Computer Aided Surgery, 2004; 9(5): 215–226

Taylor & Francis
Taylor & Francis Group

BIOMEDICAL PAPER

Patient-specific bronchoscope simulation with *pq*-space-based 2D/3D registration

FANI DELIGIANNI, ADRIAN CHUNG, & GUANG-ZHONG YANG

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

(Received 31 August 2004)

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 3D tomographic data with patient-specific bronchoscope videos. To facilitate matching of 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 was assessed by using an electromagnetic tracker and a specially constructed airway phantom. Preliminary *in vivo* validation of the proposed method was performed on a matched patient bronchoscope video sequence and 3D CT data. Comparison to existing intensity-based techniques was also 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 such areas from the registration process.

Keywords: *Shape from shading, pq-space, 2D/3D registration, virtual endoscope, virtual bronchoscopy, surgical simulation*

Introduction

Bronchoscopy is a special form of minimal-access surgical procedure for examining the trachea, bronchi, and air passages that lead to the lungs. Due to the complexity of instrument control, restricted vision and mobility, and lack of tactile perception, it requires a high degree of manual dexterity and hand-eye coordination. As with other forms of minimal-access surgery, the required surgical skills are normally obtained through practice on inanimate plastic models and subsequently on patients. However, with plastic models it is difficult to provide the high-fidelity physical responses that are necessary for advanced skills training and assessment, while practicing on real subjects prolongs the examination time, can involve considerable discomfort to the patient, and carries certain risks of complications.

With the maturity of augmented-reality systems, there has been an increasing demand for integration of computer simulation in certain aspects of this training, particularly for developing hand-eye coordination and instrument control. In most of the current simulation systems, however, the degree of visual

Correspondence: Professor G.Z. Yang, Royal Society/Wolfson Foundation MIC Lab, Department of Computing, 180 Queen's Gate, Imperial College, London SW7 2BZ, UK. Tel: (+44) 20 7594 8441. Fax: (+44) 20 7581 8024. E-mail: g.z.yang@imperial.ac.uk. Website: http://vip.doc.ic.ac.uk
This paper is based on research presented at the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Montreal, Canada, November 2003.

ISSN 1092-9088 print/ISSN 1097-0150 online ©2004 Taylor & Francis
DOI: 10.1080/10929080500144927

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 either structure or 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, and the challenge of generating realistic structures 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 will explore. In current implementations, this is created artificially, but 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 groups [1–3]. For the purpose of patient-specific simulation, the correspondence between video bronchoscope images and matched virtual views from tomographic data provides the further possibility of extracting surface texture that is immune to changes in lighting and viewing position [4]. By augmenting virtual endoscopic views with patient-specific structural and surface details, a more realistic simulation environment can be derived.

To match video bronchoscope images to the geometry extracted from 3D reconstructions of the bronchi, robust registration techniques must be developed. This is a challenging problem as it implies 2D/3D registration with certain degrees of deformation. Existing efforts in 2D/3D registration typically employ intensity- [1, 5–7] and feature-based [8–11] techniques. Both approaches involve optimizing 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 information [6] 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 images [6]. 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 in 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 lies 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 an 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 the use of 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 deformations, which offers an effective way of excluding these areas from the registration process.

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 using a linear local shape-from-shading algorithm derived from the unique camera/lighting constraints of the endoscopes; the 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.

Shape from shading

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 some basic concepts of how image intensity is related to scene geometry. Horn [13] identifies that image irradiance is related to scene radiance by the formula

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

$E(x, y)$ 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 center of the observed small solid angle, as shown in Figure 1.

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

$$E(x, y) = c \cdot R(p, q) \tag{2}$$

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 the surface normal by $\hat{N} = (-p, -q, 1)$, where the sign depends on whether the normal is pointing inside or outside 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 Figure 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 do we associate a unique value of image intensity with every surface orientation [12].

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 in references 14–18, 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, so no assumption can be made regarding α being negligible and lighting being uniform. Recently, Prados and colleagues [19] have studied the Lambertian shape-from-shading problem for a pinhole camera with a point light source located at the optical center. However,

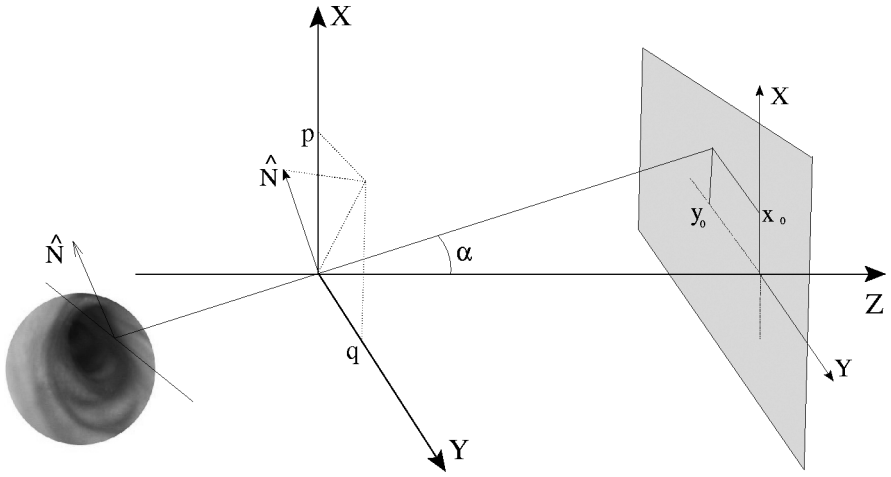

Figure 1. A schematic illustration of the image formation process. The Z-axis of the camera coordinate system is towards the image plane, and N is the normal of a surface point. The surface gradient at a given point is represented as p and q, which are the slopes of the surface in the x and y directions, 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, the angle between the viewing vector and the optical axis is not negligible.

the intensity of the image is also affected by the distance between the surface point and the light source. Rashid and Okatani [20,21] modeled 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

$$E(x, y) = s_0 \cdot \rho(x, y) \cdot \cos(i) \cdot f(r) \tag{3}$$

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 $\hat{N}$. In this study, our main interest is focused on estimating the normal vectors, rather than reconstructing the whole surface. Therefore, the linear technique of Rashid [20] was adopted because it can approximate well the gradient vectors pq 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

$$\begin{cases} A_1 \cdot p_0 + B_1 \cdot q_0 + C_1 = 0 \\ A_2 \cdot p_0 + B_2 \cdot q_0 + C_2 = 0 \end{cases} \quad \text{where}$$

$$\begin{cases} A_1 = (-x_0 \cdot R_x + 3) \cdot \left(1 + x_0^2 + y_0^2\right) - 3 \cdot x_0^2 \\ B_1 = -R_x \cdot \left(1 + x_0^2 + y_0^2\right) \cdot y_0 - 3 \cdot x_0 \cdot y_0 \\ C_1 = R_x \cdot \left(1 + x_0^2 + y_0^2\right) + 3 \cdot x_0 \\ A_2 = -R_y \cdot \left(1 + x_0^2 + y_0^2\right) \cdot x_0 - 3 \cdot x_0 \cdot y_0 \\ B_2 = \left(-y_0 \cdot R_y + 3\right) \cdot \left(1 + x_0^2 + y_0^2\right) - 3 \cdot y_0^2 \\ C_2 = R_y \cdot \left(1 + x_0^2 + y_0^2\right) + 3 \cdot y_0 \end{cases} \tag{4}$$

$R_x = E_x/E$ 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.

Extraction of pq-components from the 3D model

For tomographic images, the extraction of the pq-components from the 3D model is relatively straightforward, since the exact surface representation is known. By making use of $p = \delta z/\delta x$ and $q = \delta z/\delta y$, differentiation of the z-buffer for the rendered 3D surface will result in the required pq 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 center of the display. The frustum of the camera in the virtual world has been tuned accordingly. However, the wide angle affects the pq-space estimation derived from the z-buffer. To suppress this effect, the pq-space derived from the z-buffer was multiplied by a factor that was inversely proportional to the depth.

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 to construct 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.

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 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 extracted by differentiating the z-buffer. The similarity of the two images was determined by evaluating the dot product of corresponding pq-vectors:

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

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

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

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

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

By minimizing the above equation, the optimum pose of the camera for the video image can be derived. A weighting factor is introduced in Equation (7) because pq estimation from the 3D model is more accurate than that obtained with the shape-from-shading algorithm, since 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.

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, and localized inter-frame deformation can thus be isolated and excluded from 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.

Validation

The proposed method was implemented in Microsoft Visual C++ on a conventional PC (2-GHz Intel Pentium 4 processor, 512 MB 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).

Phantom validation. To assess the accuracy of the proposed algorithm, an airway phantom was constructed of silicone rubber and painted with acrylics, as shown in Figure 2a. The phantom has a cross-sectional diameter of 12 cm at the opening, narrowing to 5 cm 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 Figure 2b. A real-time, six-degree-of-freedom electromagnetic (EM) motion tracker (FASTRAK, Polhemus, Colchester, VT) was used to validate the 3D camera position and orientation, as seen in Figure 2c. The EM tracker has an accuracy of 0.762 mm RMS. The tomographic model of the phantom was scanned with a Siemens Somaton Volume Zoom 4-channel multi-detector CT scanner with a slice thickness of 3 mm and in-plane resolution of 1 mm. A CMOS camera and NTSC standard with frame rate of 29.97 fps was used.

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 a field of view of 120°) was used. Video images from the bronchoscopic examination were transferred to digital videotapes in PAL format at 25 fps. Since the original endoscopic video frames contain both the endoscopic image and redundant black background, only the endoscopic view was digitized and cropped to

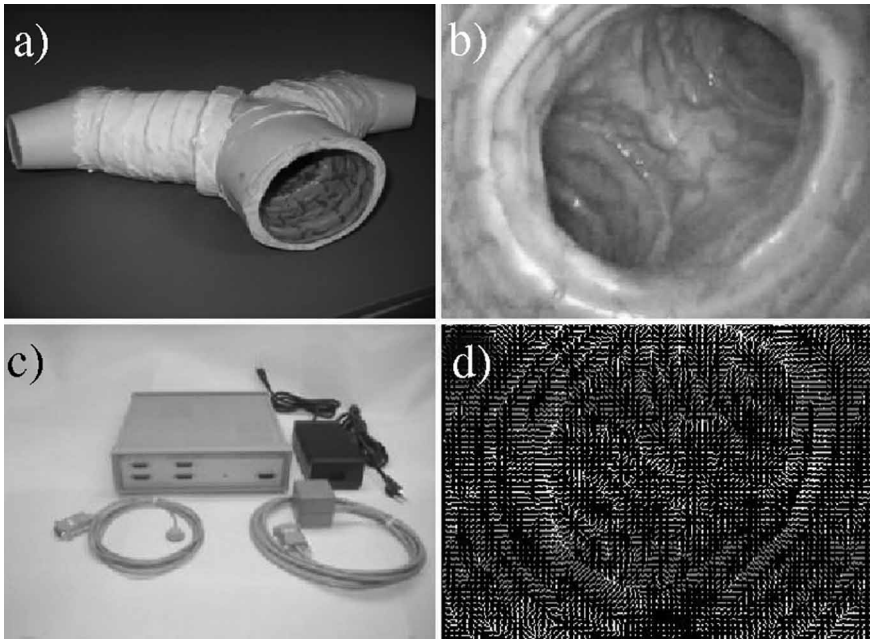

Figure 2. a) An airway phantom made of silicone rubber and painted with acrylics was constructed to assess the accuracy of the *pq*-based registration. b) A sample bronchoscope video frame from the phantom used to reproduce the airway structures. c) A real-time 6DOF EM motion tracker (FASTRAK, Polhemus) 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.

images of 454 × 487 pixels. All images were converted to grayscale before the *pq*-space analysis. As with the phantom study, the CT images were acquired from the Siemens Somaton Volume Zoom 4-channel multi-detector CT scanner with a slice width of 3 mm and collimation of 1 mm, and the acquisition volume ranged from the aortic arch to the dome of the hemi-diaphragm.

Pre-processing of the video images was necessary to alleviate the effects of interlacing, lens distortion and unnecessary texture information. These steps are schematically illustrated in Figure 3. De-interlacing is important as temporal mis-match of odd-even frames can introduce significant errors into 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 center 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 "decentering" or imperfect centering of the lens components and other manufacturing defects. The initialization of the calibration parameters follows the method proposed by Zhang et al. [23,24], in which a closed-form solution is used. To remove noise and image artefacts, anisotropic filtering was applied to each image [25]. 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 throughout the filtering process.

Results

Phantom study

Figure 2b shows a sample 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

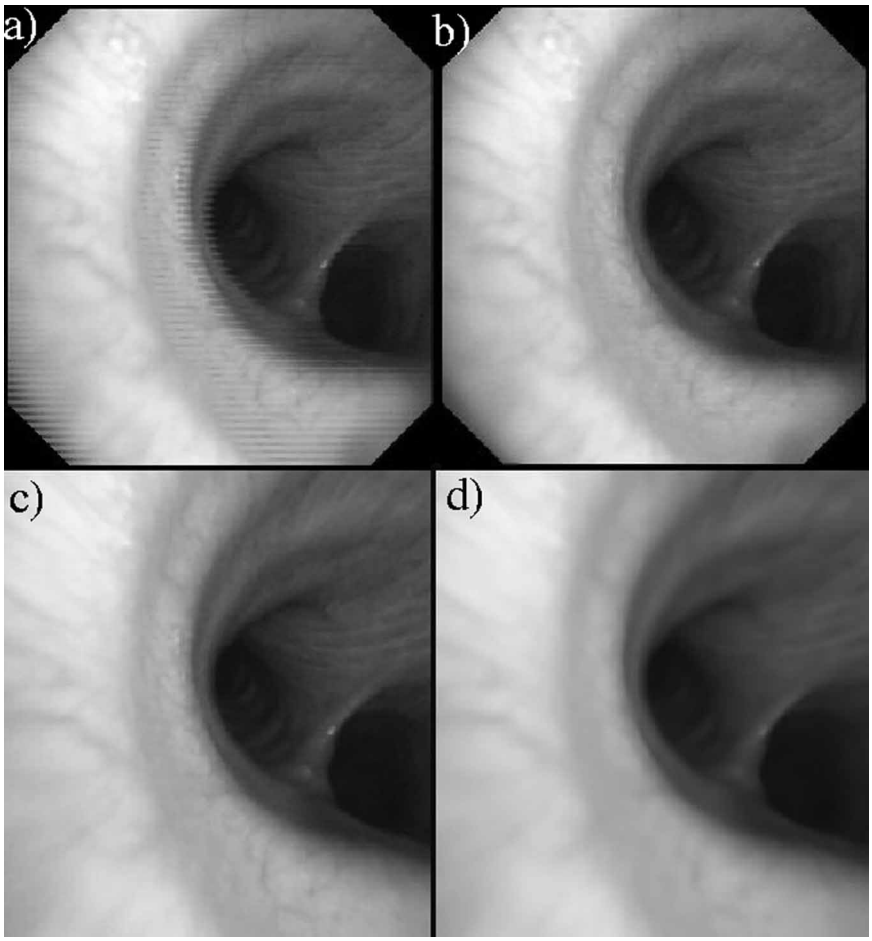

Figure 3. The pre-processing steps applied to the bronchoscope videos before 2D/3D registration. (a) Original video frame acquired from the prototype bronchoscope; (b) de-interlaced video frame; (c) after lens distortion correction; and (d) final smoothed image obtained using an anisotropic filter that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

Figure 2d. It is evident that the derived *pq*-vectors are relatively immune to lighting changes, and these vectors were then used to estimated the 3D pose of the camera used to capture the video frame, as shown in Figure 2c.

To assess the accuracy of the proposed algorithm in tracking camera poses in 3D, Figures 4 and 5 compare the relative performance of the traditional intensity-based technique and EM-tracked poses against that of the new method. Since the tracked pose has 6 degrees of freedom, we used the distance traveled and inter-frame angular difference as a means of error assessment. The video acquired had a duration of 1.73 min (25 fps). A continuous 40-second segment (1000 frames) was used for tracking. Traditional cross-correlation with the illumination conditions adjusted manually failed when the bronchoscope passed through the main bifurcation to the left bronchus and directly faced the far end of the tubular phantom airway (frame 394). 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 was highly sensitive to lighting condition changes, and manual intensity adjustments improved the convergence of this method. 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 greatly affects the performance of the method and can significantly enhance its accuracy. By choosing a weight related to the z-buffer variations, errors caused by variations in texture and surface reflection can be minimized.

The effect of localized deformation on the *pq*-space representation is illustrated in Figure 6, where 6a is the original video bronchoscope image and 6b is the derived *pq*-space deformation map. Figures 6c and 6d 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.

In-vivo validation

Qualitative results from the *in vivo* validation are shown in Figure 7, where sample frames from the video sequence are displayed. The proposed *pq*-space-based registration has been applied to a 31-second video sequence (797 frames) of the one-patient study. The bronchoscope video sequence starts from the main bifurcation and continues through the left bronchus. Visual inspection of the real and virtual endoscope images proves that the *pq*-based registration technique can track the tip of the bronchoscope relatively accurately and is stable in case of sudden movements or large rotation angles. However, as reported by Mori et al. [2,3], mis-tracking in one frame almost always leads to tracking failure in subsequent frames, as the initial starting position deviates too far from the correct result. The tracking sequence is limited to only 31 s (797 frames) because bubbles and deformation occlude and distort the anatomical features, and *pq*-based registration fails to work under these conditions.

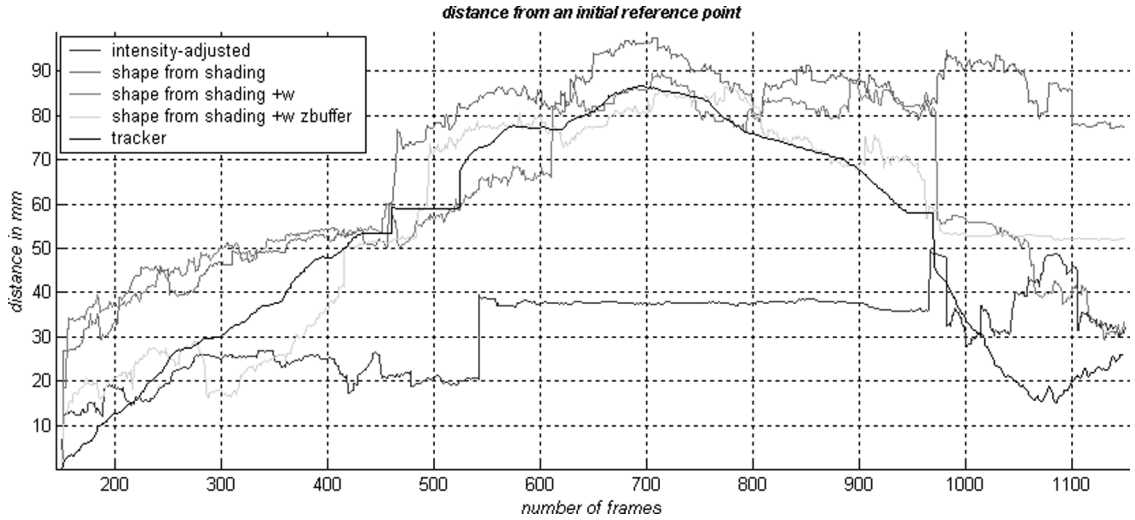

Figure 4. Euclidean distance between the first and subsequent camera positions as measured by four different tracking techniques corresponding to the conventional intensity-based 2D/3D registration with or without manual lighting adjustment, the EM tracker (which is used as the gold standard for this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

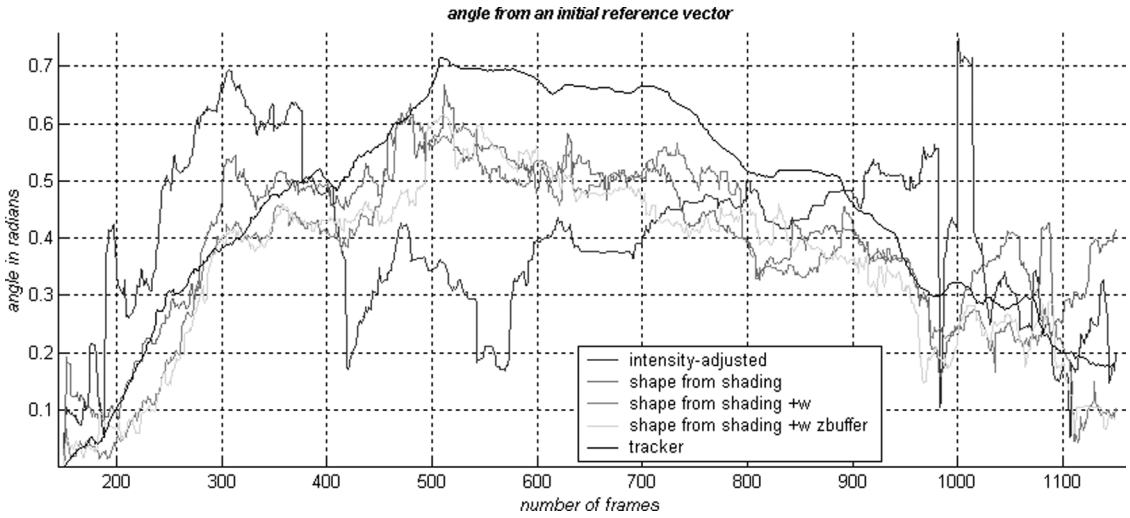

Figure 5. Inter-frame angular difference at different times in the video sequence, as measured by the four techniques listed in Figure 4.

Figures 8 and 9 show the quantitative comparison between *pq*-based registration and manual alignment, similar to that for phantom validation (Figures 4 and 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 Figure 10 for both the EM tracking data and the manual alignment. The measurements were performed using the airway phantom and EM tracker. Analogous to the phantom validation, a video sequence was tracked for more than 400 frames using the EM tracker. For the same sequence, we used manual alignment every 10 frames to obtain the position of the virtual camera that best matches the corresponding video view. The average positional error is equal to 3 mm

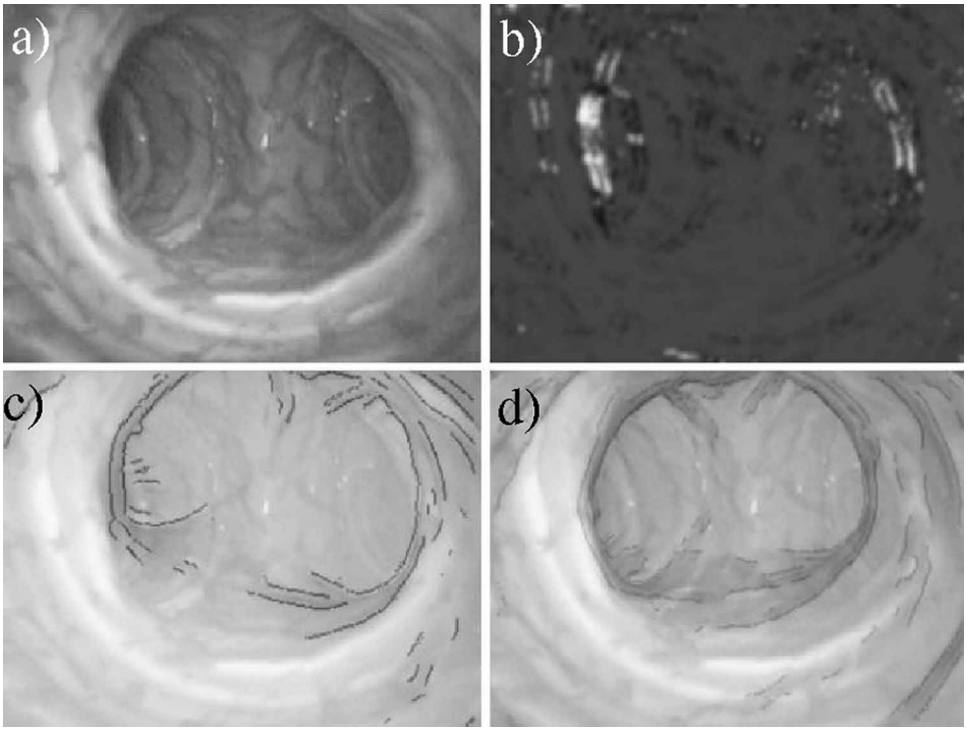

Figure 6. (a) A video frame from a deformed airway phantom. (b) The associated *pq*-space deformation map, where bright intensity signifies the amount of deformation detected. (c) and (d) The superimposed 3D rendered image with pose estimated from intensity-adjusted registration and *pq*-space registration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

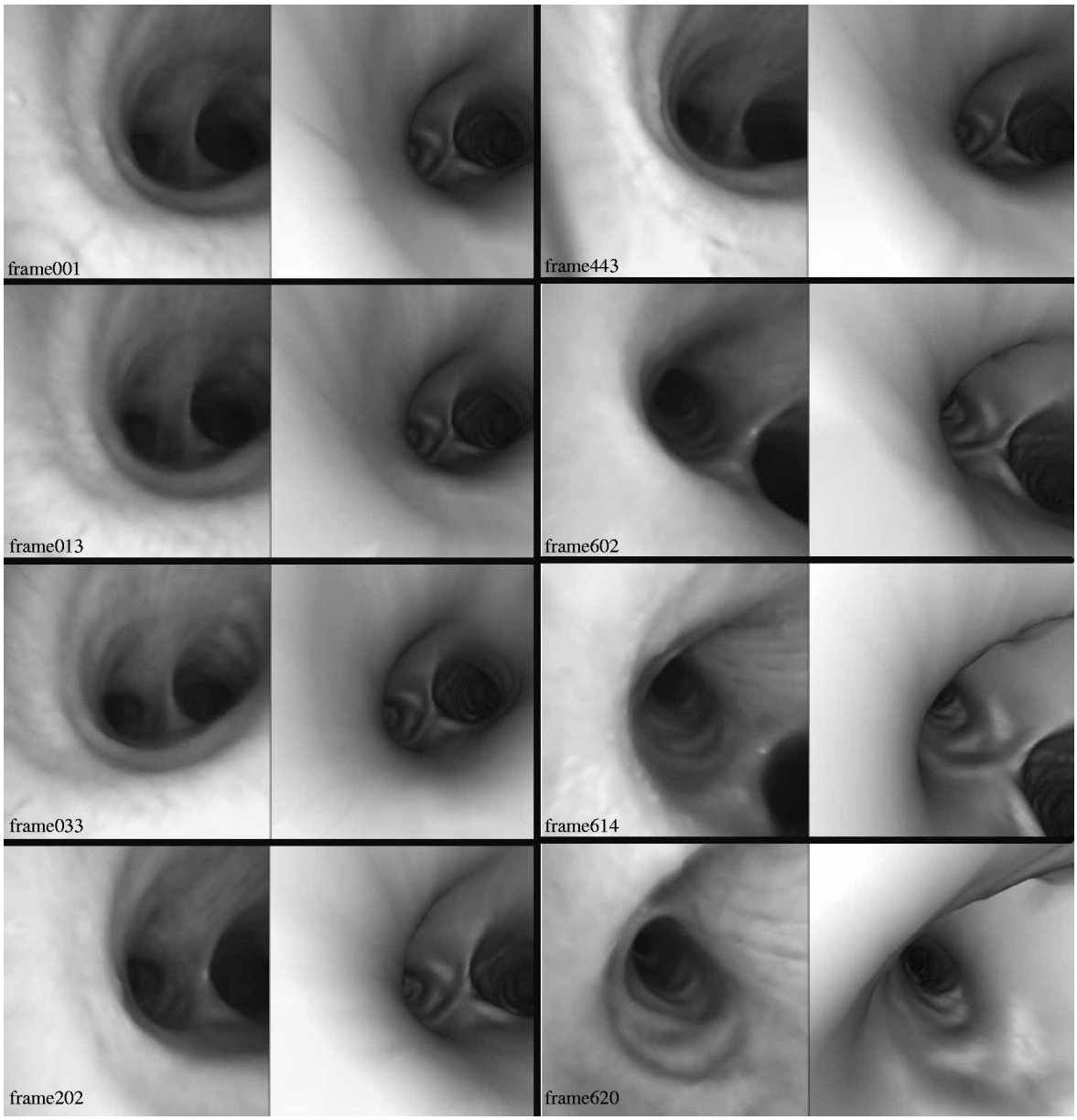

Figure 7. Sample results of *in vivo* camera tracking for the patient studied in this paper. The left column shows real bronchoscopic images and the right column presents the matched virtual bronchoscopic images after *pq*-space-based 2D/3D registration.

(SD = 2.26 mm) and the angular error is 0.0381 rad (SD = 0.0285 rad). This indicates that the error is consistent throughout the video sequence and relatively small. We expect this error to be smaller in real bronchoscope images, as the phantom is scaled larger than life size.

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 Figures 11a and 11b. Bubbles are common and usually cover the whole video frame, resulting in failure of the registration algorithm. Respiratory motion and extreme breathing patterns significantly deform the airways and severely distort the anatomical features that are essential for 2D/3D registration. An example of large tissue deformation is shown in Figures 11c and 11d. A process for identifying these phenomena combined with temporal information would facilitate tracking of the bronchoscope throughout the whole procedure.

Discussion and conclusions

We have proposed a new *pq*-space-based 2D/3D registration method for matching camera poses in bronchoscope videos. The results indicate that, based on the *pq*-space and the 3D model, reliable

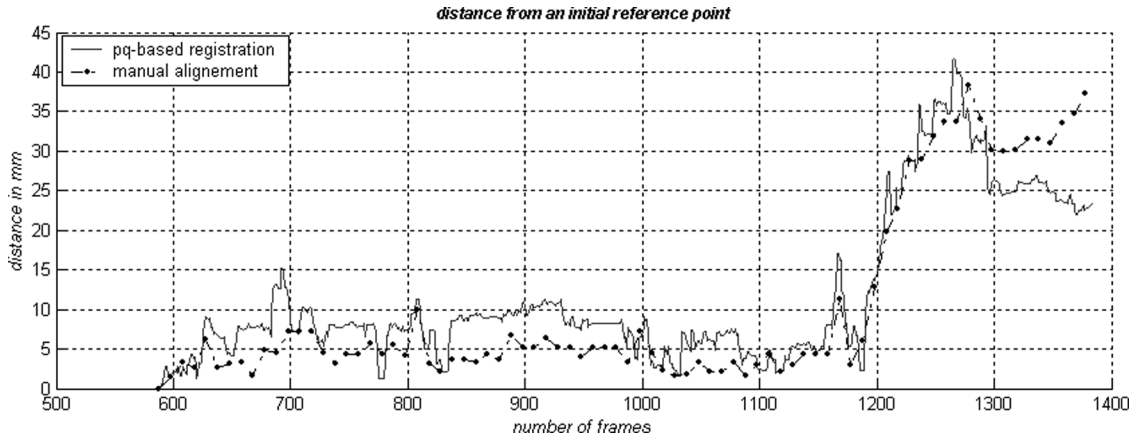

Figure 8. The accuracy of the registration result represented as the Euclidean distance between the first and subsequent camera positions as measured by the *pq*-based 2D/3D registration compared and after manual alignment.

bronchoscope tracking can be achieved. The main advantage of the method is that it is unaffected by illumination conditions. The method depends on the weighting factor that has been introduced into the similarity measure, which has been shown to be resilient to variations in 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 different anatomy, which affects the inter-reflectance lighting and subsequently results in changes in overall illumination conditions. However, the *pq*-space-based technique 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 that deformation is limited to a small portion of the video frame. Further investigation of the accuracy of this method for large tissue deformation is required.

It should also be noted that a number of factors 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 may 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 adversely effect the accuracy of *pq*-space estimation, despite the use of distortion correction. With the proposed method, the intrinsic robustness of the technique is inherently

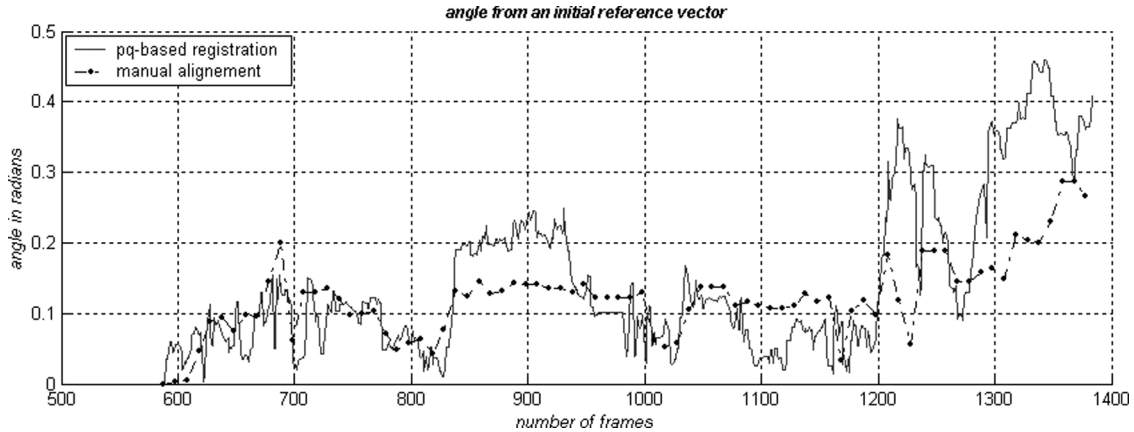

Figure 9. Inter-frame angular difference as measured by the proposed *pq*-based 2D/3D registration method and after manual alignment.

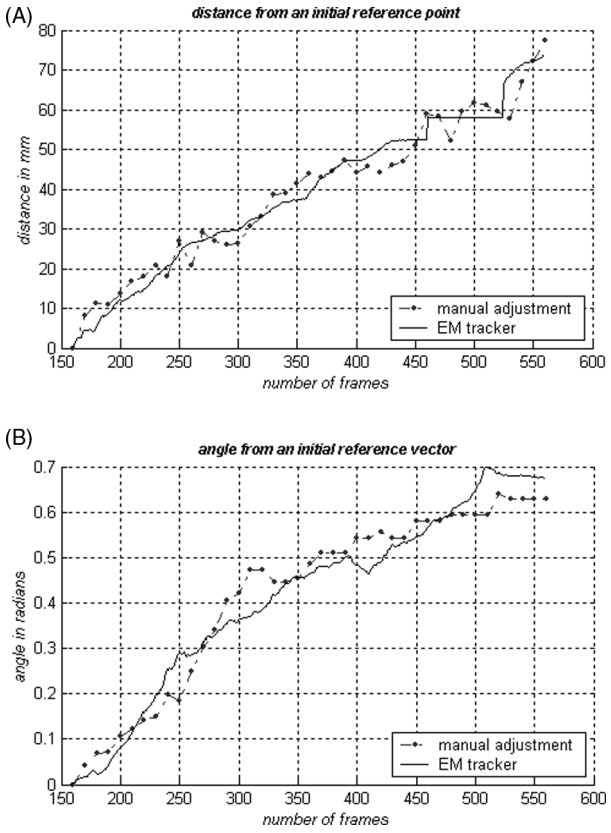

Figure 10. Assessment of manual alignment error compared to the EM tracker using the 3D bronchial model. A) Euclidean distance between the first and subsequent camera positions as measured by the EM tracker and after manual alignment with step = 10. B) Inter-frame angular difference between manual alignment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

dependent on the performance of the shape-from-shading method used. The use of camera/lighting constraints for 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 modifications can therefore be introduced to further improve the accuracy of the proposed framework by explicit incorporation of the effects 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 of 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* miniaturized 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.

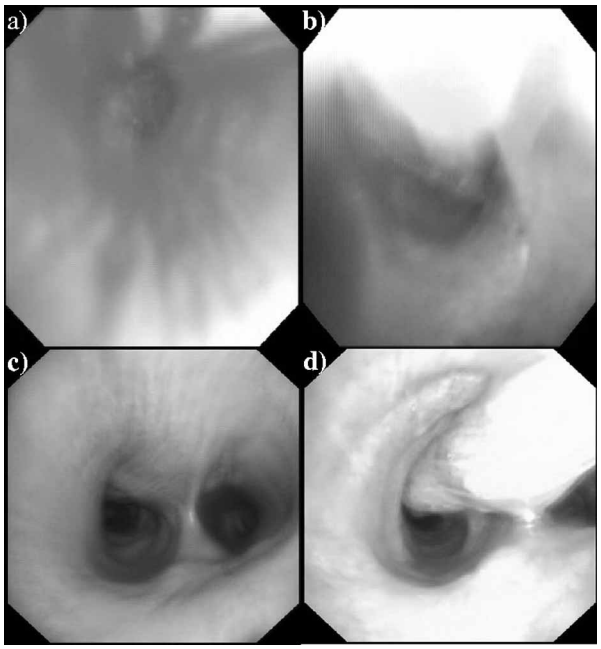

Figure 11. Common image artefacts that can affect image-based 2D/3D registration techniques: a) excessive bleeding due to pathology; b) appearance of bubbles when the patient coughs; and c-d) large tissue deformation between successive image frames.

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.

References

1. Helferty JP, Higgins WE. Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance. In: Proceedings of IEEE International Conference on Image Processing 2002, Rochester, NY, September 2002, Volume II. p 961–4.
2. Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki JI, Maurer CR Jr, Takabatake H, Natori H. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Med Image Anal 2002;6(3):321–36.
3. Mori K, Suenaga Y, Toriwaki JI, Hasegawa JI, Kataa K, Takabatake H, Natori H. A method for tracking camera motion of real endoscope by using virtual endoscopy system. Proceedings of SPIE Medical Imaging, February 2000, San Diego, CA. p 122–33.
4. Chung AJ, Deligianni F, Shah P, Wells A, Yang G-Z. Enhancement of visual realism with BRDF for patient specific bronchoscopy simulation. In: Barillot C, Haynor DR, Hellier P, editors. Proceedings of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004), Saint-Malo, France, September 2004, Part II. Lecture Notes in Computer Science 3217. Berlin: Springer, 2004. p 486–93.
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, Wells MW III. Alignment by maximization of mutual information. Int J Comput Vision 1997;24(2): 137–54.
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):114–41.
9. David P, DeMenthon D, Duraiswami R, Samet H. SoftPOSIT: Simultaneous pose and correspondence determination. Presentation at Seventh European Conference on Computer Vision (ECCV 2002), Copenhagen, Denmark, May–June 2002.
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): 1019–31.
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. Report CAR-TR-969, CS-TR-4257. May 2001.
12. Horn BKP. Understanding image intensities. Artificial Intell 1977;8(2):201–31.
13. Horn BKP. Robot Vision. Cambridge, MA: MIT Press, 1986.
14. Bichsel M, Pentland AP. A simple algorithm for shape from shading. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '92), Champaign, IL, June 1992. p 459–65.
15. Oliensis J. Shape from shading as a partially well-constrained problem. Computer Vision, Graphics, and Image Processing: Image Understanding 1991;54(2):163–83.
16. Oliensis J, Dupuis P. A global algorithm for shape from shading. In: Proceedings of IEEE International Conference on Computer Vision (ICCV '93), Berlin, Germany, May 1993. p 692–701.
17. Tsai P-S, Shah M. A fast linear shape from shading. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '92), Champaign, IL, June 1992. p 734–6.
18. Worthington PL, Hancock ER. New constraints on data-closeness and needle map consistency for shape-from-shading. IEEE Trans Pattern Anal Machine Intell 1999; 21(12):1250–67.
19. Prados E, Faugeras O. A rigorous and realistic Shape From Shading method and some of its applications. INRIA, Odyssie Lab, Sophia Antipolis, France. Report RR-5133, March 2004.
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):119–27.
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):119–31.
22. Heikkila J, Silven O. A four-step camera calibration procedure with implicit image correction. In: Proceedings of Computer Vision and Pattern Recognition (CVPR '97), San Juan, Puerto Rico, June 1997. p 1106–12.
23. Zhang C, Helferty JP, McLennan G, Higgins WE. Nonlinear distortion correction in endoscopic video images. In: Proceedings of IEEE International Conference on Image Processing, Vancouver, BC, September 2000. p 439–42.
24. Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of International Conference on Computer Vision (ICCV '99), Corfu, Greece, September 1999. p 666–73.
25. Yang GZ, Burger P, Firmin DN, Underwood SR. Structure adaptive anisotropic image filtering. Image and Vision Computing 1994;14(2):135–45.