**Noname manuscript No.**
(will be inserted by the editor)

# Robust Camera Localisation with Depth Reconstruction for Bronchoscopic Navigation

**Mali Shen · Stamatia Giannarou · Guang-Zhong Yang**

Received: date / Accepted: date

**Abstract**

*Purpose* Bronchoscopy is a standard technique for airway examination, providing a minimally invasive approach for both diagnosis and treatment of pulmonary diseases. To target lesions identified pre-operatively, it is necessary to register the location of the bronchoscope to the CT bronchial model during the examination. Existing vision-based techniques rely on the registration between virtually rendered endobronchial images and videos based on image intensity or surface geometry. However, intensity-based approaches are sensitive to illumination artefacts while gradient-based approaches are vulnerable to surface texture.

*Methods* In this paper, depth information is employed in a novel way to achieve continuous and robust camera localisation. Surface shading has been used to recover depth from endobronchial images. The pose of the bronchoscopic camera is estimated by maximising the similarity between the depth recovered from a video image and that captured from a virtual camera projection of the CT model. The normalised cross-correlation and mutual information have both been used and compared for the similarity measure.

*Results* The proposed depth-based tracking approach has been validated on both phantom and in vivo data. It outperforms the existing vision-based registration methods resulting in smaller pose estimation error of the bronchoscopic camera. It is shown that the proposed approach is more robust to illumination artefacts and surface texture and less sensitive to camera pose initialisation.

Mali Shen
The Hamlyn Centre for Robotic Surgery, Imperial College London
E-mail: mali.shen09@imperial.ac.uk

Stamatia Giannarou
The Hamlyn Centre for Robotic Surgery, Imperial College London
E-mail: stamatia.giannarou@imperial.ac.uk

Guang-Zhong Yang
The Hamlyn Centre for Robotic Surgery, Imperial College London
E-mail: g.z.yang@imperial.ac.uk

*Conclusions* A reliable camera localisation technique has been proposed based on depth information for bronchoscopic navigation. Qualitative and quantitative performance evaluations show the clinical value of the proposed framework.

**Keywords** Bronchoscopic navigation · 2D/3D registration · Shape from shading · Depth recovery

## 1 Introduction

With increasing maturity of bronchoscopy, it has become a standard care of pulmonary diseases for both diagnosis and treatment. For instance, transbronchial biopsy with image guidance is increasingly used for lung cancer staging and endobronchial valves are placed directly through a bronchoscope to treat emphysema. Emphysema is characterised by the damages to the alveoli, causing hyperinflation of the lung and resulting in less efficient gas exchange and interference with breathing mechanisms. Lung volume reduction surgery (LVRS) has been used to treat emphysema, but the high operative risks and post-operative complications associated with LVRS, as well as the high hospital costs, have motivated the development of bronchoscopic techniques for lung volume reduction [22]. However, video images acquired through a bronchoscope are of degraded resolution, distortion and small field-of-view, making the perception and 3D navigation of the operating site more difficult for the surgeons. Recent development of computer vision techniques gives rise to solutions to enhance the image quality and provide effective visualisation with augmented reality.

Thus far, approaches to bronchoscopic navigation aim at the localisation of the camera in the CT model coordinate system by applying EM tracking, image registration or a combination of both. Miniature EM sensors can be mounted to the tip of the bronchoscope or inserted into the bronchoscopic channel to provide in real-time the orientation and location of the tip. EM sensor calibration can be achieved either by using extrinsic skin fiducial markers placed near the bony structures such as clavicles and sternum [23] or intrinsic anatomical landmarks of the airways such as the carina and the origin of the left and right upper lobe bronchi [1, 23]. Since the EM field can be distorted by the presence of ferromagnetic materials around the EM systems, an optical tracker was incorporated in [16] to compensate for failure of the EM tracking system. However, precise EM sensor calibration is still required before each bronchoscopic session, and the accuracy of the EM guidance system is highly sensitive to respiratory motion and the patient's motion during the examination.

Vision-based tracking techniques, on the other hand, only depend on matching bronchoscopic video images to rendered views of the segmented CT bronchial model [23]. The similarity between virtual endobronchial images and bronchoscopic video images has been measured using either intensity-based or geometry-based approaches.

Intensity-based methods match bronchoscopic images to rendered virtual camera projections by comparing image intensities. In the multilevel registration method proposed by Bricault et al. [3], virtual bronchial views have been rendered using ray tracing techniques. Photo-realistic rendering has also been proposed by Chung et al. [4] for bronchoscopy visualisation. The accuracy of intensity-based methods is sensitive to small local intensity variations and illumination artefacts. In addition, intensity-based methods often require continuous manual lighting adjustment of the virtual camera to achieve satisfactory registration results. To optimise registration, a selective similarity measure has been proposed in [15, 5] to register real and virtual endoscopic images based on subregions with bifurcations or characteristic cartilage rings of bronchi.

Geometry-based methods extract anatomical structures from video images to be matched with 3D structures in the CT airway model. These registration approaches are less sensitive to illumination artefacts because they do not depend on the direct comparison in terms of image intensity or any detection and matching of visual features. A pq-space based 2D/3D registration approach for bronchoscopic tracking was proposed by Deligianni et al. [6,7]. They also incorporated shape modelling of airway deformation for non-rigid 2D/3D registration [8]. The surface normal for each pixel of video images were extracted using the algorithm derived by Rashid et al. [21]. Validation of the pq-space based registration showed that it is more robust to illumination artefacts and tissue deformation than intensity-based methods [6]. However, pq distributions extracted from video images are sensitive to surface texture such as small vessels on the endobronchial surface, while these visual features can not be detected on CT data.

Recovering geometrical structures from video images can be challenging. This is because conventional structure recovery approaches such as Shape from Shading (SFS) assume orthogonal projection and the light source at infinity. However, these assumptions are not applicable in the bronchoscopic scenario. In the multilevel registration strategy developed by Bricault et al. [3], a model-based depth recovery method has been employed at the last level of their registration process. The accuracy of the depth recovery method relies on relatively good pre-registration between the current virtual projection view and the video view from previous registration levels. If the virtual view and the video view are very different, the CT data used for deriving the relationship between video image intensity and surface normal can no longer hold and the surface reconstruction of the endobronchial wall may fail. Furthermore, this registration method using depth recovery has not been validated for continuous tracking.

Bronchoscopic navigation has also been achieved with hybrid registration methods. A tracking system combining EM tracking and intensity-based image registration has been developed by Luo et al. [13]. Another framework using both EM tracking and image registration based on differential surface analysis (pq-space) has also been developed by Soper et al. for improved registration accuracy [24]. To overcome the limitations of EM tracking techniques, an optical mouse sensor combined with the enhanced differential evolution algorithm has been proposed in [14]. These hybrid methods improve registration performance but complicate the overall sensor calibration and registration algorithm.

Since image registration can be time consuming, methods incorporating motion prediction have also been investigated to achieve faster convergence of the registration process. Higgins et al. [9,20] tracked the 3D motion of the bronchoscope by estimating optical flow from video images and then using the tracked 3D trajectory to assist localisation in the 3D CT virtual world. Nagao et al. [17] employed the principle of Kalman filter to predict the motion of the bronchoscope. Since the 3D motion prediction computed from 2D video image sequences may not be accurate enough, the final image registration still relies on the matching between video and virtual endobronchial images.

In this work, we propose a depth-based registration approach for robust camera localisation during bronchoscopy. The camera pose is estimated based on the similarity between depth maps extracted from bronchoscopic images and depth information extracted from pre-operative CT data. To reconstruct depth information from bronchoscopic images, a SFS approach tailored for endoscopic applications is employed based on the assumption that the light source is near the surface without restricting its position to be at the optical centre or at infinity. Depth information from pre-operative CT data is recovered by linearisation of depth buffering from perspective projections of the CT model. To measure the similarity between the bronchoscopic and CT depth maps, two different measures are employed and

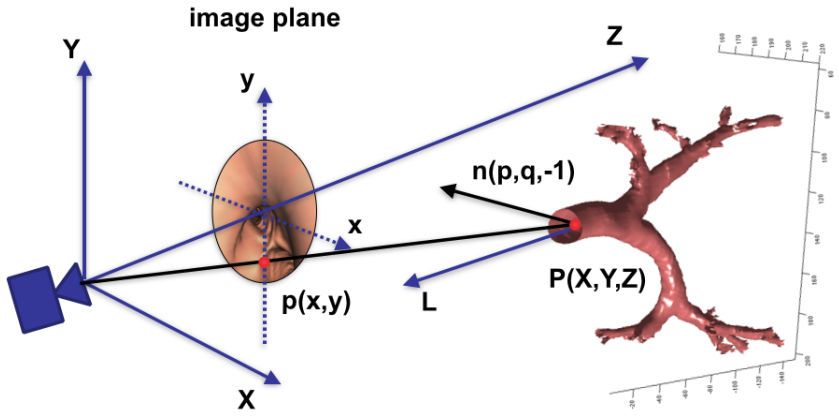



**Fig. 1** SFS in bronchoscopy with Lambertian surface for perspective camera and a light source close to surface and at the optical centre.

compared in this work namely, normalised cross-correlation and normalised mutual information. Detailed validation on phantom and in vivo data and comparison to state-of-the-art 2D/3D registration approaches have shown that the proposed framework has the advantage of being more robust to illumination artefacts and surface texture and less dependent on the initialisation of the camera pose. The results derived justify the potential clinical value of the technique.

## 2 Methods

The proposed camera localisation method for bronchoscopic navigation comprises three main steps: the reconstruction of a depth map from a single bronchoscopic video frame using SFS [25]; the extraction of depth maps from multiple virtual camera projections of the CT model; and the estimation of the camera pose that maximises the similarity between the video depth map and the corresponding virtual depth maps using normalised cross-correlation or normalised mutual information.

### 2.1 SFS Depth Recovery from Video Image

To recover depth information from the bronchoscopic environment, we use SFS as it can provide the 3D structure of a low-textured scene from a single 2D monocular image. The majority of SFS methods proposed in literature [19, 26, 12] are based on the assumption that the light source is at infinity or at the optical centre of the camera. However, this assumption is not realistic for endoscopic applications. In this work, we employ the SFS model proposed by Visentini-Scarzanella et al. [25]. This model has been tailored for endoscopic applications and its advantage compared to other popular SFS methods is that it doesn't restrict the position of the light source to be at infinity or at the optical centre but it allows the light source to be anywhere close to the surface.

For a simple SFS scenario with a Lambertian surface (Fig. 1), the image irradiance $E_i$ can be expressed as

$$E_i = \sigma \frac{cos\theta_i}{r^2} = \sigma \frac{(\mathbf{l} \cdot \mathbf{n})}{r^2} \tag{1}$$

where $\sigma$ is a constant coefficient depending on the camera parameters, the intensity of the light source and the surface albedo, $\theta_i$ is the incident angle of the light, $r$ is the distance between the light source and the surface point, $\mathbf{n}$ is the surface normal vector and $\mathbf{l}$ is the incident light direction vector [19, 10]. The image intensity $I$ is often assumed to be equal to the image irradiance $E_i$. For each pixel $\mathbf{p} = (x, y)$ in the image domain, the surface normal $\mathbf{n}$ and incident light vector $\mathbf{l}$ for its corresponding 3D surface point $\mathbf{P}$ are expressed as:

$$\begin{gathered} \mathbf{n} = \left( Z_x, Z_y, -\frac{(x+T)Z_x + (y+U)Z_y + Z(\mathbf{p})}{(f+V)} \right) \\ \mathbf{l} = (x+T, y+U, f+V) \end{gathered} \tag{2}$$

where $Z(\mathbf{p})$ is the depth at point $\mathbf{p}$ and $Z_x$ and $Z_y$ are the spatial derivatives with respect to $x$ and $y$. $\mathbf{L} = [T, U, V]$ is the position of the light source in the 3D space. In order to estimate the surface structure, the depth parameter $Z$ should be estimated at every image pixel. Using the assumption of a Lambertian surface (Fig. 1) with a point light source at the optical centre ($T = U = V = 0$), $Z$ can be recovered by solving the Hamiltonian obtained by substituting $v = \ln Z$ into Eq.(1):

$$H(\mathbf{p}, \nabla v) = I(\mathbf{p}) \frac{1}{\sigma} \sqrt{\left(v_x^2 + v_y^2\right) + J(\mathbf{p}, \nabla v)^2 \cdot Q(\mathbf{p})^{3/2}} \tag{3}$$

where:

$$\begin{cases} J(\mathbf{p}, \nabla v) = \frac{v_x x + v_y y + 1}{f} \\ Q(\mathbf{p}) = x^2 + y^2 + f^2 \end{cases} \tag{4}$$

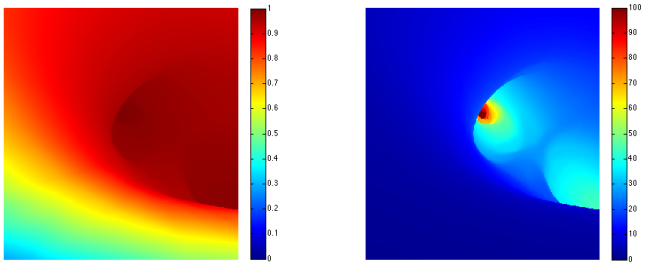
The 2D Lax-Friedrichs sweeping scheme proposed in [11] was applied to solve the above equation iteratively.

## 2.2 Depth Map from Virtual CT Projection

In order to recover depth information from the pre-operative data, perspective projections of the CT surface mesh are estimated with multiple camera poses to generate a set of virtual endobronchial views. The pose of the virtual camera is modelled as the set of rotation $R(\alpha, \beta, \gamma)$ and translation $t(x_t, y_t, z_t)$ parameters that transform the camera from the origin to the current viewing position and orientation within the world coordinate system by applying the total transformation $T$, defined as:

$$T = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & x_t \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & y_t \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma & z_t \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Each parameter combination $A(x_t, y_t, z_t, \alpha, \beta, \gamma)$ produces a unique virtual endobronchial image.



(a) (b)

**Fig. 2** Recovery of the true depth map of a CT perspective projection from depth values in Normalised Device Coordinates (NDC). (a). Sampled and normalised depth map from shader; (b). True depth map after linearisation with near and far clipping planes at Z=1 and Z=100 respectively.

The perspective projections of the CT endobronchial wall are rendered using the rasterisation technique. The intrinsic parameters of the virtual camera were set to the same values as the parameters of the bronchoscopic camera. The generated virtual images are of the same size as the bronchoscopic video images. Sampling from a depth buffer in a shader gives the associated depth value $Z_{NDC}$ in a normalised coordinate system (Normalised Device Coordinates (NDC)) for each pixel of the virtual image. The normalised depth of every pixel can be linearised to give the actual depth $Z$:

$$Z = \frac{\frac{-2fn}{f-n}}{Z_{NDC} - \left(\frac{f+n}{f-n}\right)} = \frac{-2fn}{Z_{NDC}(f-n) - (f+n)} \tag{6}$$

where $Z_{NDC}$ ranges from -1 to 1, $f$ and $n$ are the Z-coordinates of the near and far clipping planes respectively. An example of normalised and converted depth values of a virtual endobronchial image are illustrated in Fig. 2(a) and Fig. 2(b) respectively.

### 2.3 Camera Pose Estimation and Similarity measurement

To this end, depth information has been reconstructed from a bronchoscopic video frame and from a set of virtual endobronchial images generated from the CT data for multiple camera poses. The aim of the proposed camera localisation framework is to estimate the camera pose parameters that generate the virtual image with the highest similarity to the video image. The camera pose is determined by optimising the similarity between the depth maps of the video and the virtual images. For this purpose, the normalised cross-correlation and the normalised mutual information have been applied as the similarity measures.

Since the depth map extracted from the video image using SFS is not metric, it is rescaled to the same clipping frustum used for the virtual endobronchial projections as:

$$Z_{Rescaled} = n + \frac{(Z - min(Z))(f-n)}{max(Z) - min(Z)} \tag{7}$$

where $Z$ is the depth value recovered from the bronchoscopic video image and $f$ and $n$ are the Z-coordinates of the near and the far clipping planes respectively.

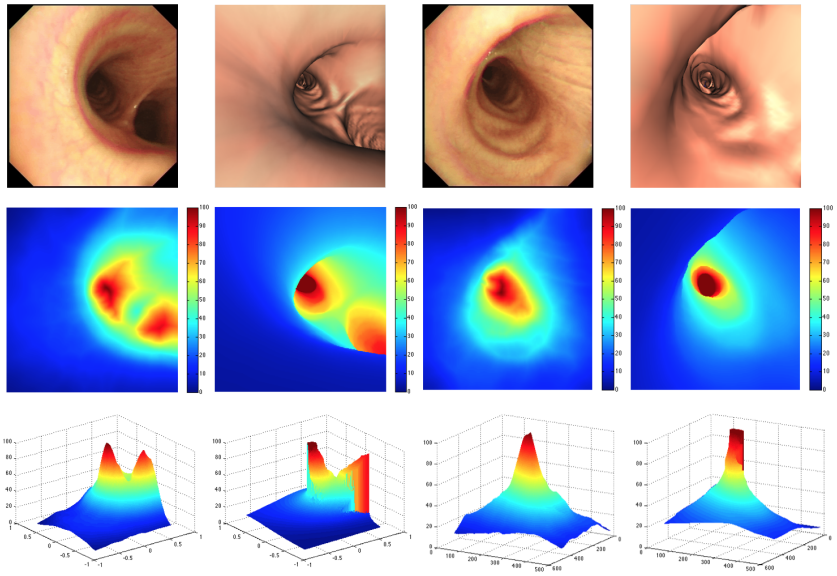


**Fig. 3** Depth maps of two pairs of matched video image and projection image of CT model, one at the carina (Columns 1& 2) and one inside the right main bronchus (Columns 3 & 4).

The precision of the depth buffer in OpenGL varies at different depth location between the near and far clipping planes. The clipping planes were adjusted to achieve optimised precision of the depth buffer for good alignment between the depth maps from the video and the virtual images.

Depth maps generated from in vivo data using the SFS depth reconstruction are illustrated in Fig. 3 for two different airway locations, one at the carina of the trachea and one inside the right main bronchus. The generated virtual endobronchial images and their corresponding depth maps for these two video images are also shown for qualitative comparison.

Both the normalised cross-correlation and the normalised mutual information were applied to measure the similarity $S$ between the depth maps. Given the registered camera pose $A_i$ for video frame $i$, the camera pose for the next frame $A_{i+1}$ is estimated by minimising the inverse of the similarity between the video depth map $Z_{i+1}$ and the generated virtual depth map $Z'(A_{i+1})$ starting with $A_i$ as the initial estimate.

$$\begin{aligned} A_{i+1} &= A_i + \Delta A_{i+1|i} \\ \Delta A_{i+1|i} &= arg\min_{\Delta A_{i+1|i}} \left\{ S^{-1}\left(Z_{i+1}, Z'\left(A_i + \Delta A_{i+1|i}\right)\right)\right\} \end{aligned} \tag{8}$$

The optimal change of the camera pose $\Delta A_{i+1|i}$ is estimated by searching for the rotation and translation updates that provide the minimal dissimilarity which is here estimated as the inverse of the similarity $S_{i+1}$. In this work, the optimisation process is based on Powell's numerical minimisation algorithm [18] which is suitable for a multi-variable function without an underlying mathematical definition as it does not require the function to be differentiable. The multiple variables in our framework correspond to the six parameters that represent the camera pose $A$. The linear search for the optimal parameters is achieved by applying the Brent's method [2] where the maximum variation of the pose was set to $5mm$ for the translation and $5^\circ$ for the rotation angles for fast convergence.

## 3 Results

The proposed depth-based registration framework was implemented using C++ with OpenGL for z-buffering and endobronchial surface rendering, and FLTK for the interface. To facilitate performance comparison, the registration algorithm was tested on the same phantom and in vivo data that has been used for validation in [7]. Both quantitative and qualitative performance evaluations were carried out and the proposed camera localisation framework was compared to the current 2D/3D registration methods.

### 3.1 Phantom validation

An airway phantom made of silicon rubber and coated with acrylics to mimic surface texture of the airway was used for validation. The pre-operative data was gathered by scanning the phantom with a Siemens Somaton Volume Zoom 4-channel multi-detector CT scanner. Video data from the inside of the phantom was recorded with a NTSC standard CMOS camera. An EM tracking system with 6 DOF was used to estimate the ground truth camera pose for every video image [6]. The proposed framework was validated on 400 frames. The camera location in the CT coordinate system was continuously estimated by registering the video frames to the CT bronchial model. The registration accuracy was estimated in terms of the Euclidean distance and angular error between the estimated camera pose and the ground truth pose given by the EM data. The performance of the proposed depth-based method is compared to the conventional intensity-based and pq-space based methods in Fig. 4. The pq components and depth maps extracted from a pair of matched video image and CT virtual image of the bronchi phantom are illustrated in Fig. 5.

Validation conducted on the static bronchi phantom (Fig.4(a)) shows that the proposed depth-based method provides lower distance error compared to the intensity-based and the pq-space based approaches for camera localisation over a continuous series of video frames. The relatively high angular errors shown in Fig.4(b) are due to the lack of strong anatomical features of the phantom that could provide information to solve the orientation ambiguity in the bronchial tubular structure.

By further comparing the accuracy estimated using the normalised cross-correlation and the normalised mutual information for the similarity measure, the normalised cross-correlation provides more robust camera pose estimation. This is expected as the normalised cross-correlation measures the linear or monotonic relationship between two distributions while the mutual information measures non-monotonic relationships. Since we require exact matching between the depth maps from the video and the CT model, the cross-correlation measure gives better registration performance.

### 3.2 In vivo validation

A video sequence of a standard bronchoscopy examination was collected with an Olympus prototype bronchoscope of 120° field of view. The pixel resolution of the video images is 454x487. Pre-operative CT data was generated in the same way as for the phantom case. Manually registered camera poses were used as ground truth for validation. The potential bias of the manually registered data was controlled by conducting an error assessment of the manually registered pose by comparing to the EM tracker data using phantom measurements. Both average distance and angular errors of the manual alignment were consistent

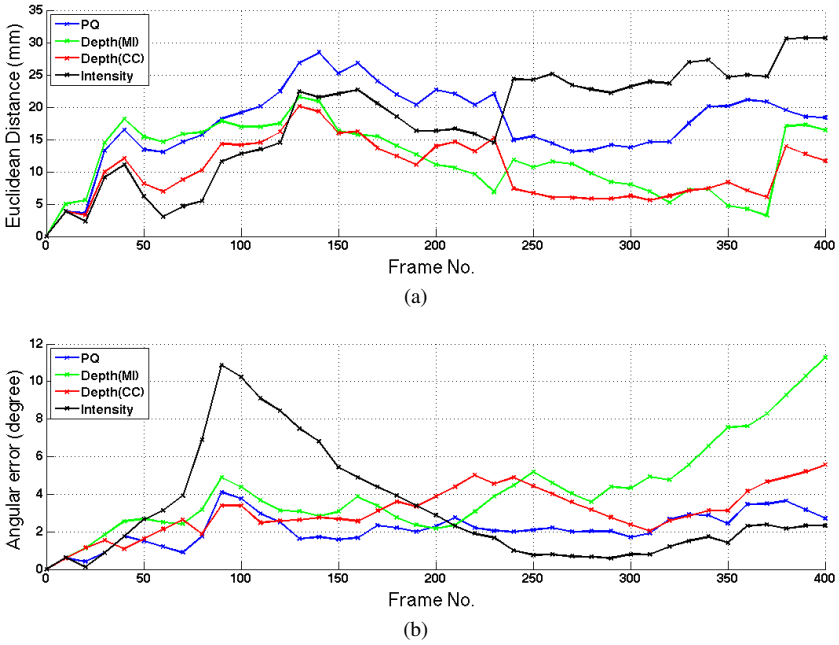



**Fig. 4** Registration accuracy of the proposed depth-based approach on the phantom data. (a). Distance error of the estimated camera location to the ground truth location over time. (b). Angular error of the estimated camera orientation to the ground truth orientation over time. Intensity: intensity-based registration with similarity measure based on normalised cross-correlation; pq-space: pq-space registration approach with similarity measure based on angle between pq vectors; Depth(MI): depth-based registration approach with similarity measure based on normalised mutual information; Depth(CC): depth-based registration approach with similarity measure based on normalised cross-correlation.

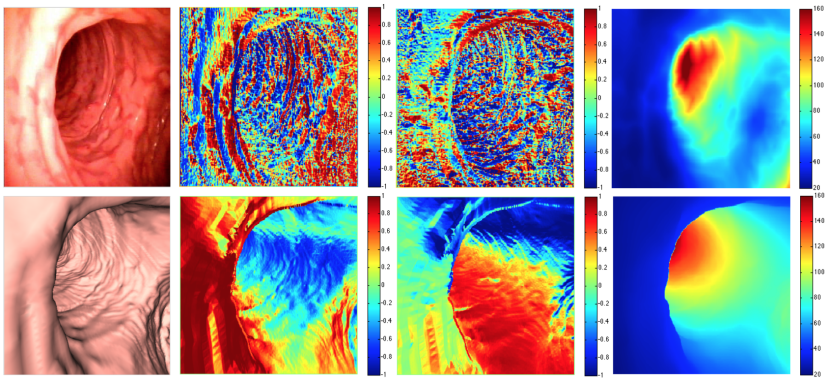


**Fig. 5** Distribution map of p component (Column 2), q component (Column 3) and depth map (Column 4) of a pair of matched video (Row 1) and CT virtual camera projection (Row 2) of the static bronchi phantom.

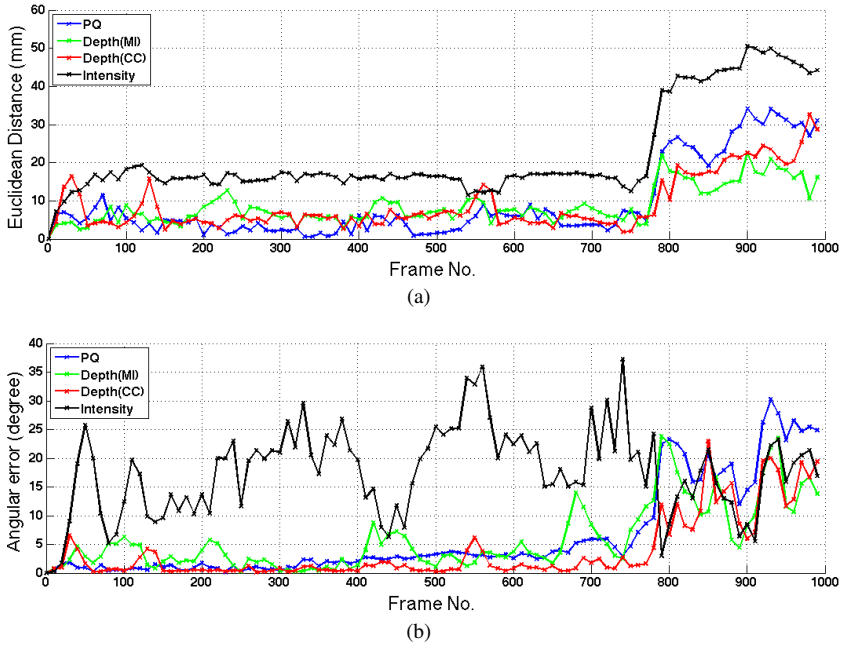



**Fig. 6** Registration accuracy of the proposed depth-based approach on the in vivo data. (a). Distance error of the estimated camera location to the ground truth location over time. (b). Angular error of the estimated camera orientation to the ground truth orientation over time.

**Table 1** Accuracy of estimated camera pose in terms of Euclidean distance error $\delta$ $(mm)$ and angular distance error $\varphi$ $(degree)$ by four compared registration methods for the in vivo validation.

| Registration method | $\mu(\delta)$ | $\sigma(\delta)$ | $\mu(\varphi)$ | $\sigma(\varphi)$ |
|---|---|---|---|---|
| Intensity | 21.294 | 12.002 | 17.576 | 7.6056 |
| pq-space | 8.7117 | 9.6660 | 5.9886 | 7.8219 |
| Depth(MI) | 8.5195 | **4.7557** | 5.6986 | 5.5984 |
| Depth(CC) | **8.4800** | 6.2981 | **3.4775** | **5.2758** |

and relatively small for the phantom that has a larger scale than real human airways, suggesting the error of the manual alignment of the in vivo data would be smaller and with reliable accuracy [7]. The radial and tangential distortions of the collected bronchoscopic data were corrected prior to processing [7]. The bronchoscopic pose was continuously estimated by registering the video frames with the CT bronchial model over 1000 video frames. The Euclidean distance and angular error for the four compared approaches are shown in Fig. 6. Sample frames of the bronchoscopic video and the corresponding virtual images registered with the compared methods are shown in Fig. 7. The mean and standard deviation of the distance error $\delta$ and the angular error $\varphi$ between the estimated camera pose and the ground truth when applying the compared methods on the in vivo data are provided in Table 1.

*Intensity-based registration* has the largest distance error and angular error for the in vivo validation as presented in Table 1. This is because variations in the lighting conditions in the real-time bronchoscopic video lead to relatively poor and unstable tracking performance for

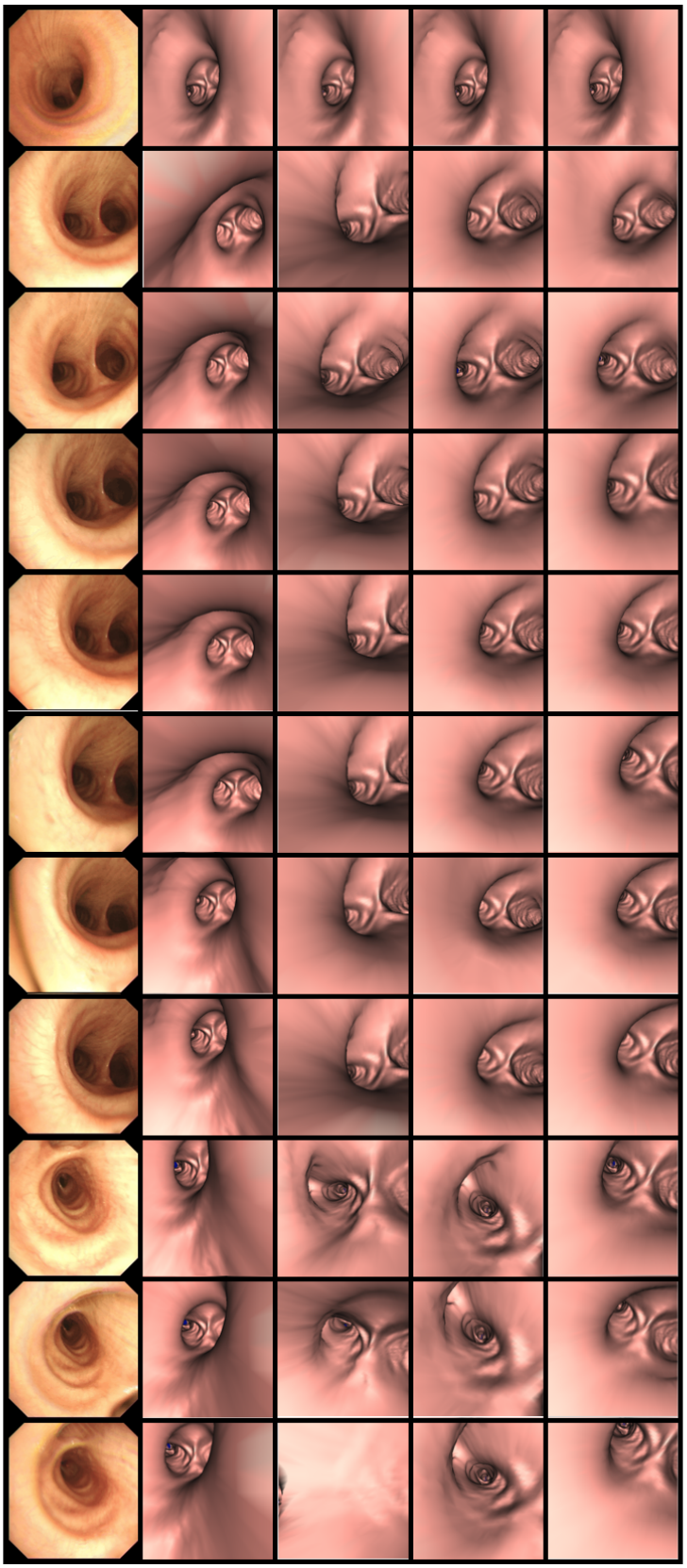


**Fig. 7** Sample frames of the bronchoscopic video (Column 1) and the corresponding virtual images registered with the four different approaches: Intensity-based (Column 2), pq-space based (Column 3), depth-based(CC)(Column 4) and depth-based(MI)(Column 5).

intensity-based registration approaches (Fig. 6). This is also verified qualitatively in Fig. 7 (Column 2), where the registered virtual endobronchial images estimated by the intensity-based approach do not match the corresponding video images.

*pq-space based registration* performs better than the intensity-based method, because it relies on surface gradient instead of intensity. Distribution maps of the p component, q component, and the flow-to-colour of the pq-space extracted from two bronchoscopic video images, one at the carina and one inside the right main bronchus are illustrated in Fig. 8. The distribution maps of the pq-space from the corresponding CT virtual camera projections were also computed for these two airway locations. The pq components calculated from the video images reveal both local surface texture and the global geometric structure of the endobronchial wall. This local texture information can not be recovered via rendering of the virtual bronchial model. As a result, the pq components extracted from the CT virtual projection only encode gradient information of the global endobronchial structure. Smoothing and anisotropic filtering have been applied to remove texture features from the video images for better matching with the CT endobronchial projections [6, 24]. However, these filters degrade the surface gradient information extracted from the video image, as well as the similarity between the pq-space derived from the video and the CT model.

*Depth-based registration* gives similar registration accuracy compared to the pq-space based method when the bronchoscope is approaching the carina of the trachea (Frame 0-800) in Fig. 6. It also shows that the curves of the depth-based approach have peaks around Frame 30, 130, 550 when there were instrumental occlusions or water bubbles blocking the bronchoscopic view. The registration accuracy plot between Frame 100-200 is shown in Fig. 9. The depth recovery becomes inaccurate when there are image artefacts, such as bubbles and blurring. However, when these artefacts disappeared, fast realignment between the video data and the virtual endobronchial images can be achieved with the depth-based method, which makes it less sensitive to registration initialisation.

Furthermore, the proposed depth-based approach provides more accurate camera pose estimation than the pq-space based approach when the bronchoscope goes into the right main bronchus from the carina (Frame 800-1000) (Fig. 6). Straight sections of the airways (i.e. trachea and main bronchi) are tubular structures with limited appearance and structure characteristics. Due to the paucity of salient information in these airway sections, the performance of the proposed method in these areas is lower than that at bronchial bifurcations. However, as shown in Fig. 6(a), this performance degradation is common to all of the compared methods after the bronchoscope enters the right main bronchus (Frame 800-1000) from the carina of the trachea. The distance error of the proposed method even in the straight airway is around 15mm which is still smaller than those of the compared intensity-based (45mm) and pq-space based (27mm) method.

Similarly to the results given in the phantom validation, the depth-based method with normalised cross-correlation gives more accurate and stable estimation of the camera pose parameters than the compared registration approaches (Table 1). Its relatively higher standard deviation of distance error is mainly caused by the mis-registrations for a few video frames when there were occlusions and blurring. Registered virtual endobronchial images given by the depth-based method using normalised cross-correlation displayed in Fig. 7 (Column 3) give the best match to the corresponding video frames among the other methods. As shown in Fig. 3, the depth recovery approach is less sensitive to small localised intensity features or local geometric structures. Therefore, the depth map generation gives

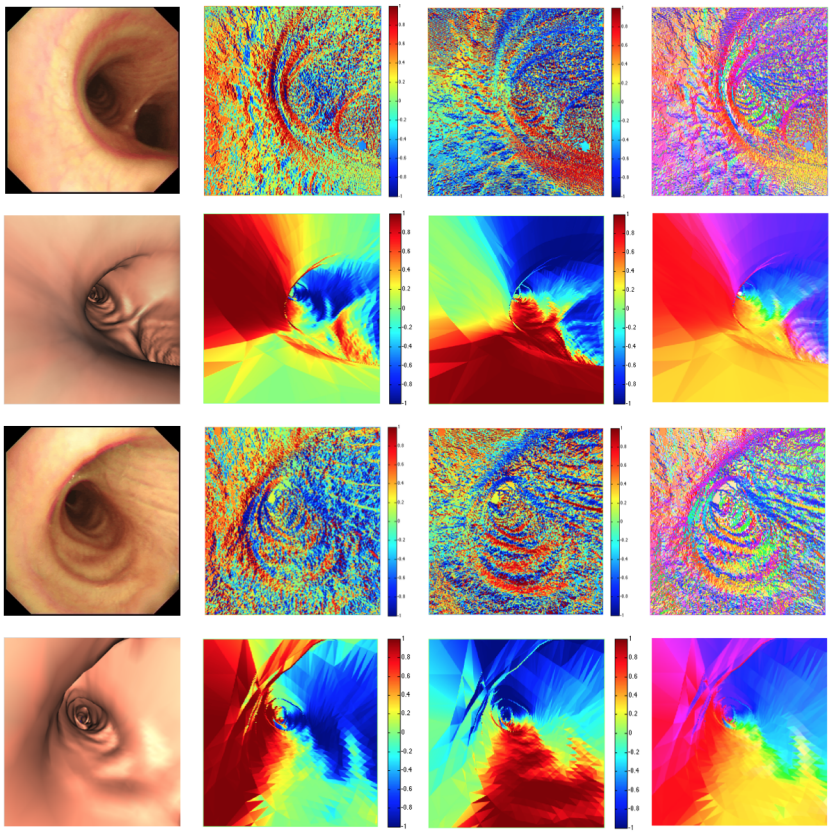


**Fig. 8** Distribution map of p component (Column 2), q component (Column 3) and flow-to-colour map (Column 4) of pq-space of two pairs of matched bronchoscopic video and CT virtual camera projections, one at the carina of the trachea (Row 1 & 2) and one inside the right main bronchus (Row 3 & 4).

more reliable estimation of the global surface structure of the endobronchial wall than the pq-space measurement.

### 3.3 Computational complexity

The localisation approaches have been tested on a PC with 3 GHz Intel Core i7 Processor, 8 GB 1600 MHz DDR3 Memory. The computational complexity and cost of the proposed depth-based camera localisation approach and the compared approaches are illustrated in Table 2. The computational cost of the intensity-based approach is mainly due to the computation of the normalised cross-correlation between the video image and the CT virtual images during the optimisation process. The complexity of calculating the normalised cross-correlation for a pair of images is considered as $O(N)$ where N is the number of pixels in the 2D video image. For the pq-space based camera localisation approach, the complexity of computing the pq components of a 2D video image is $O(N)$. In the optimisation process, computation of the pq components of the CT virtual image given by a camera pose is required before measuring the similarity between the pq components of the CT virtual image

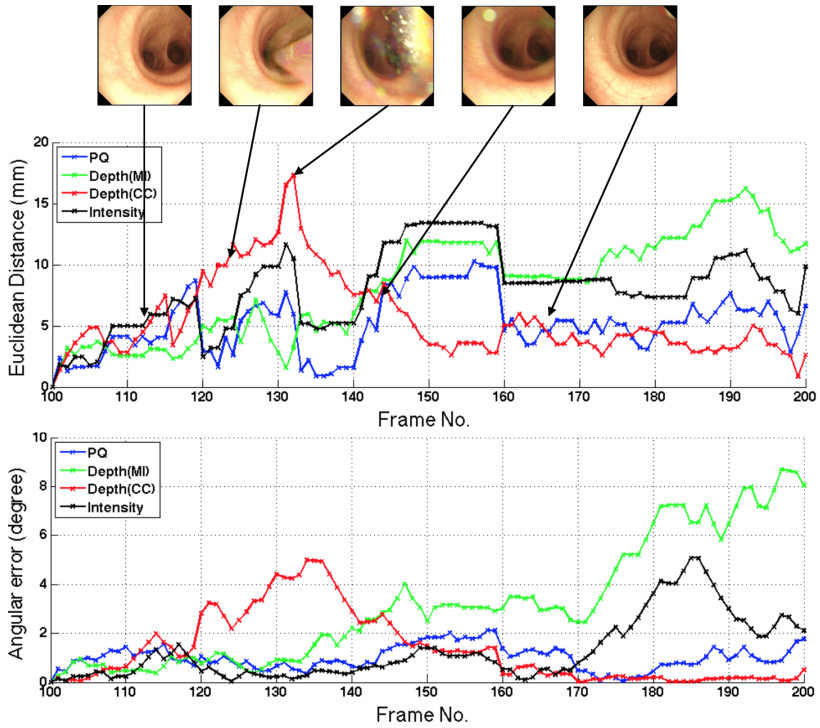



**Fig. 9** Registration accuracy between video Frame 100-200. Sample bronchoscopic video frames are displayed to indicate when the occlusions happened and disappeared.

**Table 2** Computational complexity and run-time cost of the four compared registration methods in terms of the mean processing time for registering one frame to the CT model, where N is the number of image pixels in the video image.

| Registration method | Video | CT | Similarity | Time cost (secs/frame) |
|---|---|---|---|---|
| Intensity | - | - | $O(N)$ | 0.95 |
| pq-space | $O(N)$ | $O(N)$ | $O(N)$ | 60.47 |
| Depth(MI) | $O(N)$ | - | $O(N)$ | 53.00 |
| Depth(CC) | $O(N)$ | - | $O(N)$ | 53.30 |

and the video image. Both the complexity of computing the pq values of the virtual image and that of the similarity measure are $O(N)$. For the proposed depth-based camera localisation approach, the complexity of one iteration of the depth recovery process from a video image is $O(N)$. It was found that in the in vivo validation, 150 iterations were sufficient for the complete recovery of the depth map. Since the complexity of computing the normalised mutual information is also proportional to the number of pixels $N$ in the video images, the complexity of calculating the similarity between a CT image and the video image during the optimisation process is $O(N)$. Powell's minimisation method was used for the optimisation procedure of all the compared methods, while the number of similarity measures required to find the optimal pose for each video image varies for different methods.

Intensity-based approaches are less computationally expensive than geometry-based approaches but at the expense of being more vulnerable to illumination and texture variations.

The computational burden of geometry-based approaches is mainly due to the additional step required for the estimation of the geometrical information from the video data. As shown in Table 2, the proposed depth-based approach is more efficient than the pq-space based approach. The computational cost of the proposed depth-based approach is mainly dependent on the number of iterations required to extract the depth map from the video image. The longer processing time of the pq-space based approach is due to the additional cost in estimating pq values from the depth maps of the CT data that correspond to all the virtual camera poses that are evaluated during the optimisation process.

## 4 Discussion and Conclusion

In conclusion, a novel approach has been proposed for camera localisation during bronchoscopy based on depth recovery. The proposed registration approach provides more accurate and robust camera pose estimation compared to conventional intensity-based or pq-space based registration methods. The intensity-based registration requires continuous adjustment of illumination conditions used in CT model rendering to maintain the camera position in the correct path. The pq-space based approach shows poor performance when there are distinctive surface texture on the endobronchial surface. The proposed method overcomes the limitations of the above methods and it is more robust to illumination artefacts and less sensitive to initialisation of the camera pose.

Although it is shown in Table 2 that the computational complexity of the proposed camera localisation approach is only linearly dependent on the resolution of video images, its efficiency also relies on the implementation of the method, the number of iterations required for depth reconstruction from a video image and the speed of convergence of the optimisation procedure. To improve its efficiency and enable its clinical use, parallel programming and GPU implementation can be used to optimise the implementation of the depth recovery algorithm and the optimisation method.

In our future work, tissue deformation will be considered for more accurate registration between a dynamic pre-operative CT model and the intra-operative video images. Better depth buffering technique will be investigated to improve the precision of the depth map recovery from CT models. Finally, combination of the vision-based registration with motion prediction will be used for faster convergence of the registration process.

**Acknowledgements** The authors would like to thank Dr Pallav Shah for providing us with the opportunity to observe bronchoscopic procedures at Royal Brompton Hospital and providing CT scans and bronchoscopic videos. A special thank to Dr. Fani Deligianni who provided details of the pq-space based registration approach originally proposed in [7] and the data used for validation in that publication.

**Compliance with Ethical Standards**
*Disclosure of potential conflicts of interest:* The authors declare that they have no conflict of interest.
*Research involving Human Participants and/or Animals:* For this type of study formal consent is not required. This article does not contain any studies with human participants or animals performed by any of the authors.
*Informed consent:* No Informed consent was obtained.

## References

1. Becker, H.D., Herth, F., Ernst, A., Schwarz, Y.: Bronchoscopic biopsy of peripheral lung lesions under electromagnetic guidance: a pilot study. Journal of Bronchology & Interventional Pulmonology **12**(1), 9–13 (2005)

2. Brent, R.: Algorithms for Minimization Without Derivatives. Dover Books on Mathematics. Dover Publications (1973)
3. Bricault, I., Ferretti, G., Cinquin, P.: Registration of real and ct-derived virtual bronchoscopic images to assist transbronchial biopsy. IEEE Transactions on Biomedical Engineering **17**(5), 703–714 (1998)
4. Chung, A.J., Deligianni, F., Shah, P., Wells, A., Yang, G.Z.: Patient-specific bronchoscopy visualization through brdf estimation and disocclusion correction. IEEE Transactions on Biomedical Engineering **25**(4), 503–513 (2006)
5. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. Medical Image Analysis **13**(4), 621–633 (2009)
6. Deligianni, F., Chung, A., Yang, G.Z.: pq-space based 2d/3d registration for endoscope tracking. In: Medical Image Computing and Computer-Assisted Intervention, pp. 311–318 (2003)
7. Deligianni, F., Chung, A., Yang, G.Z.: Patient-specific bronchoscope simulation with pq-space-based 2d/3d registration. Computer Aided Surgery **9**(5), 215–226 (2004)
8. Deligianni, F., Chung, A.J., Yang, G.Z.: Nonrigid 2-d/3-d registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. IEEE Transactions on Biomedical Engineering **25**(11), 1462–1471 (2006)
9. Higgins, W.E., Helferty, J.P., Padfield, D.R.: Integrated bronchoscopic video tracking and 3d ct registration for virtual bronchoscopy. pp. 80–89 (2003)
10. Horn, B.: Robot vision. MIT press (1986)
11. Kao, C.Y., Osher, S., Tsai, Y.H.: Fast sweeping methods for static hamilton–jacobi equations. SIAM journal on numerical analysis **42**(6), 2612–2632 (2005)
12. Lei, Y., Jiu-Qiang, H.: A perspective shape-from-shading method using fast sweeping numerical scheme. Optica Applicata **38**(2), 387–398 (2008)
13. Luo, X., Feuerstein, M., Sugiura, T., Kitasaka, T., Imaizumi, K., Hasegawa, Y., Mori, K.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. p. 76251B (2010)
14. Luo, X., Jayarathne, U., McLeod, A., Mori, K.: Enhanced differential evolution to combine optical mouse sensor with image structural patches for robust endoscopic navigation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 340–348 (2014)
15. Mori, K., Enjoji, T., Deguchi, D., Kitasaka, T., Suenaga, Y., Toriwaki, J., Takabatake, H., Natori, H.: New image similarity measures for bronchoscope tracking based on image registration between virtual and real bronchoscopic images. pp. 165–176 (2004)
16. Mori, K., Ishitani, K., Deguchi, D., Kitasaka, T., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Compensation of electromagnetic tracking system using an optical tracker and its application to bronchoscopy navigation system. p. 65090M (2007)
17. Nagao, J., Mori, K., Enjouji, T., Deguchi, D., Kitasaka, T., Suenaga, Y., Hasegawa, J.i., Toriwaki, J.i., Takabatake, H., Natori, H.: Fast and accurate bronchoscope tracking using image registration and motion prediction. In: Medical Image Computing and Computer-Assisted Intervention, pp. 551–558 (2004)
18. Powell, M.J.D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. The Computer Journal **7**(2), 155–162 (1964)
19. Prados, E., Faugeras, O.: Shape from shading: a well-posed problem? In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 870–877 (2005)
20. Rai, L., Helferty, J.P., Higgins, W.E.: Combined video tracking and image-video registration for continuous bronchoscopic guidance. International Journal of Computer Assisted Radiology and Surgery **3**(3-4), 315–329 (2008)
21. Rashid, H.U., Burger, P.: Differential algorithm for the determination of shape from shading using a point light source. Image and Vision Computing **10**(2), 119–127 (1992)
22. Shah, P.L., Herth, F.J.: Current status of bronchoscopic lung volume reduction with endobronchial valves. Thorax **69**, 280–286 (2013)
23. Solomon, S.B., White Jr, P., Wiener, C.M., Orens, J.B., Wang, K.P.: Three-dimensional ct-guided bronchoscopy with a real-time electromagnetic position sensora comparison of two image registration methods. CHEST Journal **118**(6), 1783–1787 (2000)
24. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Transactions on Biomedical Engineering **57**(3), 736–745 (2010)
25. Visentini-Scarzanella, M., Stoyanov, D., Yang, G.Z.: Metric depth recovery from monocular images using shape-from-shading and specularities. In: 19th IEEE International Conference on Image Processing (ICIP), pp. 25–28 (2012)
26. Zhang, L., Yip, A.M., Brown, M.S., Tan, C.L.: A unified framework for document restoration using inpainting and shape-from-shading. Pattern Recognition **42**(11), 2961–2978 (2009)