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 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.
Figure 3. Method.

Initial 3D CT Registration

Select points on real image

Get new frames and track points

Pose estimation

Update virtual view

3D CT registration

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

REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 11/437,229, filed May 19, 2006, which claims priority to U.S. Provisional Patent Application Ser. No. 60/683,595, filed May 23, 2005, the entire content of each of which is incorporated herein by reference.

GOVERNMENT SPONSORSHIP

[0002] This invention was made with government support under Grant No. R01 CA074325, awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] 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

[0004] 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.

[0005] 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.

[0006] 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 bronchoscopic 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 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.

[0007] 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 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 cannot 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 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

[0008] 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 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.

[0009] The previously acquired image data may be derived from virtual image data, 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.

[0010] 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 bronchoscopic video frames, then estimating for the new pose of the bronchoscope.

[0011] 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.

[0012] 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.

[0013] Since depth-map information is available, the 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 the 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.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. 1 shows a set of 5 consecutive bronchoscopic video (real) frames displaying motion of the bronchoscope inside the airway tree;

[0015] FIG. 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;

[0016] FIG. 3 shows the overall method of the invention;
FIG. 4A-4C demonstrate the manual registration step applied to a pair of virtual images, wherein FIG. 4A shows the initial unregistered pair of virtual images, FIG. 4B shows the 6-point correspondence given manually, and FIG. 4C shows the registered pair after the pose estimation step;

FIG. 5 illustrates the use of the manual registration step for the initialization of the virtual and real image to start the guidance method;

FIGS. 6A-6C show the result of using the method by Lu et al. for pose estimation, wherein FIG. 6A shows the virtual image $I_v$ close to the real image $I_r$, on right with the matching points, FIG. 6B shows the real image $I_r$, close to the virtual image $I_v$, on left with the matching points, and FIG. 6C shows the re-rendered virtual image $I_v$ after the pose estimation step;

FIGS. 7A and 7B show the computed corresponding matching point on the real image $I_r$, given an input point on the virtual image $I_v$;

FIGS. 8A and 8B show the results obtained by applying a registration step to a virtual image and a real image; and

FIGS. 9A-9C illustrate the optical-flow-based method for registration by Helferty et al.

DETAILED DESCRIPTION OF THE INVENTION

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 cannot avoid the fine registration step. This is because the 3D virtual surface data is not an accurate representation of the actual airway tree. 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 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 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 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.

FIG. 1 shows a set of 5 consecutive bronchoscopic video (real) frames displaying motion of the bronchoscope inside the airway tree. The first frame is considered as the current video frame $I_v$, and the last frame is considered as $I_{v_{0}}$. The frames in between are denoted by $I_{v_2}$, $I_{v_3}$, and $I_{v_4}$. FIG. 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 the current video frame $I_v$. The goal is to re-render the virtual image $I_v$ so that it looks like $I_{v_{0}}$, the real frame which is five frames apart from the current video frame $I_{v_{0}}$. 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.

FIG. 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 $I_{v_{0}}$, 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-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 $I_{v_{0}}$. $I_v$ 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 $I_{v_{0}}$ to be tracked over a plurality of consecutive frames. In the preferred embodiment 20 points are tracked across 5 frames. Since $I_{v}$ is registered with $I_{v_{0}}$, 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 $(u,v)$. The fourth step is to estimate the new pose (R,T) using the 2D motion of 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 $I_{v_{0}}$ to take care of the drift errors due to tracking and then re-render $I_v$. Finally, $I_{v_{0}}$ is assigned as the new current real frame $I_{v_{0}}$, and the algorithm goes from the second to the sixth step in a loop for continuous guidance.

Selection and Tracking of Feature Points

For fast coarse tracking of the bronchoscope, 20 feature points $p_i$ are selected on image $I_{v_{0}}$. $I_v$ is the matching virtual image for $I_{v_{0}}$, 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 $p_i$ is tracked over frames $I_{v_2}$, $I_{v_3}$, $I_{v_4}$, and $I_{v_{0}}$ to get their new image location $(u_i, v_i)$ in $I_{v_{0}}$. 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.

Tracking

Once a point is selected in image $I_{v_{0}}$, it has to be tracked over frames $I_{v_2}$, $I_{v_3}$, $I_{v_4}$, and $I_{v_{0}}$. Tracking of feature points is done by finding a matching corresponding point in the next frame $I_{v_{i+1}}$ for each feature point in the previous frame $I_{v_i}$. Matching is done by finding the local shift $(u_i, v_i)$ applied to previous location of point $(x_i, y_i)$ in $I_{v_{0}}$ which minimizes the sum of squared differences (SSD) of image intensity patch around the point $(x_i, y_i)$ in $I_{v_{0}}$ and the shifted point in $I_{v_{i+1}}$.
In (1), \(w\) is a Gaussian window function applied to get better centering or localization of a matched point, \((u, v)\) is varied over a search window \(S\) and \((p,q)\) is varied over a patch \(P\). The match of point \((x,y)\) in \(I_{0}\) is given by \((x+u, y+v)\) in \(I_{p+q}\).

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 \(I_{0}\). 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 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 \(E(u,v)\) of a shift \((u,v)\) is given by:

\[
E(u,v) = \sum_{(x,y)} [(I(x+u, y+v) - I(x,y))]^2
\]

where \((x,y)\) is varied over a patch \(P\). For a small shift \((u,v)\),

\[
E(u,v) = \sum_{(x,y)} [u_I(x,y) + v_J(x,y)]^2
\]

\[
= u_I \sum_{x} b_{x} \sum_{y} b_{y} \left( \sum_{x} b_{x} \sum_{y} b_{y} \right)
\]

\[A = \frac{\sum_{x} b_{x} \sum_{y} b_{y}}{\sum_{x} b_{x} \sum_{y} b_{y}}
\]

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 \(K_{\text{ess}}\) measures the sensitivity of \(E(u_v)\) to the perturbations \((\Delta u_v)\).

\[
K_{\text{ess}} = \frac{\epsilon}{(4\pi d)^{2/3}}
\]

where \(\epsilon\) 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 \(Z\) changes much around the point \(p_i\), in the virtual image \(I_{y}\), 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 \(I_{0}\), and its new 2D location \((u,v)\) in frame \(I_{1}\), are known. Between frames \(I_{0}\) and \(I_{1}\), the brachiosaurus has undergone a 3D motion \((R,T)\).

Given the 3D locations \(W_i\) of \(n\) points in one reference frame and their 2D images \((u,v)\), 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 \((u,v)\).

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

\[
R = R_s R_c
\]

where

\[
R_s = R_s(x, y, z)
\]
where $\theta$, $\psi$ and $\phi$ are the rotation angles around each axis. For small values of $\theta$, $\psi$ and $\phi$, the rotation matrix can be written as

$$
R = I + [w] = I + 
\begin{bmatrix}
0 & -\phi & \psi \\
\phi & 0 & -\theta \\
-\psi & \theta & 0
\end{bmatrix}
$$

[0043] A 3D world point $W$ transformed by $(R, T)$ is given by:

$$
W' = R \cdot W + T
$$

[0044] The image of $W'$ through perspective projection is given by:

$$
u = f \frac{X'}{Z'} = f \frac{Y'}{Z'}
$$

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

[0045] 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

$$
(R', T') = \arg \min_{(R,T)} \sum_{i=1}^{n} \left[ (u_i - \frac{X'_i}{Z'_i})^2 + (v_i - \frac{Y'_i}{Z'_i})^2 \right]
$$

where $(X'_i, Y'_i, Z'_i)$ are given by (11). We can solve for $(R, T)$ using following equations:

$$
u_i = \frac{X_i + \phi Z_i - \phi Y_i + t_z}{Z_i + \theta T_i - \phi X_i + t_x} \quad \psi_i = \frac{Y_i + \phi X_i - \theta Z_i + t_y}{Z_i + \theta T_i - \phi X_i + t_x}
$$

[0046] This gives an over-constrained system of linear equations:

$$
\begin{bmatrix}
0 & -\phi & \psi & 1 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\theta \\
\psi \\
\phi \\
1
\end{bmatrix}
= \begin{bmatrix}
u_i Z_i - X_i \\
u_i Y_i - Z_i
\end{bmatrix}
$$

[0047] 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 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 $(R'_i, T'_i)$ should be determined by treating $W'_i$ as the new $W_i$ in (11). Then, $(R, T)$ are updated as follows:

$$
R = R'_i - R \quad T = T'_i - T
$$

[0048] The method typically converges in 3 or 4 iterations.

3D CT Registration

[0049] 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 $I_{gs}$. But due to the presence of drift errors, $I_v$ is still not a good match for $I_{gs}$. Using correlation as a criterion for visual match and the depth-map associated with $I_{gs}$, the fine registration step estimates the residual motion $(R'_i, T'_i)$ between $I_v$ and $I_{gs}$. It is re-rendered using $(R'_i, T'_i)$ to complete one loop of the guidance algorithm.

Registration using Correspondence

[0050] 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 $I_{gs}$. The points, however, are selected on $I_{gs}$ 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 image pairs $I_v$ and $I_{gs}$. 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 $I_{gs}$. Manual registration uses this same method, but the corresponding points are provided manually.
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 \( I_{rg} \).

**[0052]** 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 \((u_i, v_i)\) is governed by the 3D rotation and translation through:

\[
\begin{align*}
\frac{\partial x_i}{\partial \theta_x} &= \phi x_i - \psi y_i + t_x \\
\frac{\partial y_i}{\partial \theta_x} &= \psi x_i + \phi y_i + t_x
\end{align*}
\]

(17)

**[0053]** Its derivation is almost the same as given above. The optical flow constraint equation used to determine \((u_i, v_i)\) is given by:

\[
u_i = \frac{\partial x_i}{\partial \theta_x}, \quad \nu_i = \frac{\partial y_i}{\partial \theta_x}
\]

(18)

**[0054]** 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].

**[0055]** In our case, the source image is \( I_v \) and the target image is \( I_{rg} \). 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 \( I_{rg} \) are from two different sources, the optical flow constraint cannot be used directly. However, if \( I_v \) and \( I_{rg} \) are normalized by subtracting the mean before registration, then (18) becomes a valid constraint and then both (17) and (18) can be used together for fine registration.

**EXAMPLES**

**[0056]** Figs. 4A-4C demonstrate 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.

**[0057]** Fig. 5 illustrates the use of the manual registration step for the initial registration of the virtual and real image to start the guidance method. Figs. 6A-6C show 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 \((R_1, T_1)\):

\[
R = \begin{bmatrix}
1 & -0.0061 & 0.0661 \\
0.0064 & 0.9991 & -0.0417 \\
-0.0058 & 0.0417 & 0.9991
\end{bmatrix},
T = \begin{bmatrix}
0.0412 \\
-0.1444 \\
-0.1171
\end{bmatrix}
\]

**[0058]** The link to the Matlab code for the pose estimation method by Lu et al. is given in the paper [17].

**[0059]** 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. Figs. 7A and 7B show the computed corresponding matching point on the real image \( I_{rg} \), 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.

**[0060]** Figs. 8A and 8B show the results obtained by applying the registration step 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 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.

**[0061]** Figs. 9A and 9B illustrate 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 residual motion. It is a gradient-based approach which can quickly recover the residual motion between \( I_v \) and \( I_{rg} \).

**[0062]** 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 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.

**[0063]** 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.

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.

**[0064]** The application has far-reaching applications, particularly in the field of image-guided endoscopy.

**[0067]** 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 registration [6]. This approach is faster than using simplex method or Powell’s method for registration.

REFERENCES


We claim:
1. A video-based method for providing a pose estimate of an endoscope in conjunction with a live endoscopic procedure, the method comprising:
   acquiring 3D image data of a target structure in advance of a live endoscopic procedure;
   receiving a frame of live endoscopic video image data including the target structure; and
   registering the frame of endoscopic video image data with acquired 3D image data to determine a pose estimate of the endoscope.
2. The method of claim 1, wherein the pose estimate is used to guide the endoscope to a suspect site.
3. The method of claim 2, wherein the suspect site is a mediastinal lymph node.
4. The method of claim 1, wherein the endoscope is a bronchoscope.
5. The method of claim 1, wherein the target structure is the airway tree.
6. The method of claim 5, wherein the registration is performed either near to, or far from, a bifurcation of an airway tree.
7. The method of claim 1, wherein the registration comprises an initial registration step to initiate guidance of the endoscope.
8. The method of claim 1, wherein the registration step includes the steps of:
   (a) estimating a three-dimensional location of the endoscope using (i) known motion information from said live endoscopic video image data, and (ii) local depth information obtained from the previously acquired 3D image data; and
   (b) determining a new pose of the endoscope based on the 3D location estimated in step (a).
9. The method of claim 8, further including the step of performing a fine registration step to minimize errors associated with estimating the 3D motion of the endoscope.
10. The method of claim 9, wherein the fine registration step is performed manually.
11. The method of claim 8, wherein the registration step further includes the steps of:
   (c) updating the previously acquired 3D image data in accordance with the new pose; and
   (d) repeating steps (a) through (c) until the guidance is terminated.
12. The method of claim 1, including the step of receiving a sequence of consecutive video image frames as opposed to a single frame.
13. The method of claim 12, wherein the sequence of video image frames is 5 consecutive video frames.
14. A video-based method for registering previously acquired 3D image data and live endoscopic video image data of a patient to obtain a current pose of an endoscope comprising:
   a) performing an initial registration to register the previously acquired 3D image data and a frame of the live video image data to obtain a current pose and a current depth map;
   b) selecting a plurality of points associated with the live video image data;
   c) tracking the points over a plurality of consecutive frames to estimate the two-dimensional (2D) motion of the tracked points;
   d) deriving a three-dimensional (3D) motion of the endoscope using the 2D motion of the tracked points and the current depth map;
   e) determining a new pose based on the current depth map and 3D motion of the endoscope; and
   f) updating the current pose based upon the new pose.
15. The method of claim 14, wherein steps b) through f) are repeated until reaching a suspect site.
16. The method of claim 15, wherein said site is along an airway.
17. A video-based method for providing a pose estimate of an endoscope in conjunction with a live endoscopic procedure, the method comprising:
   receiving a frame of live endoscopic video image data of a target structure from the endoscope; and
   registering previously acquired 3D image data of the target structure and the frame of endoscopic video image data of the target structure to provide a pose estimate of the endoscope; and
   wherein the step of registering is performed using information arising from said frame of endoscopic video image data and the previously acquired 3D image data.
18. The method of claim 17, wherein the step of registering the previously acquired 3D image data and the live video image data is used for guiding said endoscope to a suspect site.
19. The method of claim 18, wherein said suspect site is a mediastinal lymph node.
20. The method of claim 17, wherein the endoscope is a bronchoscope.
21. The method of claim 17, wherein said target structure is the airway tree.
22. The method of claim 21, wherein said registration is performed either near or far from a bifurcation.
23. The method of claim 17, wherein said registering step is an initial registration step to start guidance of the endoscope.
24. The method of claim 18, wherein the registration step includes the steps of:
   a) estimating a three-dimensional location of the endoscope using (i) known motion information from said live endoscopic video image data, and (ii) local depth information obtained from the previously acquired 3D image data; and
   b) determining a new pose of the endoscope based on the 3D location estimated in step (a).
25. The method of claim 24, further including the step of performing a fine registration step to minimize errors associated with estimating the 3D motion of the endoscope.
26. The method of claim 25, wherein the fine registration step is performed manually.
27. The method of claim 24, wherein the registration step further includes the steps of:
   (c) updating the previously acquired 3D image data in accordance with the new pose; and
   (d) repeating steps (a) through (c) until the guidance is terminated.
28. The method of claim 17, wherein said receiving comprises receiving a sequence of consecutive video image frames.
29. The method of claim 28, wherein said sequence of video image frames is 5 consecutive video frames.
30. A video-based method for registering previously acquired 3D image data and live endoscopic video image data of a patient to obtain a current pose of an endoscope comprising:
   g) performing an initial registration to register the previously acquired 3D image data and a frame of the live video image data to obtain a current pose and a current depth map;
   h) selecting a plurality of points associated with the live video image data;
   i) tracking the points over a plurality of consecutive frames to estimate the two-dimensional (2D) motion of the tracked points;
   j) deriving a three-dimensional (3D) motion of the endoscope using the 2D motion of the tracked points and the current depth map;
   k) determining a new pose based on the current depth map and 3D motion of the endoscope; and
   l) updating the current pose based upon the new pose.
31. The method of claim 30 comprising repeating steps b) through f) until reaching a suspect site.
32. The method of claim 31, wherein said site is along an airway.