

Research Article

Temporal Matching in Endoscopic Images for Remote-Controlled Robotic Surgery

Jia Gu,¹ Rolf Wolters,² and Ulf Gustafsson²

¹ Key Laboratory for Biomedical Informatics and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518067, China

² STI Medical Systems, 733 Bishop Street no. 3100, Honolulu, HI 96813, USA

Correspondence should be addressed to Jia Gu, gujia1981@hotmail.com

Received 23 October 2008; Accepted 4 February 2009

Recommended by Fei Hu

Temporal matching is applied in the frame of the formation of high-level entities in remote-controlled robotic surgery. The objective is to track tumor boundaries over time to improve the segmentation stage in each image of the sequence to facilitate the tracking and localization of the tumor. It makes use of an attributed string matching technique to find the correspondence between tumor boundaries over time. Relationships are then exploited to reconstitute the tumor boundaries and remove the inconsistencies coming from the detection errors. Input data are free form shapes of different length representing the tumor boundary, extracted at a previous stage.

Copyright © 2009 Jia Gu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

With recent advances in telemedicine technology, it is now possible to perform closed-abdomen surgery on a moving organ of a remote patient using robotic instrument to minimize patient trauma and certain side effects. For robotic assisted surgery, dexterity is enhanced by microprocessor controlled mechanical wrists, which allow motion scaling for reducing gross hand movements and the performance of microscale tasks that are otherwise not possible. So far, two commercially available master-slave manipulator devices are specifically designed for robotic cardiac surgery [1]. Both systems improve the ergonomics of laparoscopic surgery and provide high dexterity, precision, and 3D visualization of the operating field. One of the significant challenges of the in-vivo surgery is the destabilization introduced by respiratory motion, thus severely affecting precise instrument-tissue interactions and the execution of complex grafts. The other one is the precise localization of the tumor boundary, while most of the automatic image-guided algorithms yield incomplete, discontinuous, imprecise detection results. And the two above challenges always trigger each other and form a chick-egg problem. Thus in this paper, we intend to solve this problem by a novel high-level entities formation method.

Many existing methods have been developed to extract structural features [2–4] but, at the same time, included some errors (incomplete detection leading to discontinuities, imprecise localization of contours) that increased the difficulties of the further processing like tracking or 3D reconstruction. One way to overcome these difficulties has been, these last years, to incorporate some interactivity [5]. The contour detection makes use of an interactive procedure for the first image, and is then used as initial model of a snake to extract the contours in the other images of the sequence. Thus in [5], for each model curve, control points were interactively defined as the anchor points of an active contour modeling. Initial control points were interactively selected and the 2D contours represented as a B-spline. The active contour model was then applied to fit each curves of the selected contours. In [5], the segmentation of the first image made use of an interactive tool based on the mouse moving. The operator selected an initial point inside the image, and uncoiled a thread with the mouse, roughly following the boundary. An optimal path research procedure based on a dynamic programming algorithm and using a maximum multiscale response map for constraining the path search was performed between the initial point and the current position of the mouse. The contour was thus piecewise built, by successively moving the mouse along the boundary.

We focus here on the intermediate level between the segmentation and the reconstruction stages. We look for improving the automatic segmentation result in each image of the sequence by reconstituting high level entities, that is, the shape of tumors, in order to facilitate the tracking and localization of the tumor. A first intramimage analysis was previously performed to detect contours in each image of the sequence. It aimed at forming correspondences from previously extracted contours by using a string matching process [6]. The present work represents a continuation of the previous one. It aims at tracking boundaries over time to establish correspondences in the sequence. These correspondences will then be used to spatially prolong the segments in order to reconstitute the boundaries and get a symbolic description of the structures in the 2D space. We present in this paper the temporal matching method, which is based on an attributed string matching technique. The outline of this paper is as follows. Section 2 describes the temporal procedure. Section 3 provides results on real data. Some prospects for future work are then proposed in conclusion.

2. Temporal Matching

Attributed string matching techniques have been broadly applied in pattern recognition with the aim to look for specific shapes, comparing unknown figures occurring in the input image with a set of prototypic models in a classification purpose or emphasizing “landmarks” subsets common to a set of features. More recently, interest was found in image analysis in particular in shape recognition [7–9] and registration in the frame, for instance, of the 3D reconstruction from stereo views [10].

2.1. String Matching Concepts. String matching techniques allow measuring the similarity degree between two strings of symbols.

Let $S_n \in \Sigma^{l_n}$ be an arbitrary string of length $l_n = |S_n|$ defined on a discrete alphabet Σ . We denote s_n^i the i th symbol of S_n for $i \in \{1 \dots |S_n|\} \in \Sigma$ such as $S_n^l = s_n^0 s_n^1 \dots s_n^{l_n-1}$. An edit distance, based on edit operations (insertion, deletion, and substitution of symbols), is then defined as the minimum cost operation sequence over all possible sequences of operations O ($O = o_1 o_2 \dots o_m$) with o_k equals to one of the three above cited edit operations. It makes use of a cost matrix, which specifies the costs $\gamma(o_k)$ of individual elementary operations for all combination of symbols. The total cost of the transformation is then calculated by summing up the elementary costs applied to each operation in the sequence. The distance associated with the minimum cost operation sequence is given by the relation:

$$\delta(S_n^l, S_p^l) = \left\{ \frac{\text{Min}_O(\Gamma_{S_n, S_p}(O))}{O} : S_n^l \rightarrow S_p^l \right\}, \quad (1)$$

with $\Gamma_{S_n, S_p}(O) = \sum_{v=1}^m \gamma(o_v)$.

Its computation can be carried out using optimization techniques. Among these ones, the classical dynamic programming, initially designed by Wagner & Fisher [6], has largely been used to calculate the edit distance.

The process goes through the edit matrix computation $D(i, j)$, $1 \leq i \leq l_n$, $1 \leq j \leq l_p$ using the following recursive relation:

$$\begin{aligned} D(i, j) = \min \{ & (D(i-1, j-1) + \gamma(s_n^i \rightarrow s_p^j)), \\ & (D(i-1, j) + \gamma(s_n^i \rightarrow \lambda)), \\ & (D(i, j-1) + \gamma(\lambda \rightarrow s_p^j)) \}, \end{aligned} \quad (2)$$

where λ denote the empty symbol and $\gamma(s_n^i \rightarrow s_p^j)$, $\gamma(\lambda \rightarrow s_p^j)$, $\gamma(s_n^i \rightarrow \lambda)$, respectively, characterize the substitution, insertion, and deletion costs. The least cost operation sequence is obtained through a search of a least cost path in the graph. The result is a sequence of ordered pairs of positions in S_n^l and S_p^l corresponding to substitutions. With a least-cost trace from S_n^l to S_p^l , $T(S_n^l, S_p^l)$ is hence defined as the set of ordered pairs of integer (i, j) corresponding to the indices of the paired symbols between the two strings.

2.2. Shape Representation and Cost Function. Let $L_{t,n}$ and $L_{t+d,m}$ two lines to be matched with $(t, t+d)$ characterizing the image number in the sequence and (n, m) the line numbers in each image. Elements of $L_{t,n}$ and $L_{t+d,m}$ were, respectively, noted $p_{t,n}^i$ and $p_{t+d,m}^j$, i and j being the element indexes for each line. Four attributes were associated to each pixel $p_{t,n}^i$: the intensity $I(p_{t,n}^i)$, the Euclidean location $\vec{d}_{t,n}^i$, the tangent $\vec{t}_{t,n}^i$, and the local curvature $\kappa_{t,n}^i$. Each line was thus expressed as a sequence $S_{t,n}^l$ of symbols (l_n being the length of the string), each symbol corresponding to a set of attributes, we formalized as a vector $s_{t,n}^i = (I(p_{t,n}^i), \vec{d}_{t,n}^i, \vec{t}_{t,n}^i, \kappa_{t,n}^i)$. We designed a multiparametric cost function between sets of two vectors $(s_{t,n}^i, s_{t+d,m}^j)$ based on perceptual criteria: parallelism, proximity, curvature, and intensity similarity.

Deletion and insertion costs were set to 0.5, while the substitution cost, defined by the function $f(s_{t,n}^i, s_{t+d,m}^j)$, was expressed on the $[0, 1]$ interval:

$$\begin{aligned} f(s_{t,n}^i, s_{t+d,m}^j) = & \alpha \cdot f_i(I(p_{t,n}^i), I(p_{t+d,m}^j)) + \beta \cdot f_p(\vec{t}_{t,n}^i, \vec{t}_{t+d,m}^j) \\ & + \lambda \cdot f_\kappa(\kappa_{t,n}^i, \kappa_{t+d,m}^j) + \gamma \cdot f_d(p_{t,n}^i, p_{t+d,m}^j), \end{aligned} \quad (3)$$

where $\alpha, \beta, \lambda, \gamma$ were weighting coefficients such that $\alpha + \beta + \lambda + \gamma = 1$.

Each function was defined as an exponential of the form $f(\cdot) = e^{-(\Delta_{\max} - |\Delta|)/|\Delta|}$, where Δ_{\max} characterized the maximum permitted deviation between two symbols for a given criterion and Δ , the difference measured between the two considered symbols. Δ_{\max} definition depends on the limit, we set up for each parameter deviation so that $0 \leq f(\cdot) \leq 1$, value 0 depicting a good shape similarity.

2.3. Tumor Boundary Formation. This stage aimed at establishing the temporal correspondences between segments of contours along the sequence. The attributed string matching was applied on sets of pairwise images, all possible temporal associations being taken into consideration.

For each contours $L_{t,n}$ in image I_t , a contours neighborhood $V_{(t,t+d),n}$ was built in image I_{t+d} such as

$$V_{(t,t+d),n} = \left\{ \frac{L_{t+d,m}}{m} = 1 \cdots N_n, \text{distance}(L_{t,n}, L_{t+d,m}) < \Delta d_{\max} \right\}, \quad (4)$$

where Δd_{\max} sets the upper limit for the contour displacement between two instants t and $t + d$ and N_n denotes the cardinal of $V_{(t,t+d),n}$.

All the possible combinations of contours, inside this neighborhood, were then considered for the pairing. With the segment running direction for the matching being not prior known, each extremity of a segment was numbered and two runs were considered according to these extremities. Two associations were then considered among the four possible combinations $\{(L_{t,n}^1, L_{t+d,m}^1), (L_{t,n}^1, L_{t+d,m}^2)\}$, exponent 1 and 2 representing the extremity number. Attributed strings $(S_{t,n}^1, S_{t+d,m}^1)$ were then assigned to each two lines $(L_{t,n}, L_{t+d,m})$. The attributed string matching process was applied for each two pairs of lines and the right running direction corresponded to the best association among the two, when choosing the minimal edit distance as selection criterion.

Because multiple assignments can occur in some situations (when several contours are close to each other or when spurious segments are present), two criteria were thus defined in order to keep only the one-to-one matching

- (1) A threshold on the normalized edit distance aimed to remove all the inconsistent associations: matched lines whose edit distance was greater than this threshold were cancelled.
- (2) An overlapping criterion was then considered to remove redundant associations. These associations are considered as right if there is no overlapping between the matched sets. On the contrary, these associations are wrong and have to be filtered. Thus when several candidate curves were associated to the same curve with some overlapping between these associations, the matched lines were sorted out on the edit distance value and the established trace between the paired symbols. Considering for each line $L_{t,n}$, the matched line list by decreasing edit distance values, if the trace between the matched pixels intersected a curve, the pair of lines having the highest edit distance value was removed.

3. Result and Discussion

The matching procedure was applied to endoscopic images acquired from standard optical colonoscope (from our collaborator: Stanford University), and the preliminary results

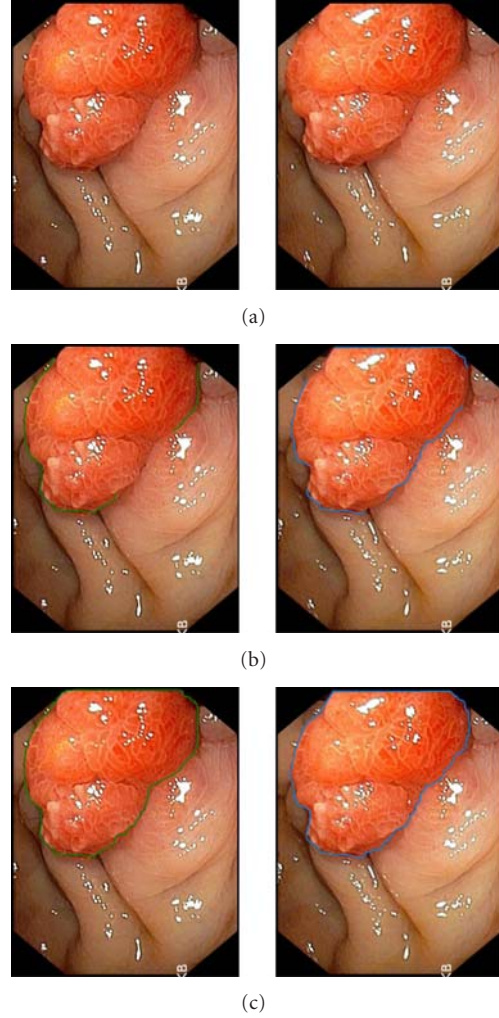


FIGURE 1: String matching process and boundary formation on two successive images of a sequence: (a) original acquired images from endoscope, (b) initial contour extraction, (c) boundary formation: the discontinuities have been reconstituted.

can be found in Figure 1. Whatever the detection operator and the decision rule, it must be accepted that no exact result can be derived. This statement, which applies to edge detection too, was the first motivation of this work. In a first stage, edges were grouped using spatial properties to construct coherent pieces [6]. This process was operated in an intramode. The next stage, that is, the temporal matching, takes advantage of the motion correspondence to solve spatial ambiguities and improve the segmentation results in presence of occlusion or discontinuities on the images. The matching process was led over time along the sequence. It was pairwise applied on the contour and allows at the same time to get a first approximation of the motion between the structures. The parameters setting resulted from a learning stage, which allowed studying the behavior of the algorithm according to the parameters choice and combination. This study showed that the results were not sensitive to the choice of the parameters. The best result was obtained with the following setting: $\alpha = 20\%$, $\beta = 15\%$,

$\lambda = 30\%$, $\gamma = 35\%$, $\Delta I_{\max} = 20$ (maximum intensity difference), $\Delta\theta_{\max} = \pi/4$ (angular difference between the tangents), $\Delta\kappa_{\max} = 0.5$ (maximum curvature with $|\Delta\kappa| = |\kappa_{t,n}^i - \kappa_{t+d,m}^j|/8$), $\Delta d_{\max} = 30$ (maximum distance). The distance parameter role is to penalize the structure displacements beyond this value. For these parameters, we computed the mean cross-correlation, the mean distance, and the mean normalized edit distance over the set of matched lines. We obtained the respective values: 0.963, 3.121, and 0.275. The branches formation was then performed over the sequence, based on the matching process results.

4. Conclusion

In this paper, we made use of an attributed string matching technique for the formation of high-level entities like boundaries of tumors in endoscopic images for remote-controlled robotic surgery. The advantage of the method is that few parameters have to be set to track the contours over time. The established temporal relationships were exploited to spatially locate the contours and reconstitute the tumor boundaries. These results will be further used to complete the tumor boundaries that were previously performed on each image of the sequence [6]. This work represents a contribution to the difficult problem of the tumor tracking and localization in robot-assisted minimal invasive surgery.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. The research was partially supported by US Army Medical Research and Material Command under Contract no. W81XWH-07-C-0072. The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation. In the conduct of research where humans are the subjects, the investigators adhered to the policies regarding the protection of human subjects as prescribed by Code of Federal Regulations Title 45, Volume 1, Part 46; Title 32, Chapter 1, Part 219; Title 21, Chapter 1, Part 50 (Protection of Human Subjects).

References

- [1] G. H. Ballantyne, "Robotic surgery, telerobotic surgery, telepresence, and telementoring," *Surgical Endoscopy*, vol. 16, no. 10, pp. 1389–1402, 2002.
- [2] C. Toumoulin, R. Collorec, and J. L. Coatrieux, "Vascular network segmentation in subtraction angiograms: a comparative study," *Medical Informatics*, vol. 15, no. 4, pp. 333–341, 1990.
- [3] Y. Sun, "Automated identification of vessel contours in coronary arteriograms by an adaptive tracking algorithm," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 78–88, 1989.
- [4] J. L. Coatrieux and C. Toumoulin, "Computational vision and structural modeling in cardiac vascular network reconstruction," in *Medical Imaging Processing: From Pixel to Structure*, Y. Goussard, Ed., pp. 83–109, L'École Polytechnique de Montréal, Montreal, Canada, 1998.
- [5] F. Mourgues, F. Devernay, G. Malandain, and È. Coste-Manière, "3D+t modeling of coronary artery tree from standard non simultaneous angiograms," in *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, vol. 2208 of *Lecture Notes in Computer Science*, pp. 1320–1322, Springer, Utrecht, The Netherlands, October 2001.
- [6] C. Toumoulin, J. Brieva, J.-J. Bellanger, and H. Shu, "String matching techniques for high-level primitive formation in 2-D vascular imaging," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 4, pp. 291–301, 2003.
- [7] M. Maes, "Polygonal shape recognition using string-matching techniques," *Pattern Recognition*, vol. 24, no. 5, pp. 433–440, 1991.
- [8] Y.-T. Tsay and W.-H. Tsai, "Attributed string matching by split-and-merge for on-line Chinese character recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 180–185, 1993.
- [9] S. W. Chen, S. T. Tung, C. Y. Fang, S. Cherng, and A. K. Jain, "Extended attributed string matching for shape recognition," *Computer Vision and Image Understanding*, vol. 70, no. 1, pp. 36–50, 1998.
- [10] R. Myers, R. C. Wilson, and E. R. Hancock, "Bayesian graph edit distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 628–635, 2000.