

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 November 2006 (30.11.2006)

PCT

(10) International Publication Number
WO 2006/127713 A2

(51) International Patent Classification:
A61B 5/05 (2006.01)

(74) Agents: WATHEN, Douglas, L. et al.; GIFFORD, KRASS, GROH, SPRINKLE, ANDERSON & CITKOWSKI, P.C., 2701 Troy Center Dr., Ste. 330 ; Pob 7021, Troy, MI 48007-7021 (US).

(21) International Application Number:
PCT/US2006/019927

(22) International Filing Date: 23 May 2006 (23.05.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/683,588 23 May 2005 (23.05.2005) US
11/437,230 19 May 2006 (19.05.2006) US

(71) Applicant (for all designated States except US): THE PENN STATE RESEARCH FOUNDATION [US/US]; 304 Old Main, University Park, PA 16802 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): HIGGINS, William, E. [US/US]; 141 Legion Lane, State College, PA 16801 (US). MERRITT, Scott, A. [US/US]; 843 Willard Circle, State College, PA 16803 (US). RAI, Lav [US/US]; 411 WAUPELANI DRIVE, Apt. 101, State College, PA 16801 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

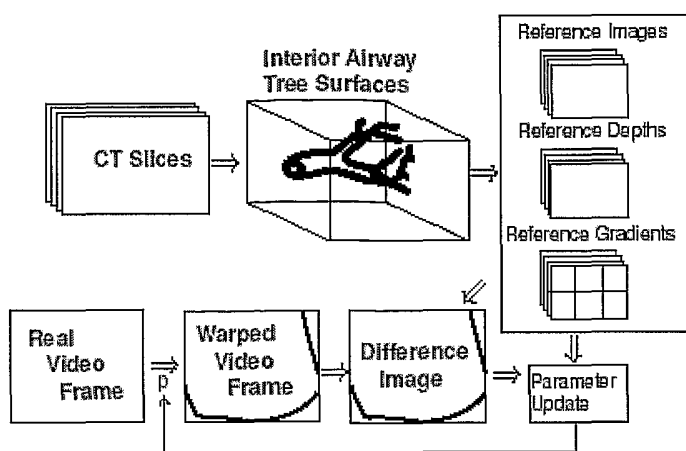
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, 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

[Continued on next page]

(54) Title: A FAST 2D-3D IMAGE REGISTRATION METHOD WITH APPLICATION TO CONTINUOUSLY GUIDED ENDOSCOPY



(57) Abstract: A novel framework for fast and continuous registration between two imaging modalities is disclosed. The approach makes it possible to completely determine the rigid transformation between multiple sources at real-time or near real-time frame-rates in order to localize the cameras and register the two sources. A disclosed example includes computing or capturing a set of reference images within a known environment, complete with corresponding depth maps and image gradients. The collection of these images and depth maps constitutes the reference source. The second source is a real-time or near-real time source which may include a live video feed. Given one frame from this video feed, and starting from an initial guess of viewpoint, the real-time video frame is warped to the nearest viewing site of

the reference source. An image difference is computed between the warped video frame and the reference image. The viewpoint is updated via a Gauss-Newton parameter update and certain of the steps are repeated for each frame until the viewpoint converges or the next video frame becomes available. The final viewpoint gives an estimate of the relative rotation and translation between the camera at that particular video frame and the reference source. The invention has far-reaching applications, particularly in the field of assisted endoscopy, including bronchoscopy and colonoscopy. Other applications include aerial and ground-based navigation.



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

A FAST 2D-3D IMAGE REGISTRATION METHOD WITH APPLICATION TO CONTINUOUSLY GUIDED ENDOSCOPY

REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application Serial No. 60/683,588, filed May 23, 2005, the entire content of which is incorporated herein by reference.

5

FIELD OF THE INVENTION

This invention relates generally to image correlation and, in particular, to a fast image registration method applicable to guided endoscopy and other fields.

BACKGROUND OF THE INVENTION

Registration and alignment of images taken by cameras at different spatial
10 locations and orientations within the same environment is a task which is vital to many applications in computer vision and medical imaging. For example, registration between images taken by a mobile camera and those from a fixed surveillance camera can assist in robot navigation. Other applications include the ability to construct image mosaics and panoramas, high dynamic range images, or super-resolution images, or the fusion of
15 information between the two sources.

However, because the structure of a scene is inherently lost by the 2D imaging of a 3D scene, only partial registration information can typically be recovered. In many applications, depth maps can be generated or estimated to accompany the images in order to reintroduce the structure to the registration problem.

20 Most currently available 2D alignment algorithms use a gradient descent approach which relies on three things: a parameterization of the spatial relationship between two images (e.g., the 2D rotation and translation between two 2D images), the ability to visualize these images under any value of the parameters (e.g., viewing a 2D reference image rotated by 30 degrees), and a cost function with associated image
25 gradient information which allows an estimate of the parameter updates to be calculated.

- 2 -

Among the most straightforward and earliest of these algorithms is the Lucas-Kanade algorithm, which casts image alignment as a Gauss-Newton minimization problem [5]. A subsequent refinement to this algorithm includes the inverse compositional alignment algorithm which greatly speeds the computation of the parameter update by recasting the problem, allowing all gradient and Hessian information to be calculated one time instead of every iteration [6]. Several other improvements have centered around the choice of parameters and the corresponding image warps these parameterizations induce. For example, images obtained from two identical cameras observing the same scene from a different location can be approximately related by an affine transformation or an 8-parameter homography [7].

The main problem with these types of parameterizations is that they do not truly capture the physically relevant parameters of the system, and, in the case of the homography, can lead to overfitting of the image. A more recent choice of parameters attempts to match two images obtained from a camera that can have arbitrary 3D rotations around its focal point [8]. This algorithm succeeds in extracting the physically relevant parameters (rotation angles about the focal point). However, while it is able to handle small translations, it cannot handle general translation and treats it as a source of error.

Little has been done to tackle the problem of registration of two images generated by cameras related by a general rigid transformation (i.e., 3D rotation and translation). The main reason for this is that the accurate visualization of a reference image as seen from a different camera location ideally requires that the depth map associated with that image be known — something which is not generally true. In certain situations, such as a robot operating in a known man-made environment, or during bronchoscopy where 3D scans are typically performed before the procedure, this information is known. Indeed, even in situations where the depth map is unknown, it can often be estimated from the images themselves.

An example of this is the aforementioned shape-from-shading problem in bronchoscopy guidance [9]. Current practice requires a physician to guide a bronchoscope from the trachea to some predetermined location in the airway tree with

- 3 -

little more than a 3D mental image of the airway structure, which must be constructed based on the physician's interpretation of a set of computed tomography (CT) films. This complex task can often result in the physician getting lost within the airway during navigation [1]. Such navigation errors result in missed diagnoses, or cause undue stress to the patient as the physician may take multiple biopsies at incorrect locations, or the physician may need to spend extra time returning to known locations in order to reorient themselves.

In order to alleviate this problem and increase the success rate of bronchoscopic biopsy, thereby improving patient care, some method of locating the camera within the airway tree must be employed. Fluoroscopy can provide intraoperative views which can help determine the location of the endoscope. However, as the images created are 2D projections of the 3D airways, they can only give limited information of the endoscope position. Additionally, fluoroscopy is not always available and comes with the added cost of an increased radiation dose to the patient.

A few techniques also exist that determine the bronchoscope's location by attempting to match the bronchoscope's video to the preoperative CT data. One method uses shape-from-shading, as in [2], to estimate 3D surfaces from the bronchoscope images in order to do 3D-to-3D alignment of the CT airway surface. This method requires many assumptions to be made regarding the lighting model and the airway surface properties and results in large surface errors when these assumptions are violated. A second method of doing this is by iteratively rendering virtual images from the CT data and attempting to match these to the real bronchoscopic video using mutual information [3] or image difference [4].

While these methods can register the video to the CT with varying degrees of success, all operate very slowly and only involve single-frame registration — none of them are fast enough to provide continuous registration between the real video and the CT volume. They rely on optimization methods which make no use of either the gradient information nor the known depth of the CT-derived images, and thus require very computationally intensive searches of a parameter space.

SUMMARY OF THE INVENTION

This invention resides in a novel framework for fast and continuous registration between two imaging modalities. A method of registering an image according to the invention comprises the steps of providing a set of one or more reference images with depth maps, and registering the image to at least one of the reference images of the set using the depth map for that reference image. The image and the reference set may both be real, virtual, or one real with the other virtual. The set of reference images may endoscopic, derived from a bronchoscope, colonoscope, laparoscope or other instrument. The registration preferably occurs in real-time or near real-time, and one or more of the images in the set of reference images can be updated before, during, or after registration.

According to a robust implementation, the set of reference images represents viewpoints with depth maps and image gradients, and the image to be registered is derived from a video feed having a plurality of consecutive frames. The method includes the steps of:

- a) warping a frame of the video to the nearest viewpoint of the reference source;
- b) computing an image difference between the warped video frame and the reference image;
- c) updating the viewpoint using a Gauss-Newton parameter update; and
- d) repeating steps a) through c) for each frame until the viewpoint converges or the next video frame becomes available.

The invention makes it possible to completely determine the rigid transformation between multiple sources at real-time or near real-time frame-rates in order to register the two sources. A disclosed embodiment involving guided bronchoscopy includes the following steps:

1. In the off-line phase, a set of reference images is computed or captured within a known environment, complete with corresponding depth maps and image gradients. The collection of these images and depth maps constitutes the reference source.
2. The second source is a real-time source from a live video feed. Given one frame from this video feed, and starting from an initial guess of viewpoint,

- 5 -

the real-time video frame is warped to the nearest viewing site of the reference source.

3. An image difference is computed between the warped video frame and the reference image.

5 4. The viewpoint is updated via a Gauss-Newton parameter update.

5. Steps 2-4 are repeated for each frame until the viewpoint converges or the next video frame becomes available. The final viewpoint gives an estimate of the relative rotation and translation between the camera at that particular video frame and the reference source.

10 The invention has far-reaching applications, particularly in the field of assisted endoscopy, including bronchoscopy and colonoscopy. Other applications include aerial and ground-based navigation.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows a block diagram of registration algorithm for guided
15 bronchoscopy;

FIGURES 2A-2F show source images and results for virtual-to-real registration; specifically, Figure 2A shows a real video frame, Figure 2B shows a warped real image at initial viewpoint, Figure 2C shows edges from final reference image overlaid, Figure 2D shows a reference virtual image corresponding to final registration, Figure 2E shows a
20 warped real image at final viewpoint, and Figure 2F shows edges of corresponding virtual image overlaid; and

FIGURES 3A-3C show source images and sample results for virtual-to-virtual registration; specifically, Figure 3A shows a real image, Figure 3B shows a reference image, and Figure 3C shows a warped real image.

25 DETAILED DESCRIPTION OF THE INVENTION

Broadly, this invention is a 2D image alignment algorithm which is augmented to three dimensions by introducing the depth maps of the images. The method provides an ideal way to extend the existing matching framework to handle general 3D camera

- 6 -

motion, allowing one to directly solve for the extrinsic parameters of the camera and localize it within its environment.

For the purpose of explaining the method in a very concrete fashion, discussion shall focus on a situation similar to that used in guided bronchoscopy. In a typical
 5 bronchoscopic procedure, a CT scan is initially performed and can subsequently be processed to extract the airway tree surfaces. The interior of the hollow airway tree constitutes the known environment. During bronchoscopy, the bronchoscope is inserted into the airways and a camera mounted on the tip transmits in real-time a sequence of real
 10 bronchoscopic (RB) video images. Assuming that the calibration parameters of the endoscope are known, virtual bronchoscopic (VB) images (endoluminal renderings) can be rendered at arbitrary viewpoints within the airway tree. It is also clear that the depths corresponding to each pixel of the VB image can be immediately calculated and form a virtual depth map (VDM).

The problem is that we have a fixed real-time RB image from an unknown
 15 location within the interior of an airway, but we also have a known VB source with known location and 3D information that enables us to create manifestations of the same hollow airway structure from arbitrary viewpoint. Given the above setup, the goal is to locate the source of the RB image by attempting to find the best match between that fixed RB image and any possible VB endoluminal rendering. A fairly straightforward
 20 approach to accomplish this is by employing a Gauss-Newton gradient descent algorithm that attempts to minimize a difference measure between the RB image and the VB image with respect to the viewing parameters (i.e., viewpoint) of the virtual image. The method for doing this is similar to the Lucas-Kanade image alignment algorithm [5].

The objective function used in [5, 6] is the sum squared difference (SSD) between
 25 the pixel intensities of the two images, although weighted SSD is equally viable, and weighted or unweighted normalized cross-correlation (CC) can be used if some additional weak assumptions are made. Using the SSD, the objective function can thus be written as

$$E = \sum_{u,v} [\mathbf{I}_v(u, v; \mathbf{p} + \Delta \mathbf{p}) - \mathbf{I}_r(u, v)]^2 \quad (1)$$

- 7 -

where p is the vector of viewing parameters, $I_v(u, v; p + \Delta p)$ is the virtual VB image rendered from the viewpoint $p + \Delta p$, u and v are the row and column indices, and I_r is the real RB image. Following the procedure of [5], it is shown that the Gauss-Newton parameter update Δp can be found as

$$\Delta p = H^{-1} \sum_{u,v} \left[\frac{\partial I}{\partial p} \right] [I_v(u, v; p) - I_r(u, v)] \quad (2)$$

where the Hessian H is approximated per Gauss-Newton as

$$H = \sum_{u,v} \left[\frac{\partial I_v}{\partial p} \right]^T_{u,v;p} \left[\frac{\partial I_v}{\partial p} \right]_{u,v;p} \quad (3)$$

10

where $\left[\frac{\partial I_v}{\partial p} \right]_{(u,v;p)}$ is a vector that gives the change in the intensity of a pixel (u, v) in a VB image I_v rendered at viewpoint p with respect to each of the components of the parameter vector Δp . $\left[\frac{\partial I_v}{\partial p} \right]_p$ can also be interpreted as a vector of steepest descent images, where each component of the vector is actually an image that describes the variation of the image intensities with respect a component of the parameter vector. Because the steepest descent images $\left[\frac{\partial I_v}{\partial p} \right]_p$ change at every viewpoint p , they, and the Hessian must be recomputed every iteration, leading to a very computationally costly algorithm.

To speed up the iteration, the inverse compositional algorithm was proposed [6]. Under this strategy, instead of moving the virtual viewpoint towards the real viewpoint using the parameter update, we instead move the real viewpoint toward the virtual viewpoint using the inverse of the parameter update. Since the computer obviously has no control over the location of the bronchoscope tip, this may seem to be an unfeasible strategy. However, using a depth-based warping, the RB image can be warped to simulate its appearance from other viewpoints. This strategy results in comparing a warped version of the real image to a stationary virtual image. Under this formulation, the objective function we seek to minimize is:

$$E = \sum_{u,v} [I_v(u, v; \Delta p) - I_r(W(u, v, Z_r; p))]^2$$

- 8 -

(4)

The warping function $W(\cdot)$ warps the image coordinates of the RB image I_r and hence warps the image itself. It is important also to note that the warp in this case is dependent on the depth map of the real image Z_r . Solving for the Gauss-Newton parameter update associated with 4 yields

$$\Delta p = H^{-1}|_{p=\vec{0}} \sum_{u,v} \left[\frac{\partial I}{\partial p} \right]_{u,v;\vec{0}} [I_r(W(u,v,Z_r;p)) - I_v(u,v)] \quad (5)$$

While this may seem to add unnecessary complexity and error to the problem, it actually serves to greatly speed the iteration and has the additional side benefit of eliminating the need to render arbitrary viewpoints on the fly if you instead have a collection of pre-rendered (or pre-captured) images and corresponding depth maps. The reason for this significant increase in speed is that the VB image and VB image gradients are always evaluated at $p = 0$, the reference viewing site, and as such allows all of the following operations to be pre-computed before iteration begins:

1. The known environment is sampled as a set of viewing sites.
2. Virtual images I_v are pre-rendered at each viewing site.
3. Virtual depth maps Z_v are computed at each site.
4. Steepest descent images $\frac{\partial I_v}{\partial p}$ are computed with respect to each of the viewing parameters in vector p .
5. The inverse Hessian H^{-1} is Gauss-Newton estimated from the steepest descent images $\frac{\partial I_v}{\partial p}$ via equation (14).

The iterative portion of the algorithm may then be carried out in the following steps:

1. Warp the real image from pose p to the nearest reference site.
2. Compute the error image $(I_r(W(u,v,Z_r;p)) - I_v(u,v;\vec{0}))$.
3. Compute the parameter update Δp via equation (5).

- 9 -

4. Find the new values of p by incrementing the old parameters by the inverse of the update $(\Delta p)^{-1}$.

These steps are illustrated in Figure 1. Ignoring the warp function, all the equations presented thus far are general and can apply equally well to 2D transformations, such as affine or homography, or 3D rotations. The focus is now narrowed, however, to the full 3D motion case with our choice of coordinate system and parameters. One may realize from inspection of the warps in (4) that the problem is defined in terms of several local coordinate systems as each reference view is defined to be at $p = \vec{0}_i$, yielding a different coordinate system for each viewing site used. It is, however, a trivial matter to relate each of these coordinate systems to a global coordinate frame in order to perform parameter conversions between frames. Therefore, given a camera pose with respect to the global camera frame, we can define our parameter vector as

$$p = [\theta_r \ \theta_p \ \theta_y \ t_x \ t_y \ t_z]^T \quad (6)$$

with three Euler rotation angles and three translations with respect to the nearest reference view.

With this parameterization, the warping $W(u,v,Z;p)$ is governed by the matrix equation

$$\begin{bmatrix} \frac{u'Z'}{f} \\ \frac{v'Z'}{f} \\ Z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} \frac{uZ}{f} \\ \frac{vZ}{f} \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (7)$$

where $\mathbf{R}$ is the rotation matrix defined by the Euler angles $(\theta_r, \theta_p, \theta_y)$, u and v are the columns and rows of the image, f is the focal length, and Z is the entry on the depth map Z corresponding to the point (u,v) . Here (u',v') gives the warped image coordinate of interest, and Z' gives the warped depth corresponding to that point. Note that in the problem statement, we assume only that the virtual depth map Z_v is known. However, when using the inverse compositional algorithm, the warp is applied to the real image I_r and the real depth map Z_r must first be calculated by warping the virtual depth map Z_v to the current estimated pose of the real camera via p . This can also be performed using (7) and then interpolating the resulting warped depth map onto the coordinate system of the

- 10 -

real image. In doing so, we are implicitly assuming that our estimate of p is relatively close to its actual value. If this is not the case, the parameter error can lead to large errors in the real depth map Z_r , and therefore large errors in the image warping. Under such circumstances, the forward gradient descent method governed by (1-2) may be better suited to the problem.

In order to apply the warping function, at each pixel coordinate (u,v) , with intensity $I(u,v)$ and depth $Z(u,v)$, a new coordinate (u',v') and depth $Z'(u',v')$ are found via (7). The original intensities and depths may then be mapped onto the new image array $I(u',v')$. Some special care must be taken when performing the warping. Firstly, the image difference in (4) requires that the coordinate locations be the same for both images. The resultant array must therefore be interpolated onto the same coordinate grid as the original arrays. Because of this interpolation, and because the depth-based warping may result in occlusion, it can be difficult to choose the proper intensity corresponding to an output pixel. This can be mitigated somewhat if the intensities corresponding to larger depths are discarded when they overlap with those of smaller depths.

Finally, we turn to the calculation of the steepest-descent images $\frac{\partial I}{\partial p}$. There are several ways to generate the steepest descent images. They may be generated numerically by taking the difference of the reference images warped to small positive and negative values of each parameter. They may also be generated analytically by expanding the derivative via the chain rule:

$$\frac{\partial I}{\partial p} = [\nabla_u I \quad \nabla_v I] J_p \quad (8)$$

where $\nabla_u I$ and $\nabla_v I$ are the image gradients with respect to the rows and columns of the image, and J_p is the Jacobian of the warped coordinates with respect to p and thus can be found by differentiating u' and v' from (7) with respect to each of the warp parameters and evaluating it at a particular current value of p . In the case of the inverse compositional algorithm, the image derivatives are always evaluated at $p = \vec{0}$ and thus the Jacobian is constant for each reference viewing site:

$$J_p = \begin{bmatrix} -v & -\frac{uv}{f} & -f - \frac{v^2}{f} & \frac{f}{Z} & 0 & -\frac{u}{Z} \\ -u & -f - \frac{u^2}{f} & -\frac{uv}{f} & 0 & \frac{f}{Z} & -\frac{v}{Z} \end{bmatrix} \quad (9)$$

- 11 -

We now have all the necessary information to calculate the iterated parameter update Δp . The final step is to invert this update, and compose it with the current estimate of p . The Euler angles can be found from the rotation matrix resulting from

$$\mathbf{R}' = \mathbf{R} \mathbf{R}_d^T \quad (10)$$

where $\mathbf{R}_d$ is the incremental rotation matrix associated with the rotation angles in Δp . The updated translations can be found from

$$\begin{pmatrix} t'_x \\ t'_y \\ t'_z \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} - \mathbf{R} \mathbf{R}_d^T \begin{pmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \end{pmatrix} \quad (11)$$

where Δt_i are the translation elements of the parameter update Δp .

In order to improve the performance when applying the above approach, several optimizing techniques are used. Operations performed on full-resolution images can be very computationally intensive. Therefore, a resolution pyramid is used wherein all
15 images, depth maps, and gradients are down-sampled, preferably by a factor of 4, at each level. As we are not particularly concerned with computation time regarding the precomputed virtual views and gradients, and most video capture hardware provides real-time hardware subsampling for the real image, the computational cost of this subsampling is inconsequential and provides much quicker iteration times.

20 When implementing the above registration algorithm using pyramid decomposition, the algorithm is begun at the lowest resolution level of the pyramid (experimental results in this paper were performed starting at level 3; i.e., a factor of 64 reduction in resolution) and run until a reasonable stopping criteria was met before proceeding to a higher resolution level. This pyramidal approach not only speeds
25 computation, it also serves to prevent convergence to local optima, because only the largest features are present in the highly subsampled images, while sharper features are introduced in higher resolution levels to aid in fine adjustment.

A second optimization that is used in practice is the use of the weighted normalized cross-correlation objective function

$$E = - \sum_{u,v} w_{u,v} \left[\frac{\mathbf{I}_v(W(u,v,Z;\Delta p)) - \mu_v}{\sigma_v} \right] \left[\frac{\mathbf{I}_r(W(u,v,Z;p)) - \mu_r}{\sigma_r} \right]$$

- 12 -

(12)

that allows images of different mean intensities and intensity ranges to be compared and
 5 also allows weighting of individual pixel values. It should be noted that in order to use
 this objective function under the inverse compositional algorithm, the weights must be
 constant and they must be chosen prior to the computation of the steepest descent images
 (i.e. they must be based off features of the virtual images). Taking advantage of the
 equivalence of normalized SSD and normalized cross-correlation, the update can be
 10 found as:

$$\Delta p = H^{-1}|_{p=0} \sum_{u,v} w_{u,v} \left[\frac{\partial \mathbf{I}}{\partial p} \right]_{p=0}^T [\bar{\mathbf{I}}_r(W(u,v,Z;p)) - \bar{\mathbf{I}}_v(u,v)] \quad (13)$$

where the Hessian in this case is

$$H = \sum_{u,v} w_{u,v} \left[\frac{\partial \mathbf{I}}{\partial p} \right]_{p=0}^T \left[\frac{\partial \mathbf{I}}{\partial p} \right]_{p=0}, \quad (14)$$

$\left[\frac{\partial \mathbf{I}}{\partial p} \right]$ is the set of mean-subtracted steepest descent images divided by the variance of the
 virtual image I_v , and $\bar{\mathbf{I}}_i$ are the normalized images.

EXAMPLES

20 To validate the algorithm, sample results for the virtual-to-real and virtual-to-
 virtual registration cases are given. In both of the cases outlined below, the virtual
 environment is a CT chest scan of a human patient designated h005. The airway surfaces
 were automatically generated using the methods of Kiraly et al. [10]. Airway centerlines
 were extracted using the methods of Swift et al. and the virtual viewing sites were chosen
 25 along these airway centerlines at intervals varying between 0.3mm and 1mm, with the
 viewing direction chosen parallel to the airway centerline [11]. Virtual images and depth
 maps were generated by an OpenGL renderer assuming a spot light source at the camera
 focal point, a field of view of 78.2 degrees and a 264x264 image size to match the
 calibration parameters of the bronchoscope camera.

Virtual-to-Real Registration

The virtual-to-real registration was performed using pyramid decomposition starting from level 3 and ending at level 1. To account for the difference in intensity characteristics between the imaging sources, the weighted normalized cross-correlation
 5 (12) was used as the objective function, with weights $w_{u,v}$ chosen as

$$w_{u,v} = 1 - I_u(u, v) \quad (15)$$

in order to emphasize dark areas, which tend to have more information in bronchoscopic
 10 video. The video frame, taken from a bronchoscopic procedure performed on h005 was first processed to remove the geometric barrel distortion from the lens to obtain the real image I_r . In the virtual-to-real registration case, it is difficult to give ground truth locations as the location of the scope tip is in practice unknown. Without external localization, the quality of a good registration is somewhat qualitative in nature. Figure 2
 15 shows a sample of the registration results, with edges from the virtual image overlaid on the unregistered and registered real views. The results show that the alignment is qualitatively very satisfying.

Virtual-to-Virtual Registration

In the virtual-to-virtual registration case, the “real” image is actually a rendering
 20 generated at a specified location in the airway, but with all depth information discarded. The algorithm uses pyramid decomposition starting from level 3 and ending at level 1, and the weighted SSD objective function was used where the weights $w_{u,v}$ were chosen as in (15) as before.

Figure 3 shows the “real” image I_r prior to registration, the virtual image I_v at the
 25 nearest reference site and the warped real image $I_r (W(u,v,Z;p))$ after registration is complete.

Viewpoint	X position (mm)	Y position (mm)	Z position (mm)	$\theta\alpha$ (deg)	$\theta\beta$ (deg)	$\theta\gamma$ (deg)
Initial	147.5	149.2	71.1	-20.2	-1.7	0
Reference Site	146.7	149.4	73.3	-7.3	5.1	-19.9
Registered	147.6	149.0	73.9	-20.9	1.2	-3.2

- 14 -

Ground Truth	147.1	148.9	73.8	-20.24	-1.8	-0.4
Error	0.6	0.1	0.5	-0.7	3.0	2.8

At least four different alternatives are available for registering the real and virtual sources in the case of bronchoscopy. These scenarios are outlined below:

1. Virtual-to-real registration: real-time or pre-recorded video images I_r from a bronchoscope at an unknown location are registered to a set of endoluminal CT renderings I_v and depth maps Z_v .
2. Virtual-to-virtual registration: an endoluminal rendering I_r with unknown location and with or without an associated depth map Z_r is registered to a set of endoluminal CT renderings I_v and depth maps Z_v .
3. Real-to-real registration: real-time video images I_r from an endoscope at an unknown location is registered to a set of previously recorded video images I_v with known or estimated depth maps Z_v .
4. Real-to-virtual registration: an endoluminal rendering I_r with unknown position and with or without an associated depth map Z_r is registered to a set of previously recorded video images I_v with known or estimated depth maps Z_v .

The application has far-reaching applications, particularly in the field of assisted endoscopy. The registration between a CT volume and real-time bronchoscopic video allows the fusion of information between the CT realm and the bronchoscope. This allows regions of interest (ROIs) defined only in the CT volume to be superimposed on the real video frame to assist the physician in navigating to these ROIs. Likewise, airway centerlines, branch labels and metric information such as distances to walls can be displayed on the video frame.

A natural extension of this concept is to other forms of endoscopy such as colonoscopy, where similar guidance information could be displayed on the registered colonoscopic image. Virtual-to-real registration can also be applied to pre-recorded endoscopic video, and opens the door to many post-processing options, such as mapping textural and color information available only in the endoscopic video onto the CT-derived surfaces to enable their visualization from viewpoints not available in the video alone.

- 15 -

An application of the real-to-real registration scenario that can be envisioned for this approach, is for aerial navigation. Satellite imagery, combined with topographic terrain information provides the known 3D environment, while real-time images from a mobile camera aboard an aircraft can be registered to this environment to give the aircraft's location and orientation without GPS or radar information. Similarly, this method also assists in ground-based robotic navigation within a known environment. Reference images and depth maps can be captured at known locations throughout the robot's working environment using a stereo camera setup, and a camera mounted on the robot can be registered to this set of images and depth maps.

REFERENCES

1. H. Minami, Y. Ando, F. Nomura, S. Sakai, and K. Shimokata, "Interbronchoscopist variability in the diagnosis of lung cancer by flexible bronchoscopy," *Chest* 105(2), pp. 1658–1662, June 1994.
2. I. Bricault, G. Ferretti, and P. Cinquin, "Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy," 17, pp. 703–714, Oct. 1998.
3. J. Helferty, *Image-Guided Endoscopy and its Application To Pulmonary Medicine*. PhD thesis, The Pennsylvania State University, 2002.
4. D. Deguchi, K. Mori, J. Hasegawa, J. Toriwaki, and H. Natori et al., "Camera motion tracking of real bronchoscope using epipolar geometry analysis and CT derived bronchoscopic images," *SPIE Medical Imaging 2002: Physiol. Func. from Multidim. Images* A Clough and C. Chen (ed.), v. 4683, pp. 30–41, 2002.
5. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.
6. S. Baker and I. Matthews, "Equivalence and efficiency of image alignment algorithms," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1, pp. 1090–1097, 2001.
7. H.-Y. Shum and R. Szeliski, "Panoramic image misaiming," *Technical Report MSR-TR-97-23*, Microsoft Research.
8. H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," *International Journal of Computer Vision* 16(1), pp. 63–84, 2000.
9. T. Okatani and K. Deguchi, "Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center," *Computer Vision and Image Understanding* 66, pp. 119–131, May 1997.
10. A. P. Kiraly, E. A. Hoffman, G. McLennan, W. E. Higgins, and J. M. Reinhardt, "3D human airway segmentation for virtual bronchoscopy," *SPIE Medical Imaging 2002: Physiology and Funct. from Multidim. Images*, A. V. Clough and C. T. Chen, eds. 4683, pp. 16–29, 2002.
11. R. Swift, A. Kiraly, A. Sherbondy, A. L. Austin, E. A. Hoffman, G. McLennan, and W. E. Higgins, "Automatic axes-generation for virtual bronchoscopic assessment of major airway obstructions," *Computerized Medical Imaging and Graphics* 26, pp. 103–118, March-April 2002.

We claim:

- 17 -

- 1 1. A method of registering an image, comprising the steps of:
2 providing a set of one or more reference images with depth maps;
3 registering the image to at least one of the reference images of the set using the
4 depth map for that reference image.
- 1 2. The method of claim 1, wherein:
2 the image is real and the set of reference images are real;
3 the image is real and the set of reference images are virtual;
4 the image is virtual and the set of reference images are real; or
5 the image is virtual and the set of reference images are virtual.
- 1 3. The method of claim 1, wherein the image or the set of reference images
2 are endoscopic.
- 1 4. The method of claim 1, wherein the registration occurs in real-time or near
2 real-time.
- 1 5. The method of claim 1, wherein at least one of the images in the set of
2 reference images further includes an image gradient.
- 1 6. The method of claim 1, wherein the images are cross-correlated.
- 1 7. The method of claim 1, wherein the image being registered is a video
2 frame.
- 1 8. The method of claim 1, wherein the image being registered is warped prior
2 to registration.

- 18 -

1 9. The method of claim 1, wherein the registration includes the step of
2 computing an image difference.

1 10. The method of claim 1, wherein the registration is updated using a Gauss-
2 Newton method.

1 11. The method of claim 1, wherein the set of reference images can be
2 updated before, during, or after registration.

1 12. The method of claim 1, wherein:
2 the set of reference images represents viewpoints with depth maps and image
3 gradients;

4 the image to be registered is derived from a video feed having a plurality of
5 consecutive frames; and the method includes the steps of:

6 a) warping a frame of the video to the nearest viewpoint of the reference source;

7 b) computing an image difference between the warped video frame and the
8 reference image;

9 c) updating the viewpoint using a Gauss-Newton parameter update; and

10 d) repeating steps a) through c) for each frame until the viewpoint converges or
11 the next video frame becomes available.

1 13. The method of claim 12, wherein one or more of the images in the set of
2 reference images can be updated before, during, or after registration.

1 14. The method of claim 12, wherein the video is real-time or near real-time.

1 15. The method of claim 12, including the step of guessing an initial
2 viewpoint.

- 19 -

1 16. The method of claim 12, wherein the final viewpoint gives an estimate of
2 the relative rotation and translation between the current frame of the video feed and the
3 reference source.

1 17. The method of claim 12, further including the step of using a resolution
2 pyramid wherein all images, depth maps, and gradients are down-sampled.

1 18. The method of claim 12, further including the step of using a weighted
2 normalized cross-correlation objective function that allows images of different mean
3 intensities and intensity ranges to be compared while allowing weighting of individual
4 pixel values.

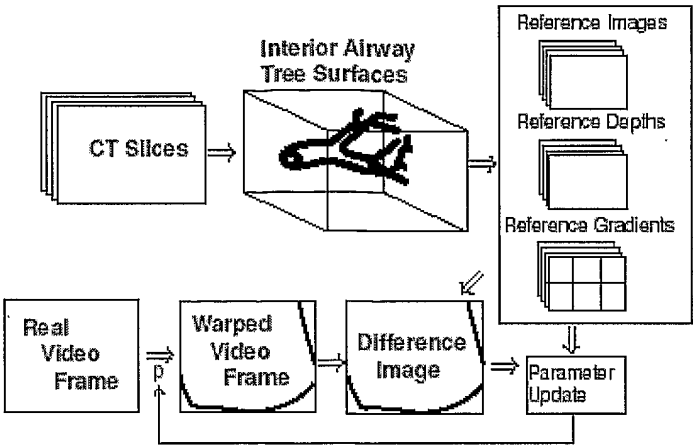


FIGURE 1

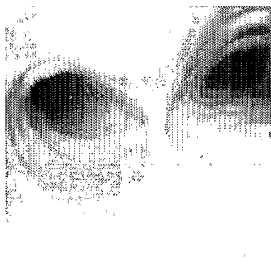


FIGURE 3A

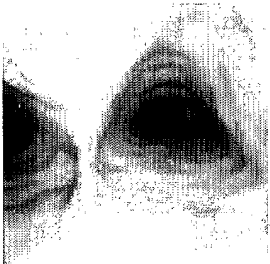


FIGURE 3B

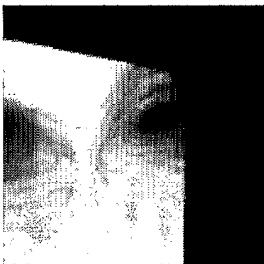


FIGURE 3C

2/2



FIGURE 2A

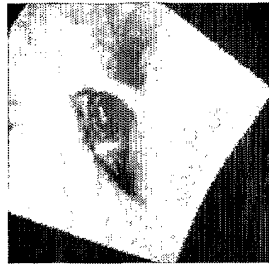


FIGURE 2B

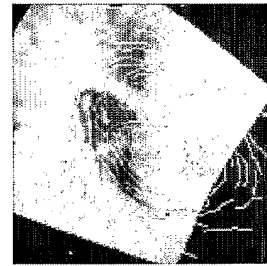


FIGURE 2C

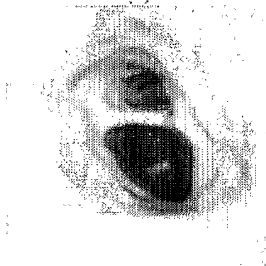


FIGURE 2D

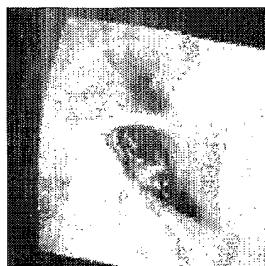


FIGURE 2E

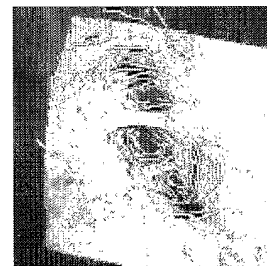


FIGURE 2F