Automatic Detection of Region of Interest Based on Object Tracking in Neurosurgical Video

Article *in* Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference · February 2005

DOI: 10.1109/IEMBS.2005.1615931 · Source: PubMed

6 authors, including:

Bing Liu

81 PUBLICATIONS **907** CITATIONS

SEE PROFILE

Mingui Sun

University of Pittsburgh

339 PUBLICATIONS **2,437** CITATIONS

SEE PROFILE

Qiang Liu

Nanjing University of Posts and Telecommunications

283 PUBLICATIONS **3,297** CITATIONS

SEE PROFILE

Amin Kassam

Aurora Health Care

502 PUBLICATIONS **15,409** CITATIONS

SEE PROFILE

Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Automatic Detection of Region of Interest Based on Object Tracking in Neurosurgical Video

Bing Liu [1,2], Mingui Sun[2,3,4], Qiang Liu[2,4], Amin Kassam[2], Ching-Chung Li[4], and Robert J. Sclabassi[2,3,4]
[1]Department of Information Security Engineering, School of Software
Yunnan University, Kunming, Yunan, 650091, China, Email: bliu@demensi.com
Laboratory for Computational Neuroscience
Departments of [2]Neurosurgery, [3]Bioengineering, and [4]Electrical Engineering
University of Pittsburgh, Pittsburgh, PA 15261, Email: lcn@neuronet.pitt.edu

Abstract - **Automatic detection of region of interest (ROIs) in a complex image or video, such as an angiogram or endoscopic neurosurgery video, is a critical task in many medical image and video processing applications. In this paper, we present a new method that addresses several challenges in automatic detection of ROI of neurosurgical video for ROI coding which is used for neurophysiological intraoperative monitoring (IOM) system. This method is based on an object tracking technique with multivariate density estimation theory, combined with the shape information of the object. By defining the ROIs for neurosurgical video, this method produces a smooth and convex emphasis region within which surgical procedures are performed. A large bandwidth budget is assigned within the ROI to archive high-fidelity Internet transmission. Outside the ROI, a small bandwidth budget is allocated to efficiently utilize the bandwidth resource . We believe this method also can be used to image-guiduance surgery (IGS) systems to track the positions of surgical instruments in the physical space occupied by the patient after some improvement.**

Index Terms - Region of Interest (ROI), Object Coding, Intraoperative Monitoring (IOM), Object Tracking.

I. INTRODUCTION

Automatic detection of ROI in a complex image or video like angiogram or endoscopic neurosurgery video, is an important task in many image and video processing applications such as image-guide surgery system [1], real-time patient monitoring system [2], and object-based video compression [3].

In telemedical applications, object-based video coding is highly useful because it produces a good perceptual quality in a specified region, i.e., a region of interest (ROI), without requiring an excessive bandwidth. By using a specially designed video encoder, such as the wavelet-based encoder combined with the MPEG-4 standard, the ROI can be coded with more bits to obtain a much higher quality than that of the non-ROI which is coded with fewer bits. This method has been successfully utilized in a prototype intraoperative monitoring system [4].

Various methods have been reported in the literature for ROI extraction. In [5], ROIs were determined in a compressed domain based on the properties of the Human Visual System (HVS). The major factors that decide ROIs are its motion, spatial frequency, edge and color content. ROI detections under different resolutions were addressed in [6]. This procedure included feature selection, region merging or splitting, determination and fusing. Reference [7] presents a scheme to automatically generate the ROI mask during the encoding process of JPEG 2000. The ROI mask is determined by using the information of significant states of different bit-planes and the Rate-Distortion optimization procedure. A wavelet based ROI encoder for angiogram video sequences was discussed in [3] based on the motion differences between arteries and background. In all these reported methods, an object in the image or video frame is the ROI. However, in our case where a surgical video is utilized for remote intraoperative monitoring, the ROI is a field within which surgical procedures take place. In general, it is not a single object. Our ROI involves both surgical instruments and a surrounding area, which consists of various types of biological tissue and deformable substances, such as fluids.

In this paper, we present a new method that addresses several challenges in automatic detection of ROI in neurosurgical video. The edge information of a selected instrument in the video is computed and utilized as an input to a histogram-based tracking algorithm which provides a critical location, e.g., the tip of the surgical instrument. This location is then employed to define an ROI which has a cornerless, smooth, and convex shape, facilitating observation at theremote site. The ROI is updated continuously to track the motion of surgical instrument. This paper is structured as follow. Section II discusses several important issues of ROI in neurosurgical video; Section III describes an approach for ROI detection; Section IV presents ROI updating method; and Section V concludes this work.

II. DEFINITION OF ROI IN NEUROSURGICAL VIDEO

Fig. 1 shows two typical frames of neurosurgical video, both are 240×352 in size. In this type of video, the region

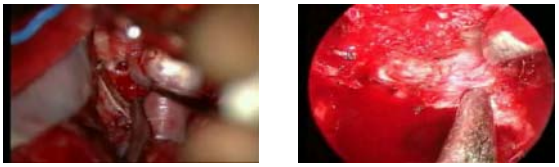

Fig. 1 Two frames of typical neurosurgery videos

at the tip of the surgical instrument is usually the focal point of observation, while the surrounding tissues close to the tip are also important for the purpose of intraoperative monitoring. The outer-regions in the frame, although less important and more tolerable for a reduced resolution, are utilized mainly for reference to both the geometry and the depth of the surgical landscape. Based on these special features, we define the ROI as a region which always includes the tip of a selected instrument. This instrument is usually a suction held by the left hand of the neurosurgeon who utilizes it to remove fluids during surgery.

At the beginning of video data transmission, our monitoring system asks the observer at the remote site to specify a circle on the video screen representing his/her interest. The center location and the radius of this circle is transmitted to the surgical site and utilized as the main component of the ROI for object-based video coding. However, the instrument may occasionally move out of the circle. When this happens and is detected by the tracking algorithm, the circle is automatically augmented to include this extra region into which the instrument has moved.

III. ROI Detection Based On Object Tracking

The basic concept of the tracking method in this study was previously reported by Comaniciu [8]. The method was mainly for applications involving non-rigid objects. While it is unpractical to adopt a circle or ellipse to match the surgery instruments such as suction in our case without rotation, since the instruments in the surgical video are rigid, rod-like, and slowly moving. We use a rectangle which is suitable for most of the instruments to track the object without complexity. At the same time, the edge information of the surgery instrument is used to adjust the tracked position.

A. Mean Shift Analysis

Given $\{x_i\}_{i=1\cdots n}$ is an arbitrary set of n points in the d dimensional Euclidean space R^d. Its estimation of probability density function (PDF) can be expressed by multivariate kernel density estimation method with the kernel $K(x)$ and a window bandwidth h as (1) [9].

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{x-x_i}{h}). \quad (1)$$

Its density gradient function $\nabla\hat{f}(x)$ also can be estimated with the given data as (2).

$$\nabla\hat{f}(x) = C_f\sum_{i=1}^{n}(x-x_i)K'(\frac{x-x_i}{h}), \quad (2)$$

where C_f is a constant. If the radially symmetric kernel $K(x) = ck(\|x_i\|^2)$ is used, then (2) can be denoted as (3),

$$\nabla\hat{f}(x) = C_g g(\left\|\frac{x-x_i}{h}\right\|^2)\left[\frac{\sum_{i=1}^{n} x_i g(\left\|\frac{x-x_i}{h}\right\|)^2}{\sum_{i=1}^{n} g(\left\|\frac{x-x_i}{h}\right\|)^2} - x\right]. \quad (3)$$

Where $g(x) = -k'(x)$. The last term is called the mean shift which always points towards the steepest ascent of direction of the density function. Based on this feature, the mean shift is used for the data cluster and object tracking.

B. Basic Tracking Method

The rectangle within which the instrument is selected interactively by the observer is in the first frame of a video along with the circle discussed previously. Within the rectangle, the instrument is represented in a featured space, which usually is color-scale space, by its estimated PDF $\hat{q}$. In the local window with n_h points of the subsequent frame, a target candidate locates at y and can be represented by the estimated PDF $\hat{p}(y)$. A set of discrete densities (m -bin histograms) is utilized to define $\hat{q}$. As (4) and (5),

$$\hat{q}_u = c_q\sum_{i=1}^{n} k(\|x_i\|^2)\delta[b(x_i)-u], \quad (4)$$

$$\hat{p}_u(y) = c_p\sum_{i=1}^{n_h} k(\left\|\frac{y-x_i}{h}\right\|^2)\delta[b(x_i)-u]. \quad (5)$$

Where, $u = 1,\cdots,m$. Constants c_q and c_p in (4) and (5) are used to normalize $\hat{q}_u$ and $\hat{p}_u(y)$ separately. Function $b(x)$ is the color index at position x.

Then, a similarity function such as (6) between $\hat{q}$ and $\hat{p}$ is defined to look for the new object position y in the tracking window which centers at the object position in the previous frame but is located in the subsequent frame.

$$\hat{\rho}(y) \equiv \rho[\hat{p}(y),\hat{q}] = \sum_{u=1}^{m}\sqrt{\hat{p}_u(y)\hat{q}_u}. \quad (6)$$

The new position y is computed by the mean shift which is discussed above, where the gradient of $\hat{\rho}(y)$ is used to seek the y to maximize the similarity function $\hat{\rho}(y)$ as (7),

$$\nabla\hat{\rho}(y) = c_\rho\sum_{i=1}^{n_h} g(\left\|\frac{y-x_i}{h}\right\|^2)\left[\frac{\sum_{i=1}^{n_h} w_i x_i g(\left\|\frac{y-x_i}{h}\right\|^2)}{\sum_{i=1}^{n_h} w_i g(\left\|\frac{y-x_i}{h}\right\|^2)} - y\right]. \quad (7)$$

Where $w_i = \sum_{u=1}^{m}\sqrt{\frac{q_u}{p_u(y_0)}}\delta[b(x_i^*)-u]$ is a weight for the given data. The mean shift term which is in the parenthesis of (7) gives the new position y in the next frame. Where,

$$y = \frac{\sum_{i=1}^{n_h} w_i x_i g(\left\|\frac{y-x_i}{h}\right\|^2)}{\sum_{i=1}^{n_h} w_i g(\left\|\frac{y-x_i}{h}\right\|^2)}. \quad (8)$$

C. *Finding the Tracked Position with Edge Detection*

As above discussed, since the instrument constantly changes its orientation, we utilize the edge information in the frame to obtain a rotation angle which updates the tracking position. First, an edge image in the current frame is computed within a local window around the object in the previous frame. Then, only the longest and the strongest edge line detected by using the Hough transform is saved and its angle is calculated. Fig. 2 gives the edge images (left column) and the detected lines (right column) using canny method and Hough transform, where only the strongest lines are used for the following processing.

In order to reduce the effect of the edge detection error and avoid unrealistic, drastic angle changes, every angle α_{i+1} obtained from the current frame is thresholded against the previous few frames, for example 5 frames. And the final angle is weighted average by w with previous angles as (9). By this way, the possible error angle and drastic changes are eliminated.

$$\alpha_{i+1} = \begin{cases} \alpha_{i+1}, & if \ \left|\alpha_{i+1} - \sum \alpha_i w_i\right| < s\left|\sum \alpha_i w_i\right| \\ \sum \alpha_i w_i, & otherwise \end{cases} \quad (9)$$

Where, s is a scale factor to adjust the threshold, $s = 0.6$ is used in our experiment. And $\{w_i\}$ is a convex weight vector and $\sum w_i = 1$. As an example, we consider 5 frames for the angle and $\{w_i\} = \{0.05, 0.13, 0.21, 0.29, 0.32\}$ was selected to be the weight vector. Then, the angle is feed back to the tracking step to adjust the rectangle which is used to match the instrument to find the accurate tip position. Fig. 3 shows the detected ROI in the 1st, 10th, 80th and 120th frame of a neurosurgery video, where, the rectangle in the frame is the matching shape for the instrument which is suction in this video. The circle, which centers at the tip of the instrument, denotes the ROI.

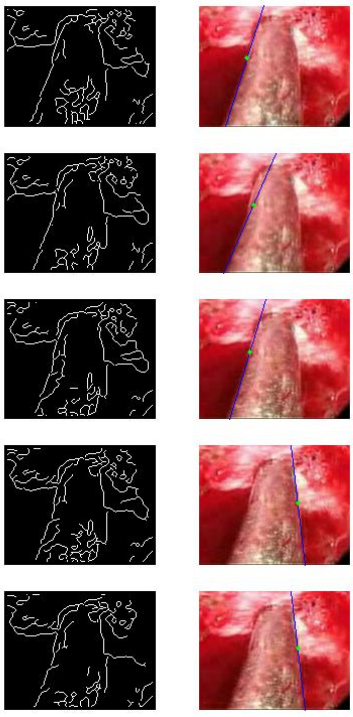

Fig.2. The edge images (left column) and the corresponding strongest lines (right column) detected by Hough transform in the local windows around the tracked objected for frame 5th – 9th of a neurosurgical video.

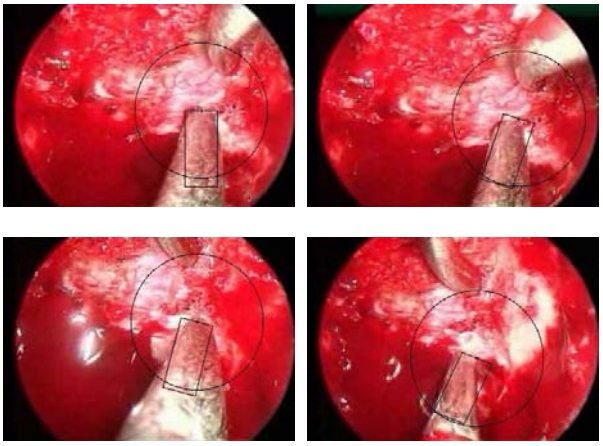

Fig. 3 Detected ROI in 1st, 10th, 80th and 120th frame of the neurosurgical video. The circle, which is center at the tip of the instrument, denotes the ROI. The rectangle is the tracked area.

IV. Update Of Roi

In order to improve video coding efficiency, the ROI is expected to have a smooth border and not to vary drastically in its shape. The following ROI updating algorithm is utilized to satisfy these requirements. As shown in Fig. 4, the ROI is updated at equal time intervals T. We assume that, at $t = 0$, the ROI is the default (solid circle). Inside this ROI, the dotted small circle centered at the cross with radius r represents the tip location of the instrument and an effective range which is predefined and required to be part of the updated ROI. At $t = 0$, no update is necessary since the effective range is entirely inside the default ROI. At $t = T$, the tip location changes and a portion of the effective range moves outside of the existing ROI. Then we define a new ROI as an enclosed region of the existing ROI, the effective range and the two tangential lines. At $t = 2T$, the ROI is similarly updated. Here we do not delete the previous update in favor of a gradual morphing of the shape. At $t = 3T$, where the effective range moves back to the position inside the default ROI, we remove the update at $t = T$. At $t = 4T$, the effective range is still within the default ROI, the update of $t = 2T$ is now removed.

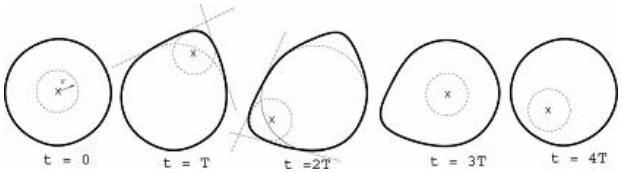

Fig. 4 Changes of ROI shapes with respect to time

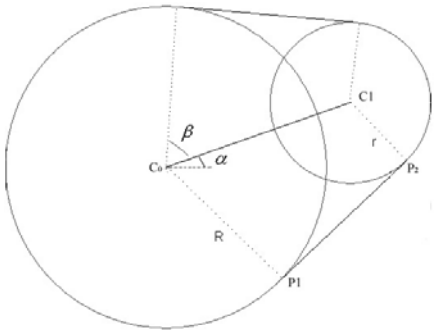

Fig. 5. Two circles are connected by tangential line

The problem here is to find the lines that connect two circles. Where, we assume the centres of the two circles are C_0 and C_1 are the coordinate in the space but in the 2×1 vector form. The radiuses are R and r respectively as Fig. 5. The distance of C_0 and C_1 are d and the angle between line C_0C_1 and x axis is α. Then the coordinate of the tangential lines which connects the two circles P_1, P_2 can be expressed as (10) and (11) with vector form,

$$P_1 = T \cdot Q_1 + C_0, \quad (10)$$

$$P_2 = T \cdot Q_2 + C_0. \quad (11)$$

Where

$$T = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix},$$

$$Q_1 = \left[R\cos(\beta) \quad \pm R\sin(\beta)\right]',$$

$$Q_2 = \left[d + r\cos(\beta) \quad \pm r\sin(\beta)\right]',$$

$$\beta = \cos^{-1}((R-r)/d).$$

Fig.6 shows the results of detected ROI with updating, where the rectangles denote the tracked instrument, the rings are the final ROIs. It can be observed that the updated ROIs are smooth, convex and without corners, which facilitates the shape coding procedure employed by the MPEG-4 encoder.

V. Conclusion

A new method to track a surgical instrument and obtain a ROI in the neurosurgical video is developed to support remote intraoperative monitoring. The developed method is also potentially useful in image-guided surgery to track the positions of surgical instruments.

Acknowledgment

This research was supported in part by USA National Institutes of Health grants No. R01EB002309, R01NS38494, R01EB002099, and Computational Diagnostics, Inc.

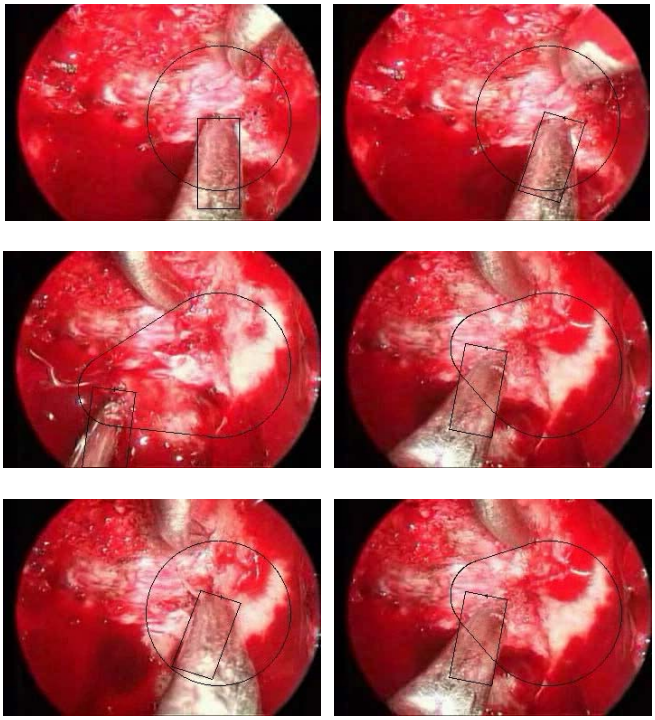

Fig. 6. Updated ROI in the 1st (up-right), 35th (up-right), 157th (middle-left),210th (middle right), 270th (bottom-left) and 320th (bottom-right) frame of a neurosurgery video. The rings are the final ROIs. The rectangles are used for show the tracked object.

References

[1] J.B West and C.R Maurer, "Designing Optically Tracked Instruments for Image-Guided Surgery," *IEEE Transactions on Medical Imaging,* vol. 23, NO. 5, pp. 533-545, May 2004.

[2] S.H Park, J.H Park, S.H R and T. Jeong, "Real-Time Monitoring of Patients on Remote sites", *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, No 3, pp.1321-1325, 1998.

[3] D. Gibson, M. Spann, and S.I. Woolley, "A Wavelete-Based Region of Interest Encoder for the Compression of Angiogram Video Sequences," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, NO. 2, pp. 103-113, June 2004.

[4] J. Xu, H. Ilgin, Q. Liu, A. Kassam, R. J. Sclabassi, and M. Sun, "Content-based video coding for remote monitoring of neurosurgery," *In Proc IEEE EMBS'04*, San Francisco, CA, 2004.

[5] A. Sinha, G. Agarwal and A. Anbu, "Region-of-interest based compressed domain video transcoding scheme," *IEEE International Conference on Acoustics, Speech, and Signal Processing 2004. (ICASSP '04)*, Vol.3, 17-21, May 2004, pp. III - 161- 164.

[6] H. Lin, J. Si and G.P. Abousleman, "Knowledge-based hierarchical region-of-interest detection," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. (ICASSP '02),* Vol. 4, 13-17 May 2002, pp. IV-3628 - IV-3631.

[7] C.C Chen and O.T.C. Chen, "Region of interest determined by perceptual-quality and rate-distortion optimization in JPEG 2000," *International Symposium on Circuits and Systems, 2004. ISCAS '04,* Vol. 3, 23-26 May, 2004 pp. III - 869-72.

[8] D. Comaniciu, V. Ramesh and P. Meer, "Kernel-Based Object Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 25, May, 2003, pp. 564 – 577.

[9] D.W. Scott, Multivariate Density Estimation, Theory, Practice and Visualization, New York: Wiley, 1992.