Patient-Specific Bronchoscope Simulation with *pq*-Space Based 2D/3D Registration

Fani Deligianni, Adrian Chung, PhD, Guang-Zhong Yang*, PhD

Royal Society/Wolfson Foundation Medical Image Computing Laboratory
Imperial College London,
London SW7 2BZ, United Kingdom

Running Title: *pq*-Space Based 2D/3D Registration

Acknowledgements:

The authors would like to acknowledge the financial support from the EPSRC, the Royal Society, and the Wolfson Foundation. They are also grateful to Pallav Shah (MD), and Athol Wells (MD) for their contribution to the clinical data collection for this paper.

Part of this paper was presented at the MICCAI 2003 meeting in Montreal, November 2003.

Address for Correspondence:

Professor G.Z. Yang
Royal Society/Wolfson Foundation MIC Lab
Department of Computing
180 Queen's Gate
Imperial College
London SW7 2BZ
United Kingdom

Tel: (+44) 20 7594 8441
Fax: (+44) 20 7581 8024
Email: [**g.z.yang@imperial.ac.uk**](mailto:g.z.yang@imperial.ac.uk)

Patient Specific Bronchoscope Simulation with pq -Space Based 2D/3D Registration

ABSTRACT

Objective: The use of patient-specific models for surgical simulation requires photorealistic rendering of 3D structure and surface properties. For bronchoscope simulation, this requires augmenting virtual bronchoscope views generated from the 3D tomographic data with patient specific bronchoscope videos. In order to match video images to the geometry extracted from 3D tomographic data, this paper presents a new pq -space based 2D/3D registration method for camera pose estimation in bronchoscope tracking.

Methods: The proposed technique involves the extraction of surface normals for each pixel of the video images by using a linear local shape-from-shading algorithm derived from the unique camera/lighting constraints of the endoscopes. The resultant pq -vectors are then matched to those of the 3D model by differentiation of the z-buffer. A similarity measure based on angular deviations of the pq -vectors is used to provide a robust 2D/3D registration framework. Localization of tissue deformation is considered by assessing the temporal variation of the pq -vectors between subsequent frames.

Results: The accuracy of the proposed method is assessed by using an electro-magnetic tracker and a specially constructed airway phantom. Preliminary *in vivo* validation of the proposed method is performed on a matched patient bronchoscope video sequence and 3D CT data. Comparison to existing intensity-based techniques has also been made.

Conclusion: The proposed method does not involve explicit feature extraction and is relatively immune to illumination changes. The temporal variation of the pq distribution also permits the identification of localized deformation, which offers an effective way of excluding these areas from the registration process.

Keywords: Shape from Shading, pq -Space, 2D/3D Registration, Virtual Endoscope, Virtual Bronchoscopy, Surgical Simulation

Key Links: <http://vip.doc.ic.ac.uk>

1. Introduction

Bronchoscopy is a special form of minimal access surgical procedure for examining the trachea, bronchi, and the air passages that lead to the lungs. Due to the complexity of instrument control, restricted vision and mobility, and the lack of tactile perception, it requires a high-degree of manual dexterity and hand-eye coordination. Similar to other forms of minimal access surgery, the acquisition of surgical skills is normally obtained through practising on inanimate plastic models and subsequently on patients. With the use of plastic models, however, it is difficult to provide high fidelity physical responses that are necessary for advanced skills training and assessment. Practising on real subjects, on the other hand, prolongs the examination time, and can involve considerable discomfort to the patient and carry certain risks of complications. With the maturity of augmented reality systems, there has been an increasing demand of using computer simulation for performing certain aspects of this training, particularly for hand eye co-ordination and instrument control. For most of the current simulation systems, however, the degree of visual realism is severely limited. In bronchoscope simulations, for example, most systems have used standard polygon rendering techniques with synthetic texture mapping. Although the results obtained are visually appealing, they are not adaptable with regard to both structure and appearance. Texture mapping is usually uniform throughout the whole simulation, and even in cases where special visual effects for certain lesions are provided they are limited in both accuracy and adaptability. *In vivo* structures show considerable diversity in shape and texture, the challenge of generating realistic structure and surface properties has hindered the production of generic patient case databases for skills assessment.

Simulators require a geometric model of the world that the trainees explore. In current implementations, this is created artificially, but they could equally well be obtained from non-invasive tomographic imaging techniques. Recently, the feasibility of implementing such an idea for tracking camera motion and navigation planning has been investigated by a number of research groups¹⁻³. For the purpose of patient specific simulation, the correspondence between video bronchoscope images with matched virtual views from tomographic data provides the further possibility of extracting surface texture that is immune to changes in lighting and viewing position⁴. By augmenting virtual endoscopic views with patient specific structural and surface details, a more realistic simulation environment can be derived.

In order to match video bronchoscope images to the geometry extracted from three-dimensional reconstructions of the bronchi, robust registration techniques have to be developed. This is a challenging problem as it implies 2D/3D registration with certain degrees of deformation. Existing effort in 2D/3D registration typically employs intensity^{1,5-7} and feature based techniques⁸⁻¹¹. Both techniques involve optimising a similarity measure, which evaluates how close a 3D model viewed from a given camera pose is to the current 2D video frame. Intensity based techniques entail comparing a predicted image of the object with the 2D image without any structural analysis. With this approach, similarity measures such as cross-correlation^{2,3} and mutual information⁶ are typically used. Mutual information exploits the statistical dependency of two datasets and is particularly suitable for multi-

modal images. Existing methods, however, are based on special illumination conditions that may not match bronchoscope images⁶. Bronchoscope images are illuminated by a light source that is close to the tissue surface and are heavily affected by inter-reflections. In this case, light power decreases with the square of the distance from the light source and it is essential to adjust the illumination conditions of the rendered 3D model in order for the intensity-based techniques to work. The method, however, is further complicated by specular reflections due to surface fluid, which is difficult to model for simulated views.

As an alternative, feature based techniques depend on the alignment of corresponding image features, which are relatively immune to changes in lighting conditions. Since the density of visual features is generally sparse, the computational performance can be efficient. Furthermore, the method also offers the potential for dealing with local and global deformation. The fundamental challenge of this approach is in the reliability of the feature extractors to be used. Features purely based on visual appearance are not reliable due to the richness of surface texture typically observed in bronchoscope views, whereas in 3D tomographic images only salient geometrical features are preserved.

The purpose of this paper is to introduce a novel pq -space based 2D/3D registration technique by exploiting the unique geometrical constraints between the camera and the light source for endoscopic procedures. In the specific case of using perspective projection with a point light source near the camera, the use of intensity gradient can reduce the conventional shape-from-shading equations to a linear form, which suggests a local shape-from-shading algorithm that avoids the complication of changing surface albedos. We demonstrate how to use the derived pq -space distribution to match that of the 3D tomographic model. The major advantages of this method are that it depends neither on the illumination of the 3D model, nor on feature extraction and matching. Furthermore, the temporal variation of the pq distribution also permits the identification of localized deformation, which offers an effective way of excluding these areas from the registration process.

2. Methods

The main process of the proposed technique comprises the following major steps: the extraction of surface normals for each pixel of the video images by using a linear local shape-from-shading algorithm derived from the unique camera/lighting constraints of the endoscopes; extraction of the pq -components of the 3D tomographic model by direct z -buffer differentiation; and the construction of a similarity measure based on angular deviations of the pq -vectors derived from 2D and 3D data sets.

2.1 Shape from Shading

2.1.1 General Concepts

Shape from shading is a classical problem in computer vision that has been well established by the pioneering work of Horn^{12,13}. It addresses the problem of extracting both surface and relative depth information from a single image. Before moving on to the special case of extracting surface information from endoscope images, we need to define a few basic concepts of how image intensity is related to scene geometry. Horn¹³ identifies that image irradiance is related to scene radiance by the formula:

$$E(x, y) = L \frac{\pi}{4} \left(\frac{d}{f} \right)^2 \cos^4 \alpha, \text{ where } \tan \alpha = \frac{1}{f} \sqrt{x^2 + y^2} \quad (1)$$

$E(x, y)$ is the image irradiance. Irradiance is defined as the amount of light falling on a surface. L is the reflected radiance of the light source. Radiance is defined as the amount of light radiated from a surface. In Equation (1), d is the diameter of the lens, f is its focal length and α is the angle between the optical axis and the light ray going through the centre of the observed small solid angle as shown in Fig. 1.

This relationship combined with a reflectance map $R(p, q)$ leads to the image irradiance equation:

$$E(x, y) = c \cdot R(p, q) \quad (2)$$

where c is a constant. The Reflectance map directly relates the reflected radiance to the surface orientation, given a type of surface and distribution of light sources. The image irradiance equation is particularly useful in describing the relationship between image irradiance and surface orientation gradients (p, q) , which are the slopes of the surface in the x and y directions, respectively. In the above equations, $p-q$ gradients are directly related to surface normal as: $\vec{n} = (-p, -q, 1)$, where the sign depends on whether the normal is pointing inside or outside of the 3D-object. The surface $p-q$ gradients can also be seen as the x and y projections, respectively, of the normal into the camera coordinate system as shown in Fig. 1. Unfortunately, the reflectance map stated above is based on the assumption that the viewer and all light sources are distant from the object surface, because only under these assumptions we associate a unique value of image intensity with every surface orientation¹².

2.1.2 Shape from Shading for Bronchoscope Images

Recovering shape from shading from bronchoscope images is a special case because the light source is close to the camera and relatively close to the lumen surface. Hence conventional algorithms, such as those described in¹⁴⁻¹⁸, are not applicable. With these techniques, the basic assumption is that the angle between the viewing vector $\hat{V}$ and the Z-axis, α , is negligible when the object size is small compared

to its distance from the camera. In the case of endoscope images, both the camera and the light source are close to the object and the direction of the illuminating light coincides with the axis of the camera, thus no assumption can be made on α being negligible and lighting being uniform. Recently, Prados¹⁹ has studied the Lambertian shape-from-shading problem for pinhole camera with a point light source located at the optical centre. However, the intensity of the image is also affected from the distance between the surface point and the light source. Rashid and Okatani^{20,21} modelled this dependency by adding one more factor, which was a monotonically decreasing function $f(r)$ between the surface point and the light source. Therefore, the image irradiance, E , can be formulated as:

$$E(x, y) = s_0 \cdot \rho(x, y) \cdot \cos(i) \cdot f(r) \quad (1)$$

where s_0 is a constant related to the camera, ρ is the surface albedo and $\cos(i)$ is the angle between the incident light ray and the surface normal $\bar{n} = [p, q, -1]$. Within the context of this study, our main interest is focused on estimating the normal vectors but not to reconstruct the whole surface. Therefore, the linear technique of Rashid²⁰ was adopted because it can approximate well the gradient vectors p, q by using a linear local shape-from-shading algorithm. It has been proven that under the assumptions of the light source being close to the viewer and the surface being smooth and Lambertian, the following two linear equations with unknown p, q components can be written as:

$$\begin{cases} A_1 \cdot p_0 + B_1 \cdot q_0 + C_1 = 0 \\ A_2 \cdot p_0 + B_2 \cdot q_0 + C_2 = 0 \end{cases} \text{ where } \begin{cases} A_1 = (-x_0 \cdot R_x + 3) \cdot (1 + x_0^2 + y_0^2) - 3 \cdot x_0^2 \\ B_1 = -R_x \cdot (1 + x_0^2 + y_0^2) \cdot y_0 - 3 \cdot x_0 \cdot y_0 \\ C_1 = R_x \cdot (1 + x_0^2 + y_0^2) + 3 \cdot x_0 \\ A_2 = -R_y \cdot (1 + x_0^2 + y_0^2) \cdot x_0 - 3 \cdot x_0 \cdot y_0 \\ B_2 = (-y_0 \cdot R_y + 3) \cdot (1 + x_0^2 + y_0^2) - 3 \cdot y_0^2 \\ C_2 = R_y \cdot (1 + x_0^2 + y_0^2) + 3 \cdot y_0 \end{cases} \quad (4)$$

In the above equation, $R_x = E_x/E$ and $R_y = E_y/E$ are the normalized partial derivatives of the image intensities, E is the intensity of the pixel under consideration and x_0 and y_0 are the normalized image plane coordinates.

2.2 Extraction of pq-components from the 3D model

For tomographic images, the extraction of the p, q -components from the 3D model is relatively straightforward, since the exact surface representation is known. By making use of $p = \partial z / \partial x$ and $q = \partial z / \partial y$, differentiation of the z -buffer for the rendered 3D surface will result in the required p, q distribution, which also elegantly avoids the tasks of occlusion detection. The effect of perspective projection has been taken into account during the rendering stage. Modern video-based endoscopes have a wide-angle field of view to allow greater detail at the centre of the display. The frustum of the camera in the virtual world has been tuned accordingly. However, the wide-angle affects the p, q -space estimation derived from the z -buffer. In order to suppress this effect the p, q -space derived from the z -buffer was multiplied by a factor that is inversely proportional to the depth.

2.3 Similarity Measure

With the proposed pq -space based approach, an intuitive approach would be to use the angle between the surface normals extracted from shape-from-shading and those from the 3D model for constructing a minimization problem for 2D/3D registration. This, however, is not possible because the pq -vectors in the shape-from-shading algorithm have been scaled. The similarity measure used in this paper depends on the pq -components alone and the cross correlation between the two pq distributions are used.

Analytically, for each pixel of the video frame, a pq -vector corresponding to $\bar{n}_{img}(i, j) = [p_{i,j}, q_{i,j}]^T$ was calculated by using the linear shape-from-shading algorithm shown above. Similarly, for the current pose of the rendered 3D model, corresponding pq -vectors $\bar{n}_{3D}(i, j) = [p'_{i,j}, q'_{i,j}]^T$ for all rendered pixels were also extracted by differentiating the z -buffer. The similarity of the two images was determined by evaluating the dot product of corresponding pq -vectors:

$$\phi(\bar{n}_{3D}(i, j), \bar{n}_{img}(i, j)) = \frac{\|\bar{n}_{3D}(i, j) \cdot \bar{n}_{img}(i, j)\|}{\|\bar{n}_{3D}(i, j)\| \cdot \|\bar{n}_{img}(i, j)\|} \quad (5)$$

By applying a weighting factor that is proportional to the norm of $\bar{n}_{3D}$, the above equation reduces to

$$\phi_w(\bar{n}_{3D}(i, j), \bar{n}_{img}(i, j)) = \frac{\|\bar{n}_{3D}(i, j) \cdot \bar{n}_{img}(i, j)\|}{\|\bar{n}_{img}(i, j)\|} \quad (6)$$

Subsequently, by incorporating the mean angular differences and the associated standard deviations σ , the following similarity function is formed:

$$S = \frac{1}{\sum \sum (\phi_w) \cdot \sum \sum (\|1 - \sigma(\phi_w)\| \cdot \|\bar{n}_{3D}\|)} \quad (7)$$

By minimising the above equation, the optimum pose of the camera for the video image can be derived. The reason for introducing a weighting factor in Equation (7) is due to the fact that pq estimation from the 3D model is more accurate than that of the shape-from-shading algorithm. This is because it is not affected by surface textures, illumination conditions or surface reflective properties. The weighting factor therefore reduces the potential impact of erroneous pq -values from the shape-from-shading algorithm and improves the overall robustness of the registration process.

2.4 Tissue Deformation

With pq -space representation, the angle between the normal vectors before and after rigid body transformation will remain the same for every surface point. Local deformation can therefore be identified at surface points where the angle diverts from the mean angle of the whole 3D model. Localized inter-frame deformation can therefore be isolated and excluded for the pose estimation

process. In this study, we used the pq deformation map, as a weighting factor during the registration process such that the weighting provided was inversely proportional to the amount of deformation detected.

2.5 Validation

The proposed method was implemented in Microsoft Visual C++ on a conventional PC machine (2 GHz Intel Pentium 4 processor, 512MByte main memory, nVidia GeForce 4 MX 440 graphics card, with Microsoft Windows 2000 operating system). Surface rendering was implemented using OpenGL. The interface was based on FLTK (www.fltk.org).

2.5.1 Phantom Validation

In order to assess the accuracy of the proposed algorithm, an airway phantom made of silicone rubber and painted with acrylics, as shown in Fig. 2(a), was constructed. The phantom has a cross sectional diameter of 12cm at the opening and narrows down to 5cm at the far end. The inside face was coated with silicone-rubber mixed with acrylic for surface texturing and left to cure in the open air. This gives the surface a specular finish that looks similar to the surface of the lumen as shown in Fig. 2(b). A real-time, six degrees-of-freedom Electro-Magnetic (EM) motion tracker (FASTRAK, Polhemus) was used to validate the 3D camera position and orientation, as illustrated in Fig. 2(c). The EM-tracker has an accuracy of 0.762mm RMS. The tomographic model of the phantom was scanned with a Siemens Somatom Volume Zoom four-channel multi-detector CT scanner with a slice thickness of 3mm and in-plane resolution of 1mm. A CMOS camera and NTSC standard with frame rate of 29.97fps was used.

2.5.2 In-vivo Validation

For preliminary *in vivo* validation, bronchoscopy examination was performed in one patient according to a conventional clinical protocol. During the bronchoscope procedure a prototype videoscope (Olympus BF Type; with field of view 120°) was used. Video images from the bronchoscopic examination were transferred to digital videotapes in PAL format at 25fps. Since the original endoscopic video frames contain both the endoscopic image and redundant black background, only the endoscopic view was digitised and cropped to images of 454×487 pixels. All images were converted to greyscale before the pq -space analysis. Similar to the phantom study, the CT images were acquired from the Siemens Somatom Volume Zoom four-channel multi-detector CT scanner with a slice width of 3mm and collimation of 1mm, and the acquisition volume covered from the aortic arch to the dome of hemi-diaphragm.

Pre-processing of the video images was necessary in order to alleviate the effects of interlacing, lens distortion and unnecessary texture information. These steps are schematically illustrated in Fig. 3. De-interlacing is important as temporal mis-match of odd-even frames can introduce significant errors to the pq -space estimation. To correct for barrel distortion of the bronchoscope camera due to the wide-angle lens the method proposed by Heikkila *et al*²² was used. ***In general, methods that correct for ‘barrel’ distortion must calculate a distortion centre and correct for both radial and tangential components. Radial distortion is the most commonly used correction and usually dominates the distortion function. It causes the actual image plane to be displayed radially in the image plane. Tangential distortion is due to ‘decentring’ or imperfect centring of the lens components and other manufacturing defects. The initialization of the calibration parameters follows the method proposed by Zhang et al^{23,24} where a closed-form solution is used. To remove noise and image artefacts, anisotropic filtering was applied to each image²⁵. The method uses a local orientation and an anisotropic measure of level contours to control the shape and extent of the filter kernel and thus ensures that corners and edges are well preserved through out the filtering process.***

3 Results

3.1 Phantom study

Fig. 2(c) demonstrates an example video frame of the bronchoscope phantom used to validate the proposed algorithm. The derived pq -vector distribution using the linear shape-from-shading algorithm is shown in Fig. 2(d). It is evident that the derived pq -vectors are relatively immune to lighting changes, and these vectors were then used to estimate the 3D pose of the camera used to capture the video frame as shown in Fig. 2(c).

To assess the accuracy of the proposed algorithm in tracking camera poses in 3D, Fig. 4 and Fig. 5 compare the relative performance of the traditional intensity based technique and EM tracked poses against those from the new method. Since the tracked pose has six degrees-of-freedom, we used the distance travelled and inter-frame angular difference as a means of error assessment. The video acquired has 1.73 minutes duration (25fps). A continuous part of 40 sec (1000 frames) has been used for tracking. Traditional cross-correlation with the illumination conditions manually adjusted failed when the bronchoscope gets through the main bifurcation to the left bronchi and faces directly the far end of the tubular phantom airway, (394 frame). Three different variations of the pq -space technique were tested and compared against the gold standard EM tracker data (black line) and the intensity based technique (blue line). As expected the intensity-based technique is highly sensitive to lighting condition changes, and with manual intensity adjustments, the convergence of this method is improved. However, the proposed pq -space registration has significantly consistent results, which were close to those measured by the EM tracker. It is also evident that the weighting factor affects greatly the

performance of the method and can enhance significantly its accuracy. By choosing a weight related to the z-buffer variations, errors caused by variations in texture and surface reflection can be minimised.

The effect of localised deformation on the pq -space representation is illustrated in Fig. 6, where Fig. 6(a) is the original video bronchoscope image and Fig. 6(b) is the derived pq -space deformation map. Fig. 6(c) and Fig. 6(d) demonstrate the accuracy of the pose estimation with the traditional intensity based technique and the proposed pq -space registration with deformation weighting, respectively. With tissue deformation, the intensity-based technique has introduced significant error, despite careful adjustment of illumination conditions.

3.2 In-vivo validation

Quantitative results from the *in vivo* validation are demonstrated in Fig. 7, where sample frames from the video sequence are displayed. The proposed pq -space based registration has been applied to a video sequence of 31sec (797 frames) of the one-patient study. The bronchoscope video sequence starts from the main bifurcation and continues through the left bronchi. Visual inspection of the real and the virtual endoscope images proves that the pq -based registration technique can track the tip of the bronchoscope relatively accurately and is stable under sudden movements or large rotation angles. However, similar to Mori^{2,3}, when mis-tracking occurs in one frame, tracking of subsequent frames almost always fails, as the initial starting position deviates too far away from the correct result. The reason that the tracking sequence is limited to only 31sec (797frames) is that bubbles and deformation occlude and distort the anatomical features, respectively and thus pq -based registration fails to work under these conditions.

Fig. 8 and Fig. 9 demonstrate the quantitative comparison between pq -based registration and manual alignment, similar to that of phantom validation (Fig. 4 and Fig. 5). The experimental results suggest that the proposed method can track the bronchoscope tip satisfactorily. Manual alignment always entails an error. To assess this, the distance from the initial reference position and the angle from the initial vector are displayed in Fig. 10 for both the EM tracking data and the manual alignment. The measurements have been done by using the airway phantom and the EM tracker. Analogous to phantom validation, a video sequence has been tracked for more than 400 frames using the EM tracker. For the same sequence we used manual alignment for every 10frames to get the position of the virtual camera that matches best the corresponding video view. The average positional error is equal to 3mm with sd 2.26mm and the angular error is 0.0381rad with sd 0.0285rad. This indicates that the error is consistent throughout the video sequence and relatively small. We expect that this error will be smaller in real bronchoscope images as the phantom's scale is greater to the real size of the airways.

Finally, we investigated the limitations of the proposed technique in a clinical environment. Fluids such as blood and mucus dynamically change the appearance of the lumen, as shown in Fig. 11(a-b). Appearance of bubbles is common and usually covers the whole video frame resulting in a failure of the registration algorithm. Respiratory motion and extreme breathing patterns deform the airways

significantly and distort severely the anatomical features that are essential in 2D/3D registration. An example of large tissue deformation is shown in Fig. 11(c-d). A process of identifying these phenomena combined with temporal information would facilitate tracking of the bronchoscope during the whole procedure.

4 Discussion and Conclusions

We have proposed a new pq -space based 2D/3D registration method for matching camera poses of bronchoscope videos. The results indicate that based on the pq -space and the 3D model, reliable bronchoscope tracking can be achieved. The main advantage of the method is that it is not affected by illumination conditions. The method depends on the weighting factor that has been introduced to the similarity measure, which has been shown to be resilient to variations of surface texture details. The results from this study show that the intensity-based technique typically fails when the bronchoscope has passed through the main bifurcation to the bronchi. This is due to the difference in anatomy, which affects the inter-reflectance lighting and subsequently results in changes in the overall illumination conditions. The pq -space based technique, however, is less sensitive to these changes and furthermore the proposed technique does not require the extraction of explicit feature vectors.

Preliminary results indicate that deformation of the airways can be identified and excluded from the similarity measure with the proposed registration framework. However, in doing so we have assumed that the bronchoscope moves smoothly through the airways and the deformation is limited to a small part of the video frame. Further investigation of the accuracy of the method for large tissue deformation is required. It should also be noted that there are a number of factors that can affect the accuracy of the pq -space algorithm. The 3D reconstruction of the tracheobronchial tree can involve artefacts due to respiratory motion and partial volume effects. Since the respiratory status during imaging and video bronchoscope examination is not matched, the 3D airway anatomy can differ significantly from the dynamic appearance of the tracheobronchial tree during the examination. In this case, image-based techniques can fail when the anatomical features of the video frames have been significantly altered due to deformation or occlusion. Furthermore, bronchoscope cameras typically have a wide angle, which can also cause an adverse effect to the accuracy of pq -space estimation despite the use of distortion correction. With the proposed method, the intrinsic robustness of the technique is inherently dependent on the performance of the shape-from-shading method used. The use of camera/lighting constraints of the bronchoscope greatly simplifies the 3D pose estimation of the camera, but it is important to note that shape-from-shading is severely affected by specularities and inter-reflectance caused by mucus on the lumen surface. A number of improvements can therefore be introduced for further improving of the accuracy of the proposed framework by explicit incorporation of the effect of mutual illumination, inter-reflectance and the specular components. Furthermore, temporal information has the potential to stabilize the registration result, enhance the accuracy of the deformation extraction, and decrease the convergence time. Restriction to the camera orientation according to a central path

can also improve the progress of the optimization algorithm and the temporal smoothness of the resulting virtual video. Finally, with the advent of *in vivo* miniaturised catheter tip tracking devices, it is possible to significantly improve the robustness and accuracy of current data registration techniques in the presence of large tissue deformation.

Acknowledgments

The authors would like to acknowledge the financial support of the EPSRC, the Royal Society, and the Wolfson Foundation. They are also grateful to Pallav Shah (MD), and Athol Wells (MD) for their contribution to the clinical data collection for this paper.

5 References

1. Helferty JP, Higgins WE. Combined Endoscopic Video Tracking and Virtual 3D CT Registration for Surgical Guidance. 2002 22-25 September 2002; Rochester, New York.
2. Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki J-i, Jr. CRM, Takabatake H, Natori H. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis* 2002;6(3):181-336.
3. Mori K, Suenaga Y, Toriwaki J-i, Hasegawa J-i, Kataa K, Takabatake H, Natori H. A method for tracking camera motion of real endoscope by using virtual endoscopy system. *Proceedings of SPIE*; 2000 12-18 February 2000; San Diego, California, USA. p 122-133. (*Proceedings of SPIE*).
4. Chung AJ, Deligianni F, Shah P, Wells A, Yang G-Z. Enhancement of Visual Realism with BRDF for Patient Specific Bronchoscopy Simulation. 2004 September 2004; Rennes, France.
5. Studholme C, Hill DLG, Hawkes DJ. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. *Pattern Recognition* 1998;32(1):71-86.
6. Viola P, III WMW. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* 1997;24(2):137-154.
7. Likar B, Pernus F. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing* 2001;19(1-2):33-44.
8. Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 2003;89(2-3):114-141.
9. David P, DeMenthon D, Duraiswami R, Samet H. SoftPOSIT: Simultaneous Pose and Correspondence Determination. 2002 27 May- 2 June 2002; Copenhagen.
10. Gold S, Rangarajan A, Lu CP, Pappu S, Mjolsness E. New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence. *Pattern Recognition* 1998;31(8):1019-1031.
11. DeMenthon D, David P, Samet H. SoftPOSIT: An Algorithm for Registration of 3D Models to Noisy Perspective Images Combining Softassign and POSIT: Center for Automation Research, Computer Science Department, UCLA; 2001 May 2001. Report nr CAR-TR-969, CS-TR-4257.
12. Horn BKP. Understanding Image Intensities. *Artificial Intelligence* 1977;8(2):201-231.
13. Horn BKP. *Robot Vision*. Cambridge: MIT Press; 1986.
14. Bichsel M, Pentland AP. A simple algorithm for shape from shading. 1992 15-18 June 1992; Champaign, IL USA. p 459 - 465.
15. Oliensis J. Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, and Image Processing: Image Understanding* 1991;54(2):163-183.
16. Oliensis J, Dupuis P. A global algorithm for shape from shading. 1993 11-14 May 1993; Berlin Germany. *IEEE*. p 692-701.
17. Tsai P-S, Shah M. A fast linear shape from shading. 1992 15-18 June 1992; Champaign, IL USA. p 734 - 736.
18. Worthington PL, Hancock ER. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1999;21(12):1250-1267.
19. Prados E, Faugeras O. A rigorous and realistic Shape From Shading method and some of its applications. Sophia Antipolis , France: INRIA, Odysse Lab; 2004 March 2004. Report nr RR-5133.
20. Rashid HU, Burger P. Differential algorithm for the determination of shape from shading using a point light source. *Image and Vision Computing* 1992;10(2):119-127.
21. Okatani T, Deguchi K. Shape Reconstruction from an Endoscope Image by shape from shading technique for a point light source at the projection centre. *Computer Vision and Image Understanding* 1997;66(2):119-131.
22. Heikkila J, Silven O. A Four-step Camera Calibration Procedure with Implicit Image Correction. 1997 17-19 June 1997; San Juan, Puerto Rico. p 1106-1112.
23. Zhang C, Helferty JP, McLennan G, Higgins WE. Nonlinear distortion correction in endoscopic video images. 2000 10-13 September 2000; Vancouver, BC Canada. *IEEE*. p 439 - 442.

24. Zhang Z. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In: IEEE, editor; 1999 20-25 September 1999; Corfu, Greece. p 666-673.
25. Yang GZ, Burger P, Firmin DN, Underwood SR. Structure adaptive anisotropic image filtering. Image and Vision Computing 1994;14(2):135-145.

6. Figure Captions

Figure 1. A schematic illustration of the image formation process. The Z-axis of the camera coordinate system is towards the image plane, and N is the normal of a surface point. The surface gradient at a given point is represented as p and q , which are the slopes of the surface in the x and y direction, respectively. In the case of endoscope images, the viewing vector V coincides with the lighting vector L . Since the camera is close to the object and the angle between the viewing vector and the optical axis is not negligible.

Figure 2. a) An airway phantom made of silicone rubber and painted with acrylics was constructed in order to assess the accuracy of the pq -based registration. b) A sample bronchoscope video frame from the phantom used to reproduce the airway structures. c) A real-time six DOF EM motion tracker (FASTRAK, Polhemus) used to validate the 3D camera position and orientation. d) The pq -vector distribution derived from the linear shape-from-shading algorithm by exploiting the unique camera/lighting constraints.

Figure 3. The pre-processing steps applied to the bronchoscope videos before 2D/3D registration. a) Original video frame acquired from the prototype bronchoscope, b) de-interlaced video frame, c) after lens distortion correction, and d) final smoothed image by using an anisotropic filter that preserves local geometrical features.

Figure 4. Euclidean distance between the first and subsequent camera positions as measured by four different tracking techniques corresponding to the conventional intensity based 2D/3D registration with or without manual lighting adjustment, the EM tracker (which is used as the gold standard for this study), and the proposed pq -space registration technique.

Figure 5. Inter-frame angular difference at different time of the video sequence, as measured by the four techniques described in Fig. 4.

Figure 6. a) A video frame from a deformed airway phantom, b) the associated pq -space deformation map where bright intensity signifies the amount of deformation detected. (c) The superimposed 3D rendered image with pose estimated from intensity-adjusted registration and pq -space registration with deformation weightings, respectively.

Figure 7. Example results of *in vivo* camera tracking for the patient studied in this paper. The left column shows samples of real bronchoscopic images and the right column presents the matched virtual bronchoscopic images after pq -space based 2D/3D registration.

Figure 8. The accuracy of the registration result represented as the Euclidean distance between the first and subsequent camera positions as measured by the pq -based 2D/3D registration compared to manual alignment.

Figure 9. Inter-frame angular difference as measured by the proposed pq -based 2D/3D registration method and manual alignment.

Figure 10. Assessment of manual alignment error compared to the EM tracker as assessed by using the 3D bronchial model. a) Euclidean distance between the first and subsequent camera positions as measured by the EM tracker and after manual alignment with step equal to 10. b) Inter-frame angular difference between manual alignment and readings from the EM tracker.

Figure 11. Common image artefact that can affect image-based 2D/3D registration techniques: a) excessive bleeding due to pathology, b) appearance of bubbles when patient coughs, and c-d) large tissue deformation between successive image frames.