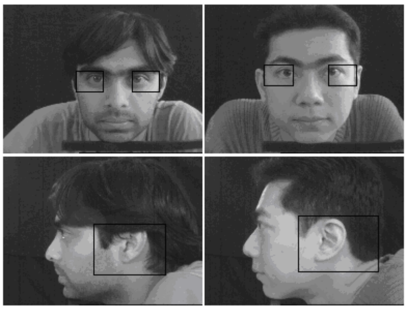

Fig. 4 Setting gross feature field

Fig. 5 shows the 3-camera configuration employed for capturing the different views of the patient's head. The three cameras are arranged over a circle's quadrant so that the working distances are equal. In this way, the regions defined in any two views are constrained to lie in a given region of the third view. With the vertical position set to be equal in all the views, this allows us to set similar AOIs for the ear tragus and the outer eye corner in the intermediate view as well.

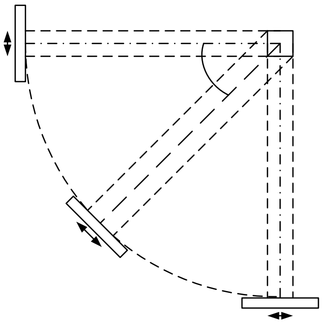

Fig. 5 Camera Geometry Constraints

The next tasks presented are the precise localisation of the outer and inner eye corners in the frontal view and ear tragus in the profile view. The precise localisation has been developed on smaller windows compared to the AOIs obtained using the ratios, especially relative to the window obtained for the ear tragus which had a large variability between the subjects. Methods to constrain the search region from the AOIs will be sought so that the techniques developed next are applied over a smaller region.

B. Eye Corner Extraction

With the AOIs for the inner and outer eye corners set based on the geometrical constraint of the initial placement of the camera system with respect to the patient, the next task is to locate the inner eye corners in this window followed by the outer eye corners. Since eye corners are salient features of the face, it is no doubt that a vast amount of literature exists related to this task. The seminal method in this respect [16] relied on grey-level processing to extract edges from the face area and using integral projections to locate the eye region. Similar methods have since been used to locate the eye centres as the latter appear as dark areas with respect to the eye sclera and thus a valley is obtained in the integral projections. Extraction of the eye corners using filter masks, once the eye field has been precisely localised, is proposed in [11] where a 5x5 filter is constructed for the inner eye corner and a 7x7 filter is used for the outer eye corner. These filters are obtained by averaging the response of Gabor filters over two scales and eight orientations. Colour-based segmentation [17] has also been applied for this task in which the eye region is first segmented from the eye field followed by finding the extremities of that region, which are assigned to the inner and outer eye corners. Use of Gabor features over grey-level and gradient values has been shown over recent years to yield more robust performance for either detection or recognition and the recent trend has seen a clear inclination for the former.

The approach adopted for the eye extraction (and ear extraction) is a combined detection and localisation methodology in which a high response from the classifier signifies both the presence of an eye and the location of the eye corner. Gabor features are used to form the feature vector for training the classifier. The 2D-Gabor filter can be represented in the normalised form as proposed by [18]:

$$\psi(x, y, f, \theta) = \frac{f^2}{\pi\gamma\eta} * e^{-(\frac{f^2}{\gamma^2}x_r^2 + \frac{f^2}{\eta^2}y_r^2)} * e^{j2\pi f x_r}$$
$$x_r = x\cos(\theta) + y\sin(\theta) \qquad (1)$$
$$y_r = -x\sin(\theta) + y\cos(\theta)$$

Where x and y are the spatial dimensions. 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. Fig. 6 shows an example of a Gabor filter's real part. γ and η are the standard deviations of the Gaussian envelope along the two spatial dimensions. Angle θ sets the orientation of the filter. The Gabor filter has been shown to provide the minimum uncertainty in time and frequency domains [19]. Furthermore, its association with the response of the visual cortical cells [20] lends more support to its use as a feature extractor as it is a constant drive of Machine Vision to mimic the capabilities of the visual system.

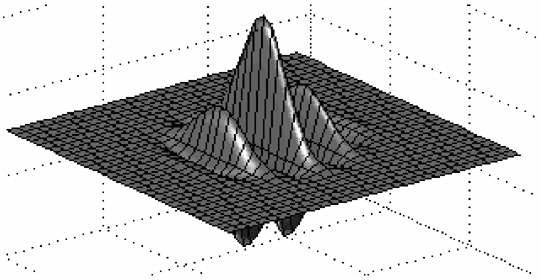

Fig. 6 Example of Real part of Gabor filter

With the flexibility of the Gabor filter to sweep a wide range of orientations, 8 orientations are normally chosen to cover a full revolution and 5 scales to vary the frequency. The scales are normally chosen to be multiples of 2 so as to keep the frequency bandwidth to 1 octave. So it is common in Machine Vision applications to use 40 masks for feature extraction. The frequencies and orientations are then given by:

$$f_k = \frac{f_{\max}}{2^{k-1}}, \quad k = 1,2,...,5$$
$$\theta_m = \frac{(m-1)*2\pi}{8}, \quad m = 1,2,...,8 \qquad (2)$$

However, the full set of 8 orientations and 5 scales have not been used in building the feature set for the craniofacial extraction; only two scales and four orientations have been used for the results presented. The maximum frequency employed was 0.25 per pixel so that wavelengths of 4 and 8 pixels were obtained for the two scales. The minimum wavelength of 4 pixels is set to satisfy the minimum wavelength of 2 pixels given by the Nyquist criterion. The four orientations chosen were 0, 45, 90 and 135 degrees. From the response of the Gabor masks over the two scales and four orientations, illumination invariance is achieved by dividing the responses by the root mean square value of the magnitudes of the responses 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 [21]:

$$G'_{k,m} = \frac{G_{k,m}}{\sqrt{\sum_{k,m}\left|G_{k,m}\right|^2}} \qquad (3)$$

Alternatively, Huang et al [13] performed illumination correction by subtracting a best-fit intensity plane from the image. The size of each of the matrix $G_{k,m}$ depends on the neighbourhood around the location chosen. For the inner and outer eye corners extraction, a 31x31 region centred at the eye corner, was selected in face images. Gabor kernels were generated for frequencies of 0.25 and 0.125 per pixel with unity values for γ and η and orientations of 0, 45, 90 and 135 degrees. These Gabor kernels were convolved with the 31x31 eye corner windows from which the central 15x15 region was selected to form the Gabor feature vector. 8 such 15x15 Gabor response matrices were obtained over the two scales and four orientations applied. These were normalised using Equation (3) and arranged into a column vector, from which the feature vector to train the PNN is derived as described later. The Gabor response, being complex, offers the possibility to use both the phase and magnitude to form the feature vector.

Face samples from the AR database [22] were employed for tests on the eye corners extraction. The face samples have been taken to give a realistic variation of facial state commonly occurring in an uncontrolled environment, such as face occlusion, emotions and illumination changes. Occlusion of facial features and emotions has not been considered in the tests carried out. However, illumination changes have been taken into account as changes in light level inside the Operating Room should be tolerated to a certain extent. The AR database offers four images of interest in this respect for a given subject. These are:

1. Normal lighting
2. Higher illumination from the left side of the face
3. Higher illumination from the right side of the face
4. Higher than normal illumination over the whole of the face

These variations in lighting were used to test the robustness of the algorithm to changes in lighting. One hundred samples of size 31x31, with the outer eye corner located at the centre, were collected from these images. A 1800-element vector was thus generated for each image sample (15x15 central pixels of 31x31 window over 2 scales and 4 orientations) from the normalised Gabor magnitude response. These samples were used to generate the feature vectors for the true positive training set for the neural network for which outputs of +1 were set. Additionally, samples where the outer eye corners were not in the central position of the 31x31 window were collected; regions like the eyebrows and cheek/hair line were also included as the windows for the eyes are set with a fair degree of tolerance and these facial features may come in the field. Outputs of -1 were set for these samples. A PNN was used as the classifier for the feature detection and localisation based on its robust discrimination by using product combinations of the input feature vector in addition to the input vector itself. A single hidden layer network architecture is employed with one neuron in the output layer.

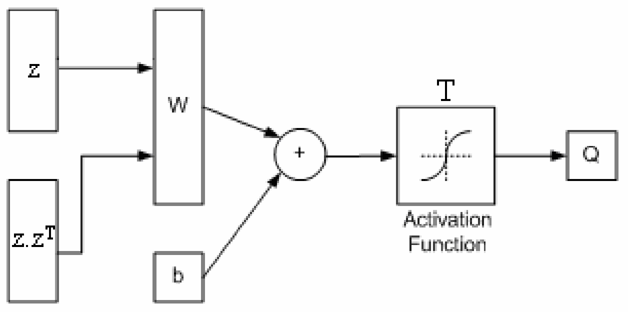

Fig. 7 PNN Architecture

The overall transformation of the network can be expressed as:

$$Q = T(W * (z, z.z^T) + b) \qquad (4)$$

where T is the activation function, W is the input weight matrix, b is the bias, z is the input vector and from it, the product combinations z.z^T is derived, which contains all the product combinations between the input vector elements. The output of the network is Q. The dimensional expansion in finding product combinations of the input vector makes it computationally intensive to directly use the 1800 elements of the Gabor feature vector as input since it would increase the number of dimensions to 1800*1800 + 1800. So the feature set obtained for the positive samples are first mapped to a lower dimensional basis using Principal Component Analysis (PCA). The output of PCA is a set of eigenvectors and eigenvalues, from which the contribution of each eigenvector as a basis for the representation of the dataset can be gauged by its corresponding eigenvalue. Although the selection of the number of dimensions to retain in the dataset is arbitrary, the minimum number of dimensions was determined from the eigenvalue spectrum by finding the eigenvalue number at which the sum of eigenvalues arranged in descending order, starting from the lowest and summing towards the largest, equals the highest eigenvalue.

With the possibility to use both the magnitude and phase of the Gabor response to form the feature vector, input vectors of size 1800 and 3600 can be constructed. However, from dimensional reduction by PCA, the combined magnitude and phase case led to a generally high lower limit for the number of dimensions, making networks based on them difficult to train as they require a much larger training set. Based on these observations, the magnitude of the Gabor response was used for further tests. The dataset is then mapped onto the reduced dimensional space and the product combinations are computed to act as the input for the PNN. Additionally, a further input is computed based on the mapped (z) and original (X) datasets by finding the following distance measure:

$$D = \sum (X - \overline{X})^2 - \sum z^2 \qquad (5)$$

This distance measure is added to the input feature vector. The resultant feature set was used to train the PNN. This procedure was followed both for the outer and inner right eye corners. 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 eye corner. A hundred positive samples taken from four images of twenty-five subjects were collected for the normal and high illumination scenarios. For subsequent testing, the trained neural network was applied on 300 sample images of subjects not included in the training set, giving a detection rate of 94%. The method of illumination correction used in this case was subtraction of an intensity plane of best-fit. Fig. 8 shows some output samples for the outer eye extraction.

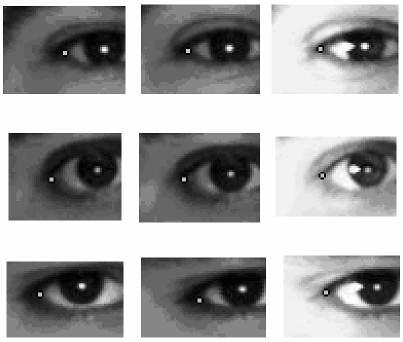

Fig. 8 Examples of Outer Eye Corners Results

It is worth mentioning that similar tests on the same training and testing datasets with a 15x15 intensity window yielded a detection rate of 86%. Comparing the performance, it was obvious that the Gabor features achieved better results. Moreover, it was evident during testing on face samples, not included in the training set, that the bright images performed poorer compared to normally lit ones, despite the presence of bright samples in 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 amounts to two per subjects in the AR database [22] giving fifty samples for training for the same 25 subjects used previously. Although lighting conditions will not be controlled during the surgery, global image statistics derived from a histogram can be effectively used to control the aperture of the cameras to prevent overly bright or dark images. The histograms of the eye windows for the brightly lit images were largely saturated, and this was the reason for trying the tests for the inner eye corner over the normally-lit images. The tolerance of the network to changes in brightness and contrast will be assessed at a later stage by applying histogram sliding and contrast stretching to the normally lit eye windows to progressively increase and decrease the contrast and brightness and assess the resulting network performance. This should enable setting of ranges over which the network can operate properly.

For the tests with the inner eye corner, illumination invariance was achieved by the method proposed in [21] where the response matrix is divided by the root mean square value of the whole responses over the different scales and orientations as given in Equation (3). When tested over 150 subjects not used as part of the training set, 149 successes were recorded (based on the same criterion as for outer eye corner). The only case where the algorithm was deemed to have failed was when the PNN response was below 0.5 although the position was correct. This can be corrected by training this particular instance as a false negative, but the aim of the experiment was to find how well the classifier generalises over data not used during training. Fig. 9 shows output samples for inner eye corner extraction.

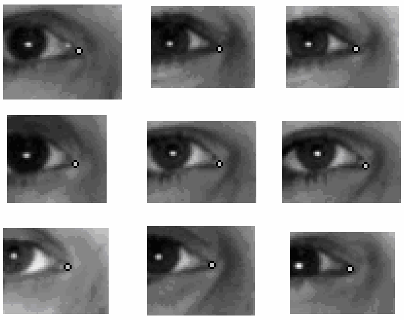

Fig. 9 Examples of Inner Eye Corner Extraction Results

C. Ear Tragus Extraction

The same methodology was adopted for ear tragus extraction, with the use of the PNN as a classifier and Gabor masks for feature extraction. Normalisation was achieved by dividing by the root mean square of the Gabor responses. However, during the training set collection, it became evident that the ear structure varies considerably more from subject to subject than the eye corner. The variation occurs in size, shape and complexion of skin. Again the same basis for lighting correction by using global image statistics was adopted, so that changes within the range of white saturation and black level clipping was considered, and any situation found to lie outside this range can be corrected during image capture by adjusting the camera aperture. With the variation in the size of the ear structure, scaling was used to make sure the chosen 31x31 window size contained the ear tragus, the anti-tragus and the valley linking these two (Fig. 10).

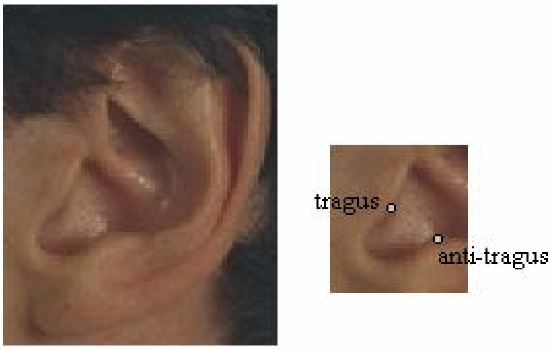

Fig. 10 General outside ear anatomy and desired ear structure to appear in 31x31 window [23]

Scales of 1, 0.9, 0.8, 0.7 and 0.6 were used during collection of the training set; for a given image, the scale at which the 31x31 window contained the desired ear structures was saved as a training sample. Samples where this structure did not appear were also collected and used as false positives in the training set. The FERET database [23] was used for the ear tragus extraction as views are available around the subject’s head at an angle of 45 degrees as well as profile images. With the complexity of the ear tragus, the training of the neural network was done in phases, as it was evident that ear shape would have a significant impact on the detection rate. The first training was done on 100 images, followed by tests on 90 samples not used during training. A detection rate of 76% was achieved under a similar criterion of 3 pixels margin from the true position. The false negatives were collected and added to the training set for the next training phase. At this point, PCA was again carried out because the basis vector and the mean had to be updated. The PNN was trained again on the new training set and the false positives collected during subsequent testing. In the second testing phase, the newly trained neural network was tested over 181 face samples not used for training, out of which 169 were successful (93% detection rate). The improvement in detection rate can be attributed to the inclusion of more shape variations of the ear in the training samples. Further inclusion of false negatives and training over more samples is expected to further improve the accuracy. Fig. 11 shows output samples for the ear extraction.

The use of several scales means that the responses have to be merged to define a single location for the ear tragus since only one ear tragus was searched for in a given image. The highest response over the different scales is chosen to be the one defining the location of the ear tragus.

Open Science Index, Biomedical and Biological Engineering Vol:3, No:9, 2009 waset.org/Publication/13963

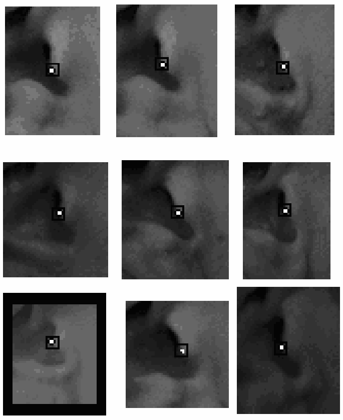

Fig. 11 Examples of Ear Tragus Extraction Results

D. Photogrammetry for Craniofacial Feature Reconstruction

Using the three camera system depicted in Fig. 5 implies that the ear tragus can be reconstructed fully from the profile and intermediate views and the inner and outer corner for one of the eyes can be reconstructed fully from the frontal and intermediate/profile views. The other eye is seen in only the frontal view and of itself, this is insufficient to reconstruct the 3D coordinates. Assuming a coordinate system with the frontal image plane giving the X-Z dimensions and the profile image plane giving the Y-Z dimensions, using a technique similar to Ansari et al. [17], the missing Y-coordinates for the eye corners on the hidden side of the face can be found in the image plane coordinate by symmetry. In this way, the 3D coordinates of the two eyes' inner and outer corners and one ear tragus can be obtained. The ear tragus on the hidden side of the head cannot be found since it does not appear in any of the views. With at least three of these five landmarks, it would be theoretically possible to perform rigid-body reconstruction. However, based on their distribution, it is inappropriate to use only the facial features; the ear tragus should be added in the registration basis to ensure wide enough coverage of the feature points.

IV. User-Interaction in Registration Framework

Since the proposed registration framework will be operated in a medical environment, as part of head injury management, it is imperative to ensure a robust operation of the developed algorithms. The validation of lighting adequacy is a way to achieve proper brightness and contrast in the camera images captured and this has been shown to yield better detection. This section introduces concepts that will be applied in the framework to ensure robustness during the running of the procedure as far as the registration is concerned.

A. Refining &Validating the Automated Response

Although the goal of using image-guided solutions for the three neurosurgical procedures is to make them as widely accessible as possible, the benefits of the system, although compelling as an outcome in the management of head injuries, do not guarantee its uptake by the medical community. The integration of the proposed concept within the medical setup to act as a powerful medical tool is an important factor in its uptake. The system is asked to give confidence to the operator by providing clear indications about its present and future course of action. Once the system is correctly put into position with respect to the patient, automated extraction of the craniofacial features is undertaken. The successful extraction of the craniofacial features leads to overlaying the results onto the CT scans in a manner that clearly shows the matching of the landmarks in the two modalities.

At this point, the operator is asked to validate whether the overlapping has been correctly performed, and with it, the registration link between CT and the patient. In the event the user sees any discrepancy in the mapping, he/she can intervene to either move the landmark(s) found by the system to position(s) thought to be more appropriate. In this way, the onus to understand and implement all the steps correctly by himself/herself is taken away from the user as the system works in synergy with the operator.

Situations where the system cannot detect the craniofacial features in one or more of the views will make it impossible to reconstruct the 3D coordinates of these landmarks. This failure can arise for several reasons, one being evidently the inability to design Machine Vision systems that operate 100% in all circumstances. Linked to this aspect of Machine Vision systems is the ability to cope with new types of datasets, e.g. a new ear shape may lead to unpredictable performance from the ear tragus detector. Another difficult scenario that can be encountered in the management of head injuries are damaged or swollen parts of the face. Although more research needs to be carried out in this respect with particular attention to anatomy and the correspondence to what is seen in CT scans and camera images, the next section gives some insight about how injuries to the face can be dealt with.

B. Non-Invasive Contact Fiducial for Landmark Localisation

The benefit of not using artificial fiducials as registration basis is unequivocal; better organisation of diagnosis and surgery is achieved by a markerless registration basis. Within the framework of the targeted emergency procedures, a markerless basis would reduce the need for additional surgery and extra scanning of the patient with fiducials implanted. The ability to rely on the automated mode while providing for an alternative way to perform the registration in the event the automated functionality fails stem again from the relatively low accuracy needed for the particular task at hand as compared to other precise neurosurgical procedures.

Although still at the conceptual stage, a possible solution proposed for managing injuries to the head/face is a handheld rod with a spherical end having good contrast from its surroundings so that it can be easily picked up by the cameras. This contrast can be achieved by having an LED in the spherical end. The operator places the sphere in contact with the eye corner or the ear tragus, which is segmented and paired in the stereo views. This should be possible in a robust and repeatable manner. The user would rely on this contact-based method either if he/she is not satisfied with the point given by the system or the system could not detect a point at all. Upon successful reconstruction of the features and registration with the CT space, the resulting mapping is again displayed to the user by overlapping the two modalities. The user is asked to validate the registration before the subsequent image-guided steps can start.

V. Discussion

The general theme of this paper has been the generation of a markerless registration basis for capturing and modelling the patient's pose intraoperatively. Although different methods of performing preoperative to intraoperative registration exist, a registration protocol has been conceived, aimed at employing relatively low cost equipment for the registration requirements. Use of sophisticated and costly techniques of registration such as opto-tracking systems and laser range scanners has been avoided in implementing the registration protocol since the main objective of the project is to design a system which can be used by local hospitals. This is a key factor to make the targeted neurosurgical procedures as widely accessible as possible.

Additionally, the retrospective nature of using natural landmarks as the registration basis decouples the organisation stages of preoperative CT scanning for diagnosis and the surgical intervention itself as the patient does not have to carry the fiducials between these two stages. Further benefits include reduced pain for the patient, no risk of infection from the placement of the fiducials and within the purview of emergency head injury management, it gives the optimal time-to-treatment.

Selection of a given registration basis puts certain constraints on the registration framework in terms of the achievable accuracy whereas the method adopted to extract the features determines the type of equipment required intraoperatively. Registration of CT and white light images using anatomical landmarks has previously been undertaken by reconstructing 3D surfaces from point clouds, but at the cost of expensive equipment and lengthy protocols. Other techniques using pixel values of the whole image in the two modalities avoid segmentation of features and by using more information, achieves more robustness. But the computational load is often high for these techniques and the modalities need to be roughly aligned for them to converge to the correct solution. The three neurosurgical procedures do not require very high accuracy, thus leading the way to simpler equipment and algorithms.

However, segmentation of features, especially by automated techniques, remains a challenge in Machine Vision applications and the uncertainty it carries with respect to intensity variations in the raw sensor data casts doubts on the reliability of setting threshold levels for segmentation automatically. This uncertainty makes semi-automated techniques for gross positioning a preferred option, where the user is allowed to intervene and set global constraints in the solution, e.g. by defining a region of interest for a feature to lie instead of allowing the algorithm to search exhaustively over the whole image. The subsequent refinement of the registration is then achieved by using more information-rich methods such as intensity-based methods.

The same methodology is implemented in the proposed framework, with the use of automated extraction of craniofacial landmarks by a neural network approach. However, the refinement of the solution has not been considered at this stage as simulation studies [9] show that the 5mm accuracy aimed at can be achieved solely by a feature-based approach. Further work in a clinical set-up will help validate this hypothesis and pave the way to improvements if deemed necessary. The validation of the landmark extraction is performed by the user on the end result, that is, on the mapping of the stereo views onto the CT head model. It cannot be ascertained that the landmark extraction algorithms will succeed in all cases,. Nonetheless, using the validation of the result at the output stage of the registration brings both confidence and robustness in using the system.

Neural network solutions for similar tasks have been shown to achieve high levels of performance, especially when trained over a large dataset. So the same can be expected for the craniofacial landmark extraction as more samples of eyes and ears are trained into the system, especially for the ear as they were found to have large variations in shape and size. In this regard, an automated paradigm as a first course of action is justified as compared to one based solely on manual extraction; a high detection rate means that most of the time the system will automatically extract the features and perform the registration correctly. Moreover, the automated extraction mode is believed to bring more synergy between the user and the system.

VI. Conclusion

The automated extraction of craniofacial landmarks in white light modality as a component of a registration framework for preoperative CT to intraoperative white light images has been described. This registration methodology has been devised to support three neurosurgical procedures that are emergency in nature. Simulation of the registration framework gave errors within the required 5mm accuracy. Clinical tests of the protocol will give definite measure of the adequacy of the framework and lead the way for any further improvements needed.

Central to the extraction of the landmarks in the white modality is an automated approach for which a neural network solution has been illustrated. An automated approach to this task is considered an important ingredient in the successful uptake of the system in the medical set-up as it brings synergy between the user and the system. The automated extraction algorithms presented employed a Polynomial Neural Network (PNN) classifier with Gabor features as input. A detection rate of 94% was obtained for the outer eye corner for tests on the AR database with frontal view samples containing images with normal and bright illumination. A 99% detection rate was obtained for the inner eye corner when the same subjects were used, excluding the brightly-illuminated cases. The influence of images with intensity distributions close to the dark and bright extremes of the dynamic range were found to be detrimental to the generalisation of the classifier. Hence a histogram-based method for adjusting the aperture of the cameras is envisaged.

Extraction of the ear tragus proved to be more complex due to the size and shape variations among subjects. Progressive training and testing yielded a detection rate of 93%. The detection rates for the landmarks in general are expected to improve further as more samples are included in the training set, thus reducing the frequency of automated incorrect registrations, and correspondingly reduced user interaction at the registration level. PNNs will be similarly developed for eye corner extraction in the profile and intermediate views and ear tragus extraction in the intermediate view.

Validation of the registration between the preoperative and intraoperative spaces by the user is an important component of the framework to bring robustness and confidence in the system and allows to cope with the unsuccessful automated registrations. Finally, methods to deal with these eventualities have been presented.

References

[1] J. M. Fitzpatrick, D. L. G. Hill and C. R. Maurer Jr, "Image registration", in *Handbook of Medical Imaging II: Medical Image Processing and Analysis* , vol. 2, M. Sonka and J. M. Fitzpatrick, Eds. Bellingham, WA: SPIE Press, 2000, pp. 447–513.

[2] J. McInerney and D. W. Roberts, "Frameless stereotaxy of the brain," *Mt. Sinai J. Med.,* vol. 67, pp. 300-310, Sep. 2000.

[3] C. R. Maurer, R. P. Gaston, D. L. G. Hill, M. J. Gleeson, M. G. Taylor, M. R. Fenlon, P. J. Edwards and D. J. Hawkes, "AcouStick: A Tracked A-Mode Ultrasonography System for Registration in Image-Guided Surgery," *LECTURE NOTES IN COMPUTER SCIENCE,* pp. 953-962, 1999.

[4] M. J. Citardi, "Computer-aided frontal sinus surgery", *Otolaryngol. Clin. North Am.,* vol. 34, pp. 111-122, 2001.

[5] A. C. Colchester, J. Zhao, K. S. Holton-Tainter, C. J. Henri, N. Maitland, P. T. Roberts, C. G. Harris and R. J. Evans, "Development and preliminary evaluation of VISLAN, a surgical planning and guidance system using intra-operative video imaging", *Med. Image Anal.,* vol. 1, pp. 73-90, Mar. 1996.

[6] M. J. Clarkson, D. Rueckert, D. L. G. Hill and D. J. Hawkes, "Using photo-consistency to register 2D optical images of the human face to a 3D surface model", *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 23, pp. 1266-1280, 2001.

[7] M. J. Clarkson, D. Rueckert, D. L. Hill and D. J. Hawkes, "Registration of multiple video images to preoperative CT for image-guided surgery", *Proceedings of SPIE,* vol. 3661, pp. 14, 2003.

[8] W. E. L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Perez, W. M. Wells III and R. Kikinis, "An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization", *Medical Imaging, IEEE Transactions on,* vol. 15, pp. 129-140, 1996.

[9] M.Gooroochurn, M.Ovinis, D.Kerr, K.Bouazza-Marouf, and M.Vloeberghs, " Preoperative to Intraoperative Space Registration for Management of Head Injuries", *Sixth International Conference on Signal and Image Processing*, submitted for publication

[10] R. Brunelli, T. Poggio, "Face recognition: Features versus templates", *IEEE transactions on pattern analysis and machine intelligence*, yr:1993 vol:15 iss:10 pg:1042

[11] Zhonglong Zheng, Jie Yanga and Limin Yang, "A robust method for eye features extraction on color image", *Pattern Recognition Letters* Volume 26, Issue 14, 15 October 2005, Pages 2252-2261

[12] Yeon-Sik Ryu, and Se-Young Oh, "Automatic extraction of eye and mouth fields from a face image using eigenfeatures and multilayer perceptrons", *Pattern Recognition*, Volume 34, Issue 12, Pages 2459-2466, December 2001

[13] Lin-Lin Huang, Akinobu Shimizu and Hidefumi Kobatake, **"**Robust face detection using Gabor filter features", *Pattern Recognition Letters* Volume 26, Issue 11, August 2005, Pages 1641-1649

[14] LIANG CHEN, C.W. ARMSTRONG, D.D. RAFTOPOULOS, "An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique", *Journal of biomechanics*, Vol. 27, No. 4, pp. 493-500, 1994

[15] W. ZHAO, R.CHELLAPA, P. J. PHILLIPS et al., "Face Recognition: A Literature Survey", ACM *Computing Surveys (CSUR)*, Volume 35 , Issue 4 (December 2003), Pages: 399 – 458

[16] Takeo Kanade, "*Picture Processing* System by *Computer Complex and Recognition of Human Faces*," doctoral dissertation, Kyoto University, November, 1973

[17] A-Nasser Ansari, Mohamed Abdel-Mottaleb, "Automatic facial feature extraction and 3D face modelling using two orthogonal views with application to 3D face recognition", *Pattern Recognition* 38 (2005) 2549 – 2563.

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

[19] D. Gabor, "Theory of communications". J. Inst. Electr., Engrs. 93, 429–457, 1946

[20] J.G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Am. A* 2 (7), 1160–1169, 1985

[21] K. Ville, J.-K. Kamarainen, H. Kalviainen, "Simple Gabor feature space for invariant object recognition", *Pattern Recognition Letters* 25 (2004) 311–318

[22] A.M. Martinez and R. Benavente "The AR Face Database*", CVC Technical Report* #24, June 1998"

[23] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", *Image and Vision Computing J*, Vol. 16, No. 5, pp 295-306, 1998.