(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

[Continued on next page]

(54) Title: GUIDANCE METHOD BASED ON 3D-2D POSE ESTIMATION AND 3D-CT REGISTRATION WITH APPLICATION TO LIVE BRONCHOSCOPY
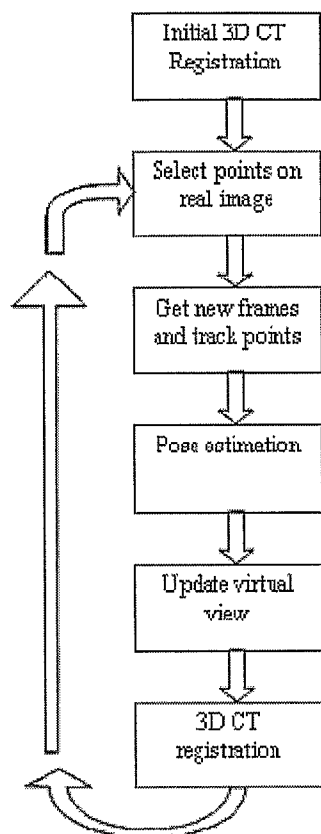
(57) Abstract: A method provides guidance to the physician during a live bronchoscopy or other endoscopic procedures. The 3D motion of the bronchoscope is estimated using a fast coarse tracking step followed by a fine registration step. The tracking is based on finding a set of corresponding feature points across a plurality of consecutive bronchoscopic video frames, then estimating for the new pose of the bronchoscope. In the preferred embodiment the pose estimation is based on linearization of the rotation matrix. By giving a set of corresponding points across the current bronchoscopic video image, and the CT-based virtual image as an input, the same method can also be used for manual registration. The fine registration step is preferably a gradient-based Gauss-Newton method that maximizes the correlation between the bronchoscopic video image and the CT-based virtual image. The continuous guidance is provided by estimating the 3D motion of the bronchoscope in a loop. Since depth-map information is available, tracking can be done by solving a 3D-2D pose estimation problem. A 3D-2D pose estimation problem is more constrained than a 2D-2D pose estimation problem and does not suffer from the limitations associated with computing an essential matrix. The use of correlation-based -cost, instead of mutual information as a registration cost^ makes it simpler to use gradient-based methods for registration.

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,
RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished*
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

- 1 -

GUIDANCE METHOD BASED ON 3D-2D POSE ESTIMATION AND 3D-CT REGISTRATION WITH APPLICATION TO LIVE BRONCHOSCOPY

REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application Serial No. 60/683,595, filed May 23, 2005, the entire content of which is incorporated herein by reference.

FIELD OF THE INVENTION

This invention relates generally to bronchoscopy and, in particular, to a method that provides guidance to the physician during a live bronchoscopy or other applications.

BACKGROUND OF THE INVENTION

For lung cancer assessment, the physician needs to perform a biopsy of the suspect cancer sites, such as the peripheral nodules or mediastinal lymph nodes. Such sites are first identified by analyzing the 3D CT image data of the chest. Later, during bronchoscopy, the physician attempts to reach these sites with the help of the live video obtained from a bronchoscope. The success of a standard bronchoscopy depends heavily on the skill level and experience of the physician. The success of the bronchoscopy could be increased if the physician received some form of guidance during the procedure.

Several guidance methods have been suggested in the past few years [1–5]. All of them use a CT-based (virtual) endoluminal rendering of the airway surface to obtain both the depth and visual data. They try to find the 3D location and orientation of the bronchoscope (pose) using the virtual renderings and incoming video frames. Bricault et al. proposed a method to register the bronchoscopic video (real) and 3D CT virtual bronchoscopic images [1]. The method uses the segmentation and shape from shading techniques to find the 3D surface for the real image and then does a 3D-3D registration of the computed surface with the virtual surface.

Mori et al. proposed a method which first tracks a set of points across the real frames to estimate the bronchoscopic motion by computing the essential matrix and then

does an estimation of the residual motion using image registration by Powell's method [3]. In [5], Mori et al. use a Kalman filter to predict bronchoscope motion and a new similarity measure to reduce the image area to be registered. Helferty et al. use a coarse tracking and fine registration approach [2,6]. The tracking is implemented by using the

5   standard optical-flow constraint equation and depth-map information from the virtual rendering to estimate the motion parameters. The registration is done by maximizing the mutual information between the real and virtual image using the simplex method.

The method proposed by Bricault et al. does not involve tracking and is limited to the bifurcation images [1]. The method of Mori et al. computes the essential matrix for

10   tracking [3] and Powell's method for registration. The approach has three limitations. Firstly, the use of Powell's method makes the registration step slow. Secondly, the essential matrix can not be determined if a subset of points are coplanar [7]. Thirdly, a translation can only be recovered up to a scale from the estimated essential matrix [7]. The optical-flow approach taken by Helferty et al. for tracking is slow since it involves

15   iterative warping and computation of gradients for the images [2, 6]. Use of simplex method makes the registration step slow as well.

SUMMARY OF THE INVENTION

This invention broadly resides in a system and method for providing guidance in conjunction with a diagnostic procedure. The method includes the steps of providing

20   previously acquired image data of a body lumen, acquiring live image data of the body lumen, and registering the previously acquired image data and the live image data in real time or near real-time. In the preferred embodiment, the registration is used to guide an instrument such as an endoscope, bronchoscope, colonoscope or laparoscope.

The previously acquired image data may be derived from virtual image data,

25   including computerized tomographic (CT) slices. Alternatively, the previously acquired image data may be derived from a prerecorded video image. The live image data may be derived from video data acquired during the diagnostic procedure or from a stream of incoming virtual images.

- 3 -

The invention has particular applicability to guidance during a live bronchoscopy. The 3D motion of the bronchoscope is estimated using a fast coarse tracking step followed by a fine registration step as necessary for correction purposes. The tracking is based on finding a set of corresponding feature points across a plurality of consecutive

5   bronchoscopic video frames, then estimating for the new pose of the bronchoscope.

In the preferred embodiment the pose estimation is based on linearization of the rotation matrix. By giving a set of corresponding points across the current bronchoscopic video image, and the CT-based virtual image as an input, the same method can also be used for manual registration.

10   The fine registration step is a gradient-based Gauss-Newton method that maximizes the correlation-based cost between the bronchoscopic video image and the CT-based virtual image. The continuous guidance is provided by estimating the 3D motion of the bronchoscope in a loop.

Since depth-map information is available, the tracking can be done by solving a

15   3D-2D pose estimation problem. A 3D-2D pose estimation problem is more constrained than a 2D-2D pose estimation problem and does not suffer from the limitations associated with computing an essential matrix. The use of correlation-based cost, instead of mutual information as a registration cost, makes it simpler to use gradient-based methods for registration.

20   BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows a set of 5 consecutive bronchoscopic video (real) frames displaying motion of the bronchoscope inside the airway tree;

FIGURE 2 shows the CT-based (virtual) endoluminal rendering of the airway surface based on the current estimate of the position and the orientation (pose) of the

25   bronchoscope;

FIGURE 3 shows the overall method of the invention;

FIGURE 4 demonstrates the manual registration step applied to a pair of virtual images;

- 4 -

FIGURE 5 illustrates the use of the manual registration step for the initial registration of the virtual and real image to start the guidance method;

FIGURE 6 shows the result of using the method for pose estimation by Lu et al. and demonstrates its unsuitability for our domain;

5   FIGURE 7 shows the computed corresponding matching point on the real image I_{R5}, given an input point on the virtual image I_V;

FIGURE 8 shows the results obtained by applying a registration step to a virtual image and a real image;

FIGURE 9 illustrates the optical-flow-based method for registration by Helferty et

10   al.; and

Figure 10 is an illustration of the fine registration step. In particular, Figure 10a shows a virtual image I_v to be registered with video image I_{Rc} (Fig. 10b) and a warped virtual image (Fig. 10c) after motion estimation.

DETAILED DESCRIPTION OF THE INVENTION

15   As discussed in the Summary of the Invention, to track the 3D motion of the bronchoscope, we use the fast coarse tracking and subsequent fine registration approach. We propose a 3D-2D pose estimation algorithm for tracking and a gradient-based Gauss-Newton method for registration which uses correlation-based cost as its cost function.

It should be noted that even if the tracking algorithm is 100 percent accurate, one

20   cannot avoid the fine registration step. This is because the 3D virtual surface data is not an accurate representation of the actual airway tree. The presence of the imaging artifacts, segmentation errors and issues related to lung capacity cause this. Hence, there will always be some drift errors during the tracking. If the drift errors are not taken care of by the registration step, they will accumulate to a point where tracking is no longer

25   successful.

In general the fine registration step takes more time. Accordingly, most of the motion should be estimated by a fast tracking method and the fine registration should only be done for correction. For tracking, we use correspondence of points between the real video frames along with the depth-map information from the virtual rendering to

- 5 -

solve a 3D-2D pose estimation problem. Since the accumulated rotation is small over a small number of consecutive real frames, linearization of the rotation matrix can be done. Thus, the 3D-2D pose estimation problem reduces to solving a linear system of equations. The same method can be used for manual registration if the manual

5   correspondence between the real and virtual image is given. For the fine registration step, we use the approach used for tracking by Helferty et al. [6]. This can be done by replacing the optical-flow constraint equation by a similar constraint based on correlation and replacing the source image with the virtual image.

Figure 1 shows a set of 5 consecutive bronchoscopic video (real) frames

10   displaying motion of the bronchoscope inside the airway tree. The first frame is considered as the current video frame I_{Rc} and the last frame is considered as I_{R5}. The frames in between are denoted by I_{R2}, I_{R3} and I_{R4}. Figure 2 shows the CT-based (virtual) endoluminal rendering of the airway surface based on the current estimate of the position and the orientation (pose) of the bronchoscope. The virtual image I_V is visually similar to

15   the current video frame I_{Rc}. The goal is to re-render the virtual image I_V so that it looks like I_{R5}, the real frame which is five frames apart from the current video frame I_{Rc}. This can be done by making use of the image motion observed in the real frames, the depth-map from the virtual rendering, and the visual similarity between the virtual image and real images.

20   Figure 3 shows the overall method. The first step is to do an initial registration of the virtual image I_V with the current real image I_{Rc}, either manually or automatically. The manual registration is done by giving corresponding points across the real and virtual image. In the preferred embodiment 6 points are used. Since the points in the virtual image also have the depth data associated with them denoted by W_i or (X_i, Y_i, Z_i), the 3D-

25   2D pose estimation method is applied to get the current pose or the 3D motion of the bronchoscope (R,T), which will make virtual image I_V look the same as the current real image I_{Rc}. I_V is re-rendered using the pose estimate. Automatic registration is done by the fine registration step.

The second step is to choose a multiplicity of points from the current real frame

30   I_{Rc} to be tracked over a plurality of consecutive frames. In the preferred embodiment 20

- 6 -

points are tracked across 5 frames. Since I_V is registered with I_{Rc}, we know the depths W_i associated with each point from the current depth-map. The third step is to track these 20 points using pairwise correspondence over the next 5 frames to get their new 2D locations (u_i,v_i). The fourth step is to estimate the new pose (R,T) using the 2D motion of

5   tracked points and their initial depths W_i. In the fifth step, the virtual image I_V is re-rendered using the new pose (R,T). The sixth step is to do fine registration between I_V and I_{R5} to take care of the drift errors due to tracking and then re-render I_V. Finally, I_{R5} is assigned as the new current real frame I_{Rc} and the algorithm goes from the second to the sixth step in a loop for continuous guidance.

10   <u>Selection and Tracking of Feature Points</u>

For fast coarse tracking of the bronchoscope, 20 feature points p_i are selected on image I_{Rc}. I_V is the matching virtual image for I_{Rc} and hence provides the depth-map information for each p_i. Every p_i has an associated depth given by the depth-map and its 3D location is given by W_i or (X_i,Y_i,Z_i). Each feature point pi is tracked over frames I_{R2},

15   I_{R3}, I_{R4}, and I_{R5} to get their new image location (u_i, v_i) in I_{R5}. The selection criterion for a feature point is entirely dependent on the method used for tracking it. It is for this reason that we explain the tracking method before the selection method.

<u>Tracking</u>

Once a point is selected in image I_{Rc}, it has to be tracked over frames I_{R2}, I_{R3}, I_{R4},

20   and I_{R5}. Tracking of feature points is done frame by frame by finding a matching corresponding point in the next frame I_{Ri+1} for each feature point in the previous frame I_{Ri}. Matching is done by finding the local shift (u_x^*, v_y^*), applied to previous location of point (x,y) in I_{Ri}, which minimizes the sum of squared differences (SSD) of image intensity patch around the point (x,y) in I_{Ri} and the shifted point in I_{Ri+1}:

$$(u_x^*, v_y^*) = \arg \min_{(u_x, v_y)} \sum_{(p,q)} w(p,q) \left[I_{Ri+1}(x + u_x + p, y + v_y + q) - I_{Ri}(x + p, y + q) \right]^2$$

25   (1)

- 7 -

In (1), w is a Gaussian window function applied to get better centering or localization of a matched point, (u_x, v_y) is varied over a search window S and (p,q) is varied over a patch P. The match of point (x,y) in I_{Ri} is given by $(x + u_x^t, y + v_y^t)$ in I_{Ri+1}.

Since the camera motion is assumed to be small between the frames, a simple translational image motion model is used, as justified by Shi and Tomasi [8]. To accommodate larger motion, a Gaussian pyramid is constructed.

The larger motion is estimated at a coarser level. This reduces the computation, since a smaller window P can be used for a template intensity patch and the search space S remains small at all the levels in the pyramid.

Selection

Before tracking, feature points p_i are chosen from frame I_{Rc}. A feature-based approach tries to use a small amount of image data to save computation and in some cases improve robustness. For a feature-based tracking, the first step is to select a set of feature points. A point is considered better for selection if it can promise to give a good match in the next frame. According to Triggs, each image-matching method defines a corresponding self-matching-based feature point detector and if a point cannot be accurately matched with itself then it cannot be matched robustly with any other point [9]. Hence the sharpness of a correlation or SSD peak obtained by matching a shifted image patch with itself under small motion has been the key criterion for many methods [8-10].

The SSD of an image patch with itself as a function $E(u_x, v_y)$ of a shift (u_x, v_y) is given by:

$$E(u_x, v_y) = \sum_{(x,y)} [I(x + u_x, y + v_y) - I(x,y)]^2$$

$$\hspace{6cm} (2)$$

where (x,y) is varied over a patch P. For a small shift (u_x, v_y),

$$E(u_x, v_y) = \sum_{(x,y)} [u_x I_x(x,y) + v_y I_y(x,y)]^2 = \begin{bmatrix} u_x & v_y \end{bmatrix} \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \begin{bmatrix} u_x \\ v_y \end{bmatrix}$$

$$(3)$$

and

- 8 -

$$A = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}$$

(4)

is known as the autocorrelation matrix. This form of the autocorrelation matrix is valid only for a simple translational motion model. For other motion models - e.g., affine motion, the number of parameters and number of dimensions are large. The eigenvalues of the autocorrelation matrix have been used to analyze the local image structure and classify a feature as a corner or an edge [8, 10].

Zuliani et al. have analyzed the relationship between different detectors based on the eigenvalues of the autocorrelation matrix [11]. They give a criterion for feature-selection called the condition number. The condition number K_{trans} measures the sensitivity of $E(u_x, v_y)$ to the perturbations ($\Box u_x, \Box v_y$). It is given by:

$$K_{trans} = \|(A + \epsilon I)^{-1}\|$$

(5)

where $\Box$ is a small number used for numerical stability. High value of a condition number means high sensitivity of the autocorrelation to the perturbations, which in turn means that the autocorrelation has a sharp peak at the point of interest.

For implementation, around 60 points are short-listed as feature-point candidates based on the strength of the image gradient at that point. If depth Z_i changes much around the point p_i in the virtual image I_V, the point may be close to a 3D edge and therefore, is not good for tracking or subsequently for pose estimation. Hence, thresholding is applied on the standard deviation of depths around the selected points to reject few more. These points are then sorted according to their condition number. Finally, the top 20 points are selected for tracking.

Pose Estimation

After a feature point P_i has been selected and tracked, its 3D location W_i in frame I_{Rc} and its new 2D location (u_i, v_i) in frame I_{R5} are known. Between frames I_{Rc} and I_{R5}, the bronchoscope has undergone a 3D motion (R,T).

Given the 3D locations W_i of n points in one reference frame and their 2D images (u_i, v_i), through perspective projection in another reference frame, solving for the rotation

and translation (R,T) between the reference frames is known as 3D-2D pose estimation problem. Thus, the goal of the pose estimation step is to estimate (R,T) given W_i and (u_i,v_i).

Many different classes of algorithms have been developed to solve this problem. Closed-form solutions exist for three or four points unless they are in a critical configuration [12–14]. These methods make use of the rigid geometrical constraints between the points to solve for a polynomial system of equations. For more than 4 points, one class of methods express a system of higher-order equations as a system of linear equations (over-dimensioning) to solve for depths first and then use the solution to absolute orientation problem to solve for the pose [15, 16]. Lu et al. give a fast iterative algorithm to determine the pose [17]. However, the method introduces large bias errors in the estimate of the translation when the object is very close to the camera or the depth of the object is comparable to the distance between the object and the camera, which holds true in our domain of application.

Since the feature tracking is done over a few frames at a time, it can be assumed that the accumulated rotation is small. Our method uses this assumption to linearize the rotation matrix. Our method is very close to Lowe's method [18] and the least-squares adjustment step done by Haralick et al. [19].

A 3D rotation matrix R is given by

$$R = R_x R_y R_z \qquad (6)$$

where

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \qquad (7)$$

$$R_y = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \qquad (8)$$

$$R_z = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

where θ, ψ and ϕ are the rotation angles around each axis. For small values of θ, ψ and ϕ, the rotation matrix can be written as

$$R = I + [\omega]_\times = I + \begin{bmatrix} 0 & -\phi & \psi \\ \phi & 0 & -\theta \\ -\psi & \theta & 0 \end{bmatrix} \tag{10}$$

A 3D world point W transformed by (R, T) is given by:

$$W' = R * W + T = (I + [\omega]_\times) * W + T = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} \theta \\ \psi \\ \phi \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \begin{bmatrix} X + \psi Z - \phi Y + t_x \\ Y + \phi X - \theta Z + t_y \\ Z + \theta Y - \psi X + t_z \end{bmatrix} \tag{11}$$

The image of W' through perspective projection is given by:

$$u = f\frac{X'}{Z'} \qquad v = f\frac{Y'}{Z'} \tag{12}$$

where f is the focal length. Henceforth, without loss of generality, f will be assumed to be 1.

Given n world points (X_i, Y_i, Z_i) and their image points (u_i, v_i) in another reference frame, we have to find

$$(R^*, T^*) = \arg\min_{(R,T)} \sum_{i=1}^{n} \left[(u_i - \frac{X_i'}{Z_i'})^2 + (v_i - \frac{Y_i'}{Z_i'})^2 \right] \tag{13}$$

where (X_i', Y_i', Z_i') are given by (11). We can solve for (R, T) using following equations:

$$u_i = \frac{X_i + \psi Z_i - \phi Y_i + t_x}{Z_i + \theta Y_i - \psi X_i + t_z} \qquad v_i = \frac{Y_i + \phi X_i - \theta Z_i + t_y}{Z_i + \theta Y_i - \psi X_i + t_z} \qquad i = 1...n \tag{14}$$

This gives an over-constrained system of linear equations:

- 11 -

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -u_iY_i & u_iX_i+Z_i & -Y_i & 1 & 0 & -u_i \\ v_iY_i-Z_i & v_iX_i & X_i & 0 & 1 & -v_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \theta \\ \psi \\ \phi \\ t_x \\ t_y \\ t_z \end{bmatrix} = \begin{bmatrix} \vdots \\ u_iZ_i-X_i \\ v_iZ_i-Y_i \\ \vdots \end{bmatrix} \qquad (15)$$

The linear system of equations (15), can be solved using singular value decomposition (SVD), although care should be taken to make very small singular values equal to zero while solving. Since the linearized form (10) of R is an approximation, we have to a iterate few more times to reach the correct solution for (R,T). Using the current solution for (R,T), the 3D points W_i are transformed to get a new estimate for W_i'. The residual transformation (R$\square$,T$\square$) should be determined by treating W_i' as the new W_i in (11). Then, (R,T) are updated as follows:

$$R = R' * R \qquad T = R' * T + T \qquad (16)$$

The method typically converges in 3 or 4 iterations.

3D CT Registration

After the pose estimation step, the virtual image I_V is re-rendered using the estimate for (R,T). This brings I_V visually closer to I_{R5}. But due to the presence of drift errors, I_V is still not a good match for I_{R5}. Using correlation as a criterion for visual match and the depth-map associated with I_V, the fine registration step estimates the residual motion (R$\square$,T$\square$) between I_V and I_{R5}. I_V is re-rendered using (R$\square$,T$\square$) to complete one loop of the guidance algorithm.

Registration using Correspondence

A fast way to register the two sources together is to use the same method as used for tracking. The only difference being that the correspondence will be found between the virtual image I_V and real image I_{R5}. The points, however, are selected on I_{R5} using the autocorrelation criterion. Since most of the information is contained in dark areas, the points are selected so that they sample all the dark regions. The selected points are matched with I_V using the correlation as the matching criterion in a Gaussian pyramid set up. The next step is to run the pose estimation algorithm and update I_V using the estimated pose. Although this method is fast, the matching does not work well for all

- 12 -

image pairs I_V and I_{R5}. The accuracy of the method depends on the distance of the bronchoscope from the branching point in the airway and the number of branches seen in I_{R5}. Manual registration uses this same method, but the corresponding points are provided manually.

5   <u>Registration using Maximization of Correlation</u>

Helferty et al. use the optical flow constraint equation along with the linearization of rotation matrix and the depth-map from the virtual image to do tracking [6]. We propose to use the same approach for fine registration of the virtual image I_V with the real image I_{R5}.

10   In the method given by Helferty et al., the goal is to register a real source image with a real target image by iteratively warping the source image towards the target image [6]. The 2D image motion of a point in the source image or optical flow (u_{xi}, v_{yi}) is governed by the 3D rotation and translation through:

$$u_{x_i} = \frac{\psi Z_i - \phi Y_i + t_x}{\theta Y_i - \psi X_i + t_z} \qquad v_{y_i} = \frac{\phi X_i - \theta Z_i + t_y}{\theta Y_i - \psi X_i + t_z}$$

(17)

Its derivation is almost same as given in Section 2.3. The optical flow constraint equation used to determine (u_x, v_y) is given by:

20   (18)

$$u_x I_x + v_y I_y + I_t = 0$$ Using (17) and (18), a

system of linear equations is set up to iteratively solve for (R, T). After each step, warping and computation of the gradients of the source image is done for the next iteration until convergence. The details can be found in [6].

25   In our case, the source image is I_V and the target image is I_{R5}. The optical flow constraint (18) is based on the SSD criterion for matching. Since in our case, both the virtual image I_V and the real image I_{R5} are from two different sources, the optical flow constraint cannot be used directly. However, if I_V and I_{R5} are normalized by subtracting the mean before registration, then (18) becomes a valid constraint and then both (17) and

30   (18) can be used together for fine registration.

- 13 -

EXAMPLES

Figure 4 demonstrates the manual registration step applied to a pair of virtual images. The six corresponding points are given manually across the two images. The unknown pose is computed by the pose estimation method using the correspondence and the depth-map associated with the virtual image on left. The virtual image on left is re-rendered using the estimate for pose. This results in a match between the left and the right image. The pose estimation method is very fast and generates the match instantaneously. The accuracy of the registration is dependent on the quality of the correspondence.

Figure 5 illustrates the use of the manual registration step for the initial registration of the virtual and real image to start the guidance method. Figure 6 shows the result of using the method for pose estimation by Lu et al. and demonstrates its unsuitability for our domain [17]. Although the correspondence has small errors (on the order of one pixel), we get large errors in the computed translation. Given below is a comparison between the correct pose (R,T) and computed pose (R1,T1):

$$R = \begin{bmatrix} 1 & -0.0061 & 0.0061 \\ 0.0064 & 0.9991 & -0.0417 \\ -0.0058 & 0.0417 & 0.9991 \end{bmatrix} \quad T = \begin{bmatrix} 0.0412 \\ -0.1444 \\ -0.1171 \end{bmatrix}$$

$$R1 = \begin{bmatrix} 0.8154 & 0.5763 & 0.0542 \\ 0.5781 & -0.8155 & -0.0262 \\ -0.0291 & -0.0527 & 0.9982 \end{bmatrix} \quad T1 = \begin{bmatrix} 2.2667 \\ -0.1280 \\ 16.0598 \end{bmatrix}$$

The link to the Matlab code for the pose estimation method by Lu et al. is given in the paper [17].

After feature selection, tracking and pose estimation, the fine registration step is required to take care of the drift errors. The fine registration step can either be based on correspondence or on optical-flow. Figure 7 shows the computed corresponding matching point on the real image I_{R5}, given an input point on the virtual image I_V. On the real image, the white point shows the initial guess for the match. The black point shows the final match obtained using the correlation criterion in a Gaussian pyramid set up. The use of Gaussian pyramid takes care of a large motion and saves on computation time by reducing the search space S.

- 14 -

Figure 8 shows the results obtained by applying the registration step in Section 2.4.1 to a virtual image and a real image. The points used for correspondence are displayed, too. Although for these two cases, the registration result is good, in general this is not the case. The accuracy of the registration step depends on the quality of the

5   correspondence. Good correspondence is not found, when the bronchoscope is either near or far from the bifurcations. In that case, the optical-flow-based fine registration step is used.

Figure 9 illustrates the optical-flow-based method for registration by Helferty et al. [6]. The source image is warped towards the target image, iteratively to recover the

10 residual motion. It is a gradient-based approach which can quickly recover the residual motion between I_V and I_{R5}. Figures 10a-c illustrate the fine registration step for a virtual and video image pair, with an example warped virtual image after motion estimation.

Fast tracking is an essential step in keeping the two sources together for guidance during bronchoscopy. It is not possible to escape from drift errors due to tracking, as

15 they arise partially from small errors in the 3D image data. A fine registration step is then necessary to take care of drift errors. Feature-based 3D-2D pose estimation is a fast and stable technique to do tracking. It does not suffer from instability associated with computing an essential matrix. If correspondence is computed across both the real and virtual images, then this same set up can be used for registration as well.

20 At least two other alternatives are available for guidance in the case of bronchoscopy. These alternatives include:

1.   The previously acquired image data is a prerecorded bronchoscopic video image sequence with associated depth information and the live source is incoming video from a bronchoscope.

25 2.   The previously acquired image data is a prerecorded bronchoscopic video image sequence with associated depth information and the live source is a stream of incoming virtual images, as may be acquired when interactively navigating through a 3D CT image.

The application has far-reaching applications, particularly in the field of image-guided

30 endoscopy.

- 15 -

In summary, we disclose a new 3D-2D pose estimation method based on linearization of the rotation matrix. The method is iterative and has fast convergence in case of small rotation. Using normalized images in the optical-flow constraint equation makes it possible to use the gradient-based registration method by Helferty et al. for fine

5   registration [6]. This approach is faster than using simplex method or Powell's method for registration.

- 16 -

REFERENCES

1. I. Bricault, G. Ferretti, and P. Cinquin, "Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy," IEEE Transactions on Medical Imaging 17, pp. 703–714, Oct 1998.

2. W. E. Higgins, J. P. Helferty, and D. R. Padfi, "Integrated bronchoscopic video tracking and 3D CT registration for virtual bronchoscopy," SPIE Medical Imaging 2003: Physiology and Function: Methods, Systems, and Applications, A. Clough and A. Amini (eds) 5031, pp. 80–89, May 2003.

3. K. Mori, D. Deguchi, J. Hasegawa, Y. Suenaga, J. Toriwaki, H. Takabatake, and H. Natori, "A method for tracking the camera motion of real endoscope by epipolar geometry analysis and virtual endoscopy system," MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 1–8, 2001.

4. H. Shoji, K. Mori, J. Sugiyama, Y. Suenaga, J. Toriwaki, H. Takabatake, and H. Natori, "Camera motion tracking of real endoscope by using virtual endoscopy system and texture information," SPIE Medical Imaging 2001: Physiology and Function from Multidimensional Images, Chin-Tu Chen and Anne V. Clough (eds) 4321, pp. 122–133, May 2001.

5. K. Mori, T. Enjoji, D. Deguchi, T. Kitasaka, Y. Suenaga, J. Toriwaki, H. Takabatake, and H. Natori, "New image similarity measures for bronchoscope tracking based on image registration between virtual and real bronchoscopic images," SPIE Medical Imaging 2004: Physiology and Function from Multidimensional Images, Amir A. Amini and Armando Manduca (eds) 5369, pp. 165–176, April 2004.

6. J. P. Helferty and W. E. Higgins, "Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance," IEEE Int. Conference on Image Processing, pp. II–961—II–964, Sept 2002.

7. R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, ISBN: 0521623049, 2000.

8. J. Shi and C. Tomasi, "Good features to track," IEEE Conf. Computer Vision and Pattern Recognition, pp. 593–600, June 1994.

9. B. Triggs, "Detecting keypoints with stable position, orientation and scale under illumination changes," European Conference on Computer Vision, pp. IV 100–113, May 2004.

10. C. Harris and M. Stephens, "A combined corner and edge detector," Alvey Vision Conference, pp. 147–151, 1988.

11. M. Zuliani, C. Kenney, and B. S. Manjunath, "A mathematical comparison of point detectors," IEEE Image and Video Registration Workshop, June 2004.

12. M. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Comm. ACM 24(6), pp. 381–395, 1981.

13. R. Horaud, B. Canio, and O. Leboullenx, "An analytic solution for the perspective 4-point problem," Computer Vision, Graphics, and Image Processing (1), pp. 33–44, 1989.

14.   R. M. Haralick, C. Lee, K. Ottenberg, and M. Nolle, "Analysis and solutions of the three point perspective pose estimation problem," Computer Vision and Pattern Recognition, pp. 592–598, 1991.

15.   A. Ansar and K. Daniildis, "Linear pose estimation from points or lines," IEEE Transactions on Pattern Analysis and Machine Intelligence 25, pp. 578–589, May 2003.

16.   L. Quan and Z. Lan, "Linear n-point camera pose determination," IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8), pp. 774–780, 1999.

17.   C. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," IEEE Transactions on Pattern Analysis and Machine Intelligence 22(6), pp. 610–622, 2000.

18.   D. G. Lowe, "Fitting parametrized three-dimensional models to images," IEEE Transactions on Pattern Analysis and Machine Intelligence 13(5), pp. 441–450, 1991.

19.   R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Analysis and solutions of the three point perspective pose estimation problem," IEEE Transactions on Systems, Man, and Cybernetics 19(6), pp. 1426–1446, 1989.

We claim:

1

- 18 -

1   1.   A method of providing guidance in conjunction with a diagnostic
2   procedure, comprising the steps of:
3   providing previously acquired image data of a body lumen;
4   acquiring live image data of the body lumen; and
5   registering the previously acquired image data and the live image data in real time
6   or near real-time.

1   2.   The method of claim 1, wherein the step of registering the previously
2   acquired image data and the live image data is used to guide an instrument during the
3   diagnostic procedure.

1   3.   The method of claim 2, wherein the instrument is an endoscope,
2   bronchoscope, colonoscope or laparoscope.

1   4.   The method of claim 1, wherein the previously acquired image data is
2   derived from virtual image data.

1   5.   The method of claim 1, wherein the previously acquired image data is
2   derived from computerized tomographic (CT) image data.

1   6.   The method of claim 1, wherein the previously acquired image data is
2   derived from a prerecorded video source.

1   7.   The method of claim 1, wherein the live image data is derived from video
2   data acquired during the diagnostic procedure.

1   8.   The method of claim 1, wherein the live image data is derived from a
2   stream of incoming virtual images.

1   9.   The method of claim 1, including the steps of:

2   a)   providing previously acquired image data including depth information;

3   b)   using an instrument to provide the live image data;

4   c)   estimating the three-dimensional (3D) motion of the instrument using the

5   depth information and the tracking of points associated with the live video image;

6   d)   determining a new pose by solving a 3D-2D pose estimation problem;

7   e)   updating the previously acquired image data based upon the new pose; and

8   f)   repeating steps c) through e) in system to complete the procedure.

1   10.   The method of claim 9, wherein pose estimation is based on the

2   linearization of a rotation matrix.

1   11.   The method of claim 9, further including the step of performing a fine

2   registration step to minimize errors associated with estimating the 3D motion of the

3   instrument.

1   12.   The method of claim 11, wherein the fine registration step is based upon a

2   gradient-based Gauss-Newton method.

1   13.   The method of claim 11, wherein the fine registration step includes the

2   steps of:

3   computing a virtual image of a body lumen; and

4   minimizing the cost between the virtual image and a live video image of the body

5   lumen.

1   14.   The method of claim 9, wherein the previously acquired images are

2   derived from computerized tomographic (CT) slices.

1   15.   The method of claim 9, wherein the instrument is an endoscope,

2   bronchoscope, colonoscope or laparoscope.

1   16.   The method of claim 9, including the steps of:

2   registering the previously acquired and live image data to obtain a current depth

3   map;

4   choosing a plurality of points associated with the live image data;

5   tracking the points over a plurality of consecutive frames to estimate the two-

6   dimensional (2D) motion of the tracked points; and

7   using the 2D motion of the tracked points and the current depth map to derive the

8   three-dimensional (3D) motion of the instrument.

1   17.   The method of claim 16, wherein:

2   the 3D locations (W_i) of n points in one of the consecutive frames and their 2D

3   images (u_i,v_i) are known through perspective projection in another reference frame; and

4   pose estimation solves for rotation and translation (R,T) using W_i and (u_i,v_i).

1   18.   The method of claim 16, wherein R is a three-dimensional matrix that is

2   linearized during pose estimation.

1   19.   A system for guiding a viewing instrument through a body lumen,

2   comprising:

3   a database storing previously acquired image data of the body lumen;

4   an input for receiving live image data of the body lumen; and

5   a processor in communication with the database and the input, the processor being

6   operative to register the previously acquired image data and the live image data in real

7   time or near real-time.

1   20.   The system of claim 19, wherein the database storing previously acquired

2   image data of the body lumen includes depth information.

1   21.   The system of claim 19, wherein the processor is further operative to:

2   a)   estimate three-dimensional motion using the depth information and the
3   tracking of points associated with the live image data; and
4   b)   solve a 3D-2D pose estimation problem to determine a new pose.

1   22.   The system of claim 21, wherein the processor is further operative to
2   linearize a rotation matrix to solve the 3D-2D pose estimation problem.

1   23.   The system of claim 19, wherein the processor is further operative to
2   perform a fine registration step to minimize errors associated with estimating three-
3   dimensional motion.

1   24.   The system of claim 23, wherein the fine registration uses a gradient-based
2   Gauss-Newton method.

1   25.   The system of claim 23, wherein the fine registration includes:
2   computing a virtual image of a body lumen;
3   minimizing the cost between the virtual image and the live image data.

1   26.   The system of claim 19, wherein the previously acquired image data are
2   derived from computerized tomographic (CT) slices.

1   27.   The system of claim 19, further including an endoscope, bronchoscope,
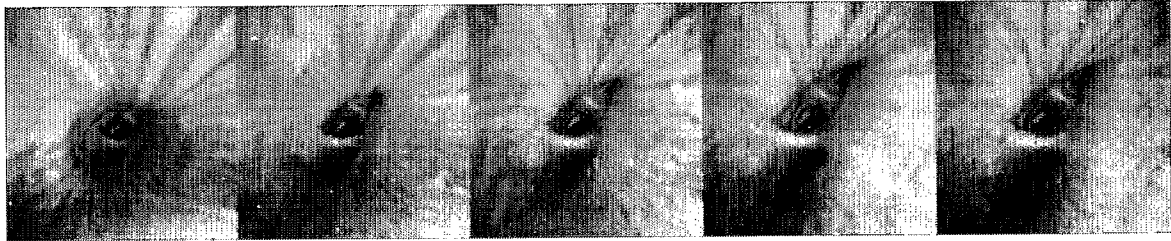2   colonoscope or laparoscope to obtain the live image data.

Figure 1. A set of 5 consecutive bronchoscopic video frames displaying motion of the bronchoscope inside the airway tree.
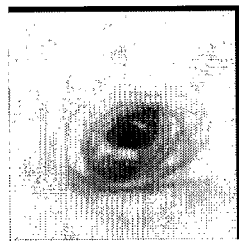

Figure 2. The CT-based endoluminal rendering of the airway surface based on the current estimate of the position and the orientation of the bronchoscope.
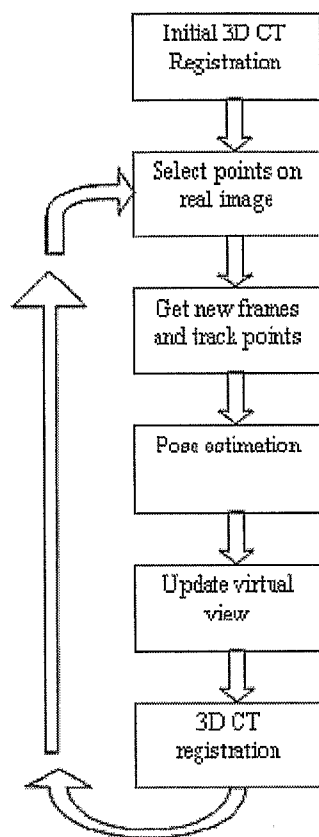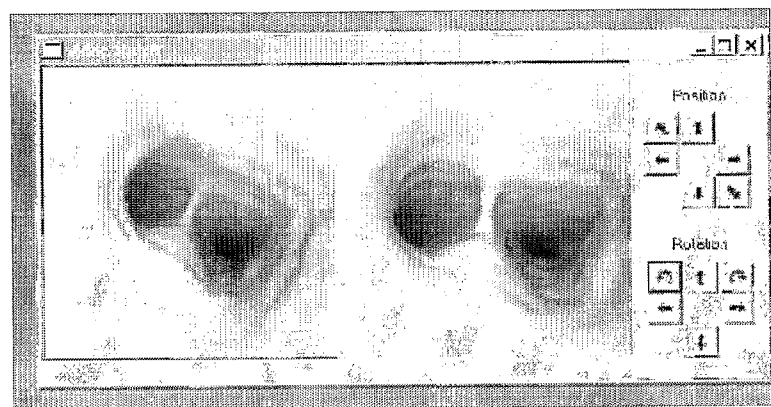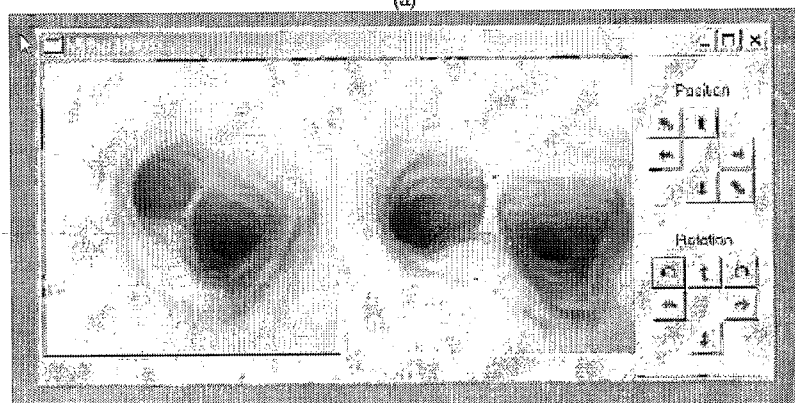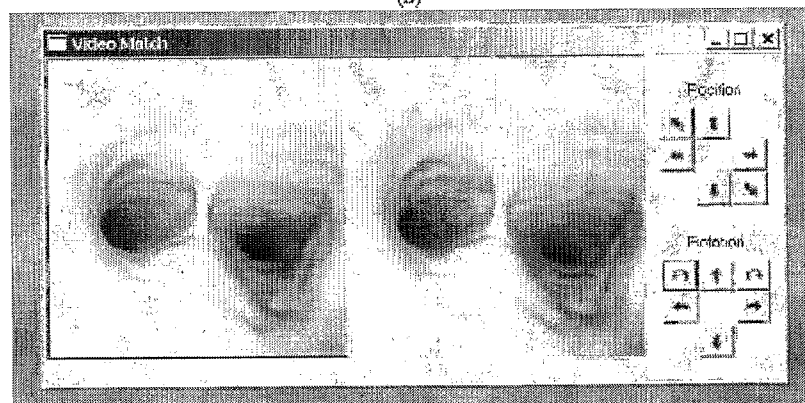
Figure 3. Method.

Figure 4. Illustration of the manual registration step applied to a virtual image pair. (a) Initial unregistered pair of virtual images. (b) 6 point correspondence given manually. (c) Registered pair after the pose estimation step.
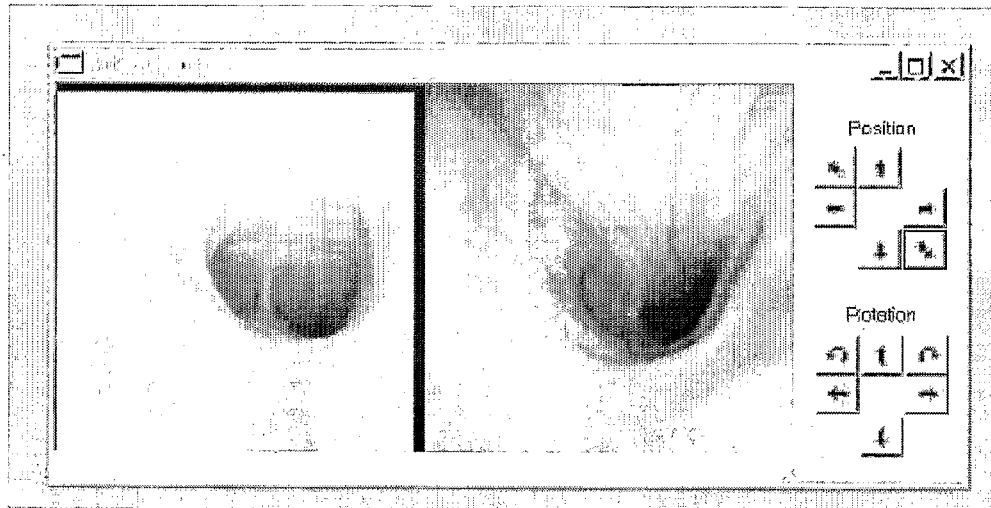
Figure 5. Result obtained after the initial manual registration is applied to a virtual image on the left and a real image on the right.
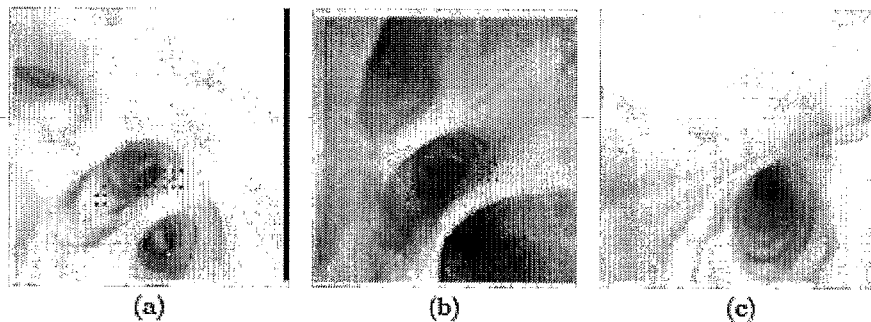

(a)   (b)   (c)

Figure 6. Failure of manual registration using the method by Lu *et al.* for pose estimation [17]. (a) Virtual image I_V close to the real image I_{Re} on right with the matching points. (b) Real image I_{Re} close to the virtual image I_V on left with the matching points. (c) Re-rendered virtual image I_V after the pose estimation step. I_V is now far off from I_{Re}.

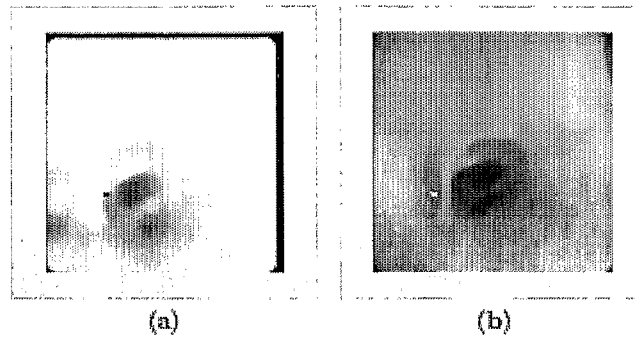(a)   (b)

Figure 7. Illustration of the matching method used to compute the correspondence between a virtual and real image for the registration step in Section 2.4.1. (a) Virtual image with a black point to be matched in the real image on right. (b) Real image with the white point showing the initial guess for the match and the black point showing the final match.

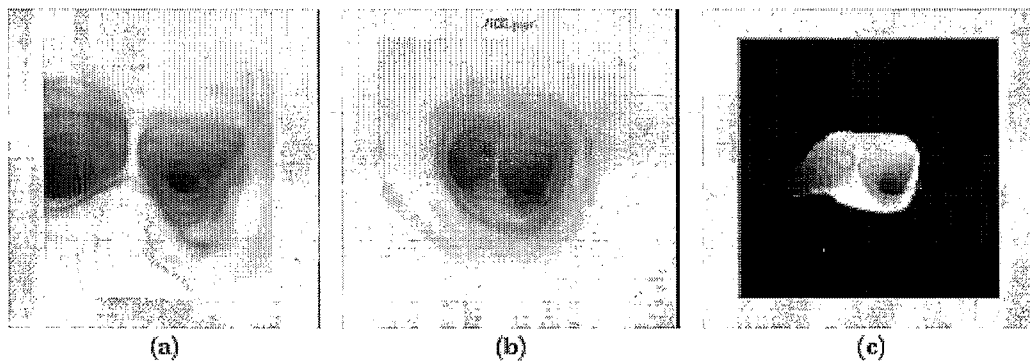(a)   (b)   (c)

Figure 9. Illustration of the fine registration step from Section 2.4.2 applied on a virtual image pair. (a) Source image to be registered with the target image in center. (b) Target image. (c) Warped source image after the application of the fine registration step.
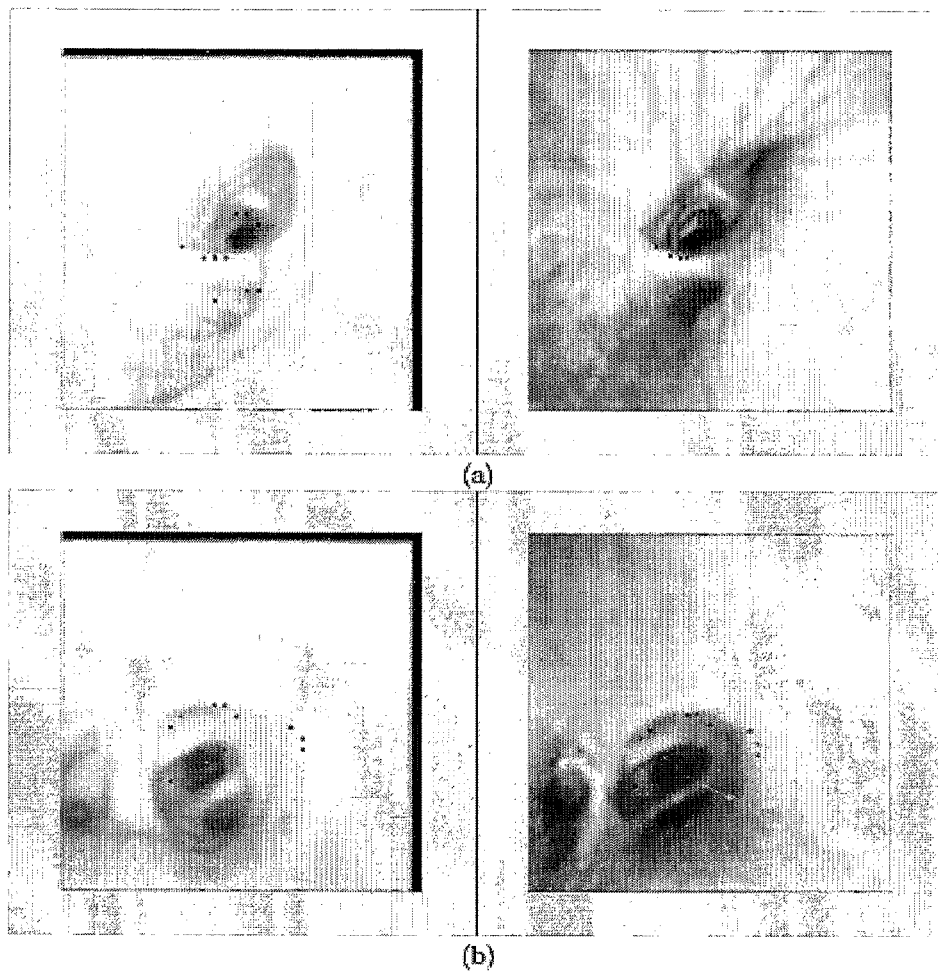
Figure 8. Results obtained by applying the registration step in Section 2.4.1 to a virtual image on the left an a real image on the right. (a) A registered pair of a virtual image on left and a real image on right with the matching points. (b) Another example of a registered pair of a virtual image on left and the matching real imag on right with their matching points.

专利名称(译)	支气管镜检查的3d-2d姿势估计和3d-CT配准		
公开(公告)号	EP1887931A2	公开(公告)日	2008-02-20
申请号	EP2006799926	申请日	2006-05-23
[标]申请(专利权)人(译)	宾州研究基金会		
申请(专利权)人(译)	宾州州立大学研究基金会		
当前申请(专利权)人(译)	宾州州立大学研究基金会		
[标]发明人	HIGGINS WILLIAM E MERRITT SCOTT A RAI LAV		
发明人	HIGGINS, WILLIAM, E. MERRITT, SCOTT, A. RAI, LAV		
IPC分类号	A61B5/05		
CPC分类号	A61B5/06 A61B1/00009 A61B5/415 A61B5/418 A61B6/032 A61B6/466 A61B6/50 A61B34/10 A61B34/20 A61B90/36 A61B90/361 A61B2034/105 A61B2034/2065 A61B2090/364 A61B2090/3762 G06K9/00208 G06K9/00214 G06K9/6206		
优先权	11/437229 2006-05-19 US 60/683595 2005-05-23 US		
外部链接	Espacenet		

摘要(译)

一种方法在实况内窥镜手术期间向医师提供指导。使用快速粗略跟踪步骤然后进行精细配准步骤来估计支气管镜的3D运动。跟踪基于在多个连续的支气管镜视频帧上找到一组对应的特征点，然后估计支气管镜的新姿势。通过在循环中估计支气管镜的3D运动来提供连续引导，并且可以通过解决3D-2D姿势估计问题来完成跟踪。