pq-Space Based 2D/3D Registration for Endoscope Tracking

Fani Deligianni, Adrian Chung, and Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory
Imperial College London, United Kingdom
{f.deligianni,a.chung,g.z.yang}@imperial.ac.uk

Abstract. This paper presents a new *pq*-space based 2D/3D registration method for camera pose estimation for endoscope tracking. The proposed technique involves the extraction of surface normals for each pixel of the video images by using a linear local shape-from-shading algorithm derived from the unique camera/lighting constrains of the endoscopes. We illustrate how to use the derived *pq*-space distribution to match to that of the 3D tomographic model, and demonstrate the accuracy of the proposed method by using an electro-magnetic tracker and a specially constructed airway phantom. Comparison to existing intensity-based techniques has also been made, which highlights the major strength of the proposed method in its robustness against illumination and tissue deformation.

1 Introduction

With the maturity of minimal access surgery in recent years, there has been an increasing demand of patient specific simulation devices for both training and skills assessment. This is due to the fact that for minimal access surgery the complexity of the instrument controls, restricted vision and mobility, difficult hand-eye co-ordination, and lack of tactile perception are major obstacles in performing minimal access surgeries. They require a high degree of manual dexterity from the operator. Computer simulation provides an attractive means of performing certain aspects of this training, particularly the hand eye co-ordination and instrument control. For most of the current simulation systems, however, the degree of visual realism is severely limited. In endoscope simulations, most systems have used standard polygon rendering techniques with synthetic texture mapping. Although these can produce visually appealing results, they are not adaptable to either structure or appearance. Texture mapping is usually uniform throughout the whole simulation, and even in cases where special visual effects, such as polyps or inflammation, are provided, they are limited in both accuracy and adaptability. Natural objects, such as the colon or the bronchi show considerable diversity of shape and texture. The problem of generating realistic structure and surface properties has hindered the production of generic test case databases. These drawbacks highlight the importance of augmenting virtual endoscopic

R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2878, pp. 311–318, 2003.

views with patient specific endoscopic videos. Simulators require a geometric model of the world that the trainees explore. In current implementations, these are created artificially, but they could equally well be obtained from non-invasive tomographic imaging techniques. Recently, the feasibility of implementing such an idea for tracking camera motion and navigation planning has been investigated by a number of research groups [1, 2].

In order to match video endoscopic images to the geometry extracted from three-dimensional reconstructions of the bronchi, robust registration techniques have to be developed. This is a challenging problem as it implies the registration of 3D to 2D data from different sensors with certain degrees of deformation. Intensity based techniques based on mutual information [3, 4] or cross correlation can be problematic due to deformation and the difficulties in accurately modeling of illuminations [5, 6]. Endoscopic images are illuminated by a light source very close to the internal wall and are heavily affected by inter-reflections. This is further complicated by specular highlights caused by shiny mucus on the surface of the lumen. Although the alternative of using feature-based approaches may circumvent some of the problems mentioned above, especially in the handling of deformation [7-11], reliable feature extraction is proven to be difficult, especially when the surface of the lumen is textured. Photo-consistency is a promising technique but requires multiple calibrated cameras, [12].

The purpose of this paper is to introduce a novel *pq*-space based 2D/3D registration technique that exploits the unique geometrical constraints between the camera and the light source for endoscopic procedures. In the specific case of using perspective projection with a point light source near the camera, the use of intensity gradient can reduce the conventional shape-from-shading equations to a linear form, which suggests a local shape-from-shading algorithm that avoids the complication of changing surface albedos. We demonstrate how to use the derived *pq*-space distribution to match to that of the 3D tomographic model. The major advantages of this method are that it depends neither on the illumination of the 3D model, nor on feature extraction and matching. Furthermore, the temporal variation of the *p-q* distribution also permits the identification of localised deformation, which offers an effective way of excluding these areas from the registration process.

2 Methods

The basic process of the proposed technique comprises the following major steps: the extraction of surface normals for each pixel of the video images by using a linear local shape-from-shading algorithm derived from the unique camera/lighting constraints of the endoscopes; extraction of the *p-q* components of the 3D tomographic model by direct *z*-buffer differentiation; and the construction of a similarity measure based on angular deviations of the *p-q* vectors derived from 2D and 3D data sets.

2.1 Shapes-from-Shading in Endoscope

Shape from shading is a classical problem of computer vision that has been well established by the pioneering work of Horn [13], [14], [15]. It addresses the problem of extracting both surface and relative depth information from a single image. Horn in [14] relates the image irradiance to the scene radiance with the formula:

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

where, E is the image irradiance, L is the scene radiance, d is the diameter of camera's lens and f is the focal length. However, the above analysis is based on the assumption that the angle between the viewing vector $\hat{V}$ and the Z-axis, α, is negligible when the object size is small compared to its distance from the camera. In the case of endoscope images, both the camera and the light source are close to the object and the direction of the illuminating light coincides with the axis of the camera, thus no assumption can be made on α being negligible and lighting being uniform. Furthermore, the intensity of the image is also affected from the distance between the surface point and the light source. Rashid *et al* [16] and Okatani *et al* [17], modeled this dependency by adding one more factor, which was a monotonically decreasing function $f(r)$ between the surface point and the light source. Therefore, the image irradiance, E ,can be formulated as:

$$\mathrm{E}(x,y) = s_0 \cdot \rho(x,y) \cdot \cos(i) \cdot f(r) \tag{2}$$

where, s_0 is a constant related to the intrinsic parameters of the camera, ρ is the surface albedo and $\cos(i)$ is the angle between the incident light ray and the surface normal.

Since in the current problem of *p-q* space registration we are mainly concerned with surface normals, it can be proved that under the assumption that the light source is close to the camera, Equation (2) can be reduced to a linear form [16], such that

$$\begin{cases} A_1 \cdot p_0 + B_1 \cdot q_0 + C_1 = 0 \\ A_2 \cdot p_0 + B_2 \cdot q_0 + C_2 = 0 \end{cases} \tag{3}$$

where

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

In Equation (3) $R_x = E_x/E$ and $R_y = E_y/E$ are the normalised partial derivatives of the image intensities, E represents the intensity of the pixel under consideration,

and x_0 and y_0 are the normalized image coordinates. The partial derivatives can be estimated by applying standard intensity gradient operators.

2.2 Extracting p-q Components from the 3D Model

As for tomographic images, the extraction of the *p-q* components from the 3D model is relatively straightforward, since the exact surface representation is known. Since $p = \partial z/\partial x$ and $q = \partial z/\partial y$, differentiation of the *z*-buffer for the rendered 3D surface will result in the required *p-q* distribution, which also elegantly avoids the tasks of occlusion detection. The effect of perspective projection has been taken into account during the rendering stage.

2.3 Similarity Measure

One would expect to use the angle between the surface normals extracted from shape-from-shading and those from the 3D model for constructing a minimization problem for 2D/3D registration. This, however, is not possible because the *p-q* vectors in the shape-from-shading algorithm have been scaled. The similarity measure used in this paper depends on the *p-q* components alone and the cross correlation between the two *p-q* distribution are used.

Analytically, for each pixel of the video frame, a *p-q* vector corresponding to $\overline{n}_{img}(i,j) = [p_{i,j}, q_{i,j}]^T$ was calculated by using the linear shade-from-shading algorithm shown above. Similarly, for the current pose of the rendered 3D model, corresponding *p-q* vectors $\overline{n}_{3D}(i,j) = [p'_{i,j}, q'_{i,j}]^T$ for all rendered pixels were also extracted by differentiating the *z*-buffer. The similarity of the two images was determined by evaluating the dot product of corresponding *p-q* vectors:

$$\varphi(\overline{n}_{3D}(i,j), \overline{n}_{img}(i,j)) = \frac{\left\| \overline{n}_{3D}(i,j) \cdot \overline{n}_{img}(i,j) \right\|}{\left\| \overline{n}_{3D}(i,j) \right\| \cdot \left\| \overline{n}_{img}(i,j) \right\|} \tag{4}$$

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

$$\varphi_w(\overline{n}_{3D}(i,j), \overline{n}_{img}(i,j)) = \frac{\left\| \overline{n}_{3D}(i,j) \cdot \overline{n}_{img}(i,j) \right\|}{\left\| \overline{n}_{img}(i,j) \right\|} \tag{5}$$

By incorporating the mean angular differences and the associated standard deviations σ, the following similarity function can be derived

$$S = \frac{1}{\sum\sum(\varphi_w) \cdot \sum\sum\left(\left\| 1 - \sigma(\varphi_W) \right\| \cdot \left\| \overline{n}_{3D} \right\| \right)} \tag{6}$$

By minimizing Equation (6), the optimum pose of the camera for the video image can be derived. The reason for introducing a weighting factor for Equation (4) is due to the fact that *p-q* estimation from the 3D model is more accurate than that of the shape-from-shading algorithm. This is because it is not affected by surface textures, illumination conditions or surface reflective properties. The weighting factor therefore reduces the potential impact of erroneous *p-q* values from the shape-from-shading algorithm and improves the overall robustness of the registration process.

2.4 Tissue Deformation

With *p-q* space representation, the angle between the normal vectors before and after rigid body transformation will remain the same for every surface point. Local deformation can therefore be identified at surface points where the angle diverts from the mean angle of the whole 3D model. Localized inter-frame deformation can therefore be isolated and excluded for the pose estimation process. In this study, we used the *p-q* deformation map as a weighting factor during the registration process such that the weighting provided was inversely proportional to the amount of deformation detected.

2.5 Validation

In order to assess the accuracy of the proposed algorithm, an airway phantom made of silicon rubber and painted with acrylics was constructed. The inside face was coated with silicon-rubber mixed with acrylic to give it color/texture and left to cure in the open air. This gives the surface a specula finish that looks similar to the surface of the lumen. A real-time, six degrees-of-freedom Electro-Magnetic (EM) motion tracker (FASTRAK, Polhemus) was used to validate the 3D camera position and orientation. The EM-tracker has an accuracy of 0.762 *mm* RMS. The tomographic model of the phantom was scanned with a Siemens Somaton Volume Zoom four-channel multi-detector CT scanner with a slice thickness of 3 *mm* and in-plane resolution of 1 *mm*. The NTSC camera used is a CCD model CCN-2712YS, YTECH Design Co. Ltd. The intrinsic parameters of the camera have been found through calibration [18], while lens distortion has been ignored.

3 Results

Figure 1 demonstrates an example video frame of the bronchoscope phantom used to validate the proposed algorithm. The derived *p-q* vector distribution using the linear shape-from-shading algorithm is shown in Fig 1(b). It is evident that the derived *p-q* vectors are relatively immune to lighting changes, and these vectors were then used to estimated the 3D pose of the camera used to capture the video frame as shown in Fig 1(a). The effect of localised deformation on the *p-q* space representation is illustrated

in Fig (2), where (a) is the original video bronchoscope image and (b) is the derived *p-q* space deformation map. Figs (c) and (d) demonstrate the accuracy of the pose estimation with the traditional intensity based technique and the proposed *p-q* space registration with deformation weighting, respectively. Under deformation the intensity based technique has introduced significant error, despite the fact that the illumination conditions have been carefully adjusted.


a)

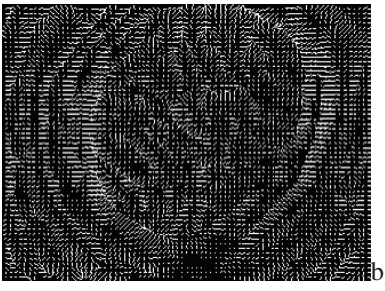
b)

Fig. 1. a) A sample bronchoscope video frame from the phantom used to reproduce the airway structures. b) The *p-q* vector distribution derived from the linear shape-from-shading algorithm by exploiting the unique camera/lighting constraints.

To assess the accuracy of the propose algorithm in tracking camera poses in 3D, Figs (3) and (4) compare the relative performance of the traditional intensity based technique and EM tracked poses against those from the new method. Since the tracked pose has six degrees-of-freedom, we used the distance traveled and inter-frame angular difference as a means of error assessment. As expected the intensity-based technique is highly sensitive to lighting condition changes, and with manual intensity adjustments, the convergence of this method is improved, as evident from the much reduced angular errors for all the image frames tested. The proposed *pq*-space registration, however, has much more consistent results which were very close to those measured by the EM tracker.

4 Discussions and Conclusion

In this paper, we have proposed a new *pq*-space based 2D/3D registration method for matching camera poses of bronchoscope videos. The results indicate that based on the *pq*-space and the 3D model, reliable bronchoscope tracking can be achieved. The main advantages of the method are that it is not affected by illumination conditions and does not require the extraction of feature vectors. The intrinsic robustness of the proposed technique is dependent upon the performance of the shape-from-shading method used, and the use of camera/lighting constraints of the bronchoscope greatly simplifies the 3D pose estimation of the camera. There are, however, a number improvements can be introduced for further enhance the accuracy of the proposed framework. For example, in this study the effect of mutual illumination, inter-reflectance and the specular components was not explicitly considered. Further investigation is needed to assess their relative impact to the accuracy of the algorithm.

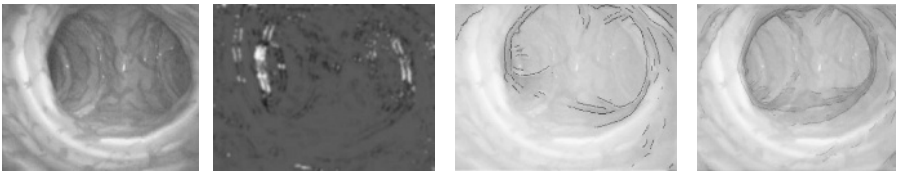

Fig. 2. a) A video frame from a deformed airway phantom, b) the associated *p-q* space deformation map where bright intensity signifies the amount of deformation detected. (c) The superimposed 3D rendered image with pose estimated from intensity-adjusted registration and *p-q* space registration with deformation weightings, respectively.

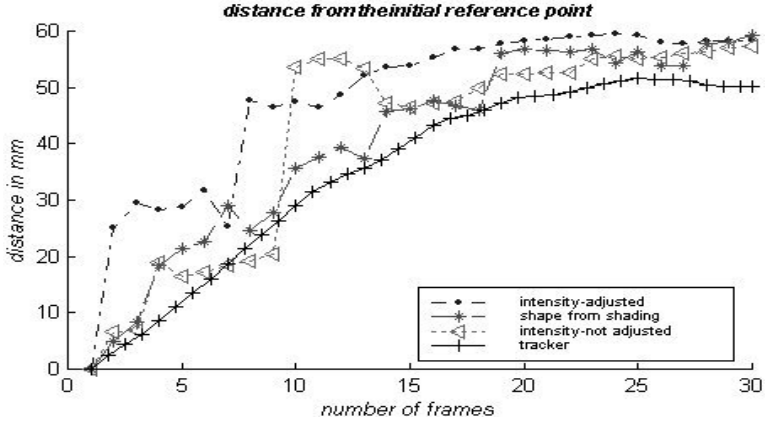

Fig. 3. Euclidean distance between the first and subsequent camera positions as measured by four different tracking techniques corresponding to the conventional intensity based 2D/3D registration with or without manual lighting adjustment, the EM tracker and the proposed pq space registration technique.

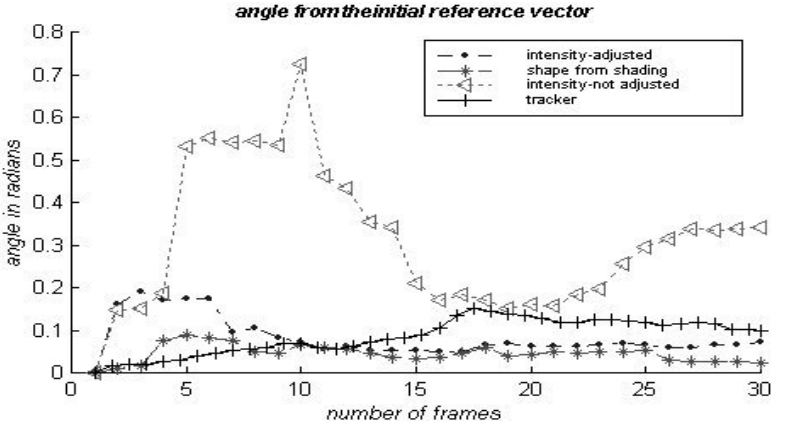

Fig. 4. Inter-frame angular difference at different time of the video sequence, as measured by

the four techniques described in Fig. 2.

References

1. Mori K., Suenaga Y., Toriwaki J., Hasegawa J., Kataa K., Takabatake H., Natori H.: A method for tracking camera motion of real endoscope by using virtual endoscopy system. Conference on Medical Imaging 2000: Physiology and Function from Multidimensional Images, Proceedings of SPIE, 3978, (2000), 122–133
2. Helferty J.P. and Higgins W. E.: Technique for Registering 3D CT Images to Endoscopic Video. IEEE International Conference, on Image Processing, 7–10 Oct., (2001), 893–896
3. Viola P.A.: Alignment by Maximization of Mutual Information. International Journal of Computer Vision, 24(2), (1997), 137–154
4. Studholme C., Hill D. L. G. and Hawkes D. J.: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition, 32(1), Dec (1998), 71–86
5. Likar B. and Pernus F.: A Hierarchical Approach to Elastic Registration Based on Mutual Information. Image and Vision Computing, 19(1-2), Jan (1999), 33–44
6. Tominaga S. and Tanaka N.: Estimating Reflection Parameters from a Single Color Image. IEEE Compute Graphics and Applications, 20(5), (2000), 58–66
7. Chui H. and Rangarajan A.: A new point-matching algorithm for non-rigid registration. Computer Vision and Image Understanding, 89(2-3), (2003), 114–141
8. Gold S., Rangarajan A., Lu C.P., Pappu S. and Mjolsness E.: New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence. Pattern Recognition 31(8), (1998), 1019–1031
9. David P., DeMenthon D., Duraiswami R. and Samet H.: SoftPOSIT: Simultaneous Pose and Correspondence. European Conference on Computer Vision 2002 (ECCV'02), Copenhagen, May (2002)
10. DeMenthon D. and Davis L.S.: Model-Based Object Pose in 25 Lines of Code. International Journal of Computer Vision, 15, (1995), 123–141
11. Pennec X., Ayache N. and Thirion J.P.: Landmark-based registration using features identified through differential geometry. Handbook of Medical Imaging – Processing and Analysis, I. Bankman Editor, Academic Press, Sept. (2000), 499–513
12. Clarkson M.J., Rueckert D., Hill D.L.G., Hawkes D.J.: Using Photo-Consistency to Register 2D Optical Images of the Human Face to a 3D Surface Model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(11), (2001), 1266–1280
13. Horn B. K. P.: Understanding Image Intensities. Artificial Intelligence, 8(2), April (1977), 201–231
14. Horn B. K. P.: Robot Vision. MIT Press, Cambridge, (1986)
15. Horn B. K. P. and Berthold K. P.: Shape from Shading. MIT Press, Cambridge, (1989)
16. Rashid H. U. and Burger P.: Differential algorithm for the determination of shape from shading using a point light source. Image and Vision Computing, 10(2), (1992), 119–127
17. Okatani T. and Deguchi K.: Shape Reconstruction from an Endoscope Image by shape from shading technique for a point light source at the projection centre. Computer Vision and Image Understanding, 66(2), May (1997), 119–131
18. Zhang Z.: A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), (2000), 1330–1334