TRACKING ABNORMALITIES IN VIDEO CAPSULE ENDOSCOPY USING SURROUNDING FEATURES WITH A TRIANGULAR CONSTRAINT

Yukiko YANAGAWA[⋆†] Tomio ECHIGO[‡] Hai VU[⋆] Hirotoshi OKAZAKI[∗]
Yasuhiro FUJIWARA[∗] Tetsuo ARAKAWA[∗]
Yasushi YAGI[⋆]

[⋆] Department of Intelligent Media, Institute of Scientific and Industrial Research, Osaka University
[‡] Department of Engineering Informatics, Osaka Electro-Communication University
[†] OMRON Corporation
[∗] Graduate School of Medicine, Osaka City University

ABSTRACT

This paper proposes a method to track abnormalities in successive frames in a capsule endoscopic image sequence. Exact tracking of an abnormality in the gastrointestinal tract is useful in preparing the content for educational systems. However, if the abnormality is deformable over continuous frames and its features are not highly distinct, it is difficult to track abnormalities precisely. The proposed method uses not only the abnormality image features, but also surrounding features, called supporters. In a capsule endoscopic image sequence, however, both the supporters and the target are difficult to track using only their image features, because these features are indistinguishable. Since estimation using only surrounding features is hard, the proposed method uses triangular constraints among supporters as well. The proposed method is able to track successfully even if the target and supporter features are indistinguishable. Finally, we evaluate the proposed method using eight major types of abnormalities.

***Index Terms*—**

Tracking, Wireless Capsule Endoscopy Images, Gastrointestinal, Education System

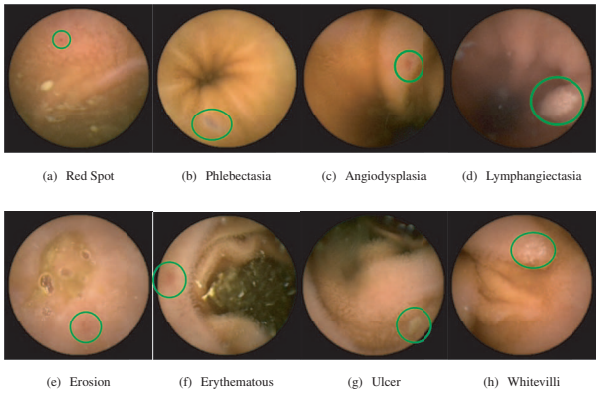


Fig. 1. Abnormal sample

1. INTRODUCTION

Diagnosis by Wireless Capsule Endoscopy (WCE) utilizes a swallowable device that is propelled by peristalsis through the Gastrointestinal (GI) tract. Since this requires different skills to those of conventional endoscopy, medical doctors must acquire new skills. Therefore, an educational system is needed to train unskilled medical doctors to identify abnormal regions from capsule endoscopic images. When a medical doctor unskilled in WCE captures an abnormality image in the education system, ground-truth data, which includes the period between the abnormality's appearance and disappearance as well as its type and location, is needed to decide whether the answer is correct. Currently, medical doctors annotate each frame manually. However, as an abnormality can appear in any number of frames, ranging from only a few frames to hundreds of frames, this is a very difficult task for medical doctors. Therefore, it is useful to be able to automatically track the region of an abnormality demarcated by a medical doctor.

Conventional tracking methods [1, 2], like Mean Shift [3] and SIFT [4], rely on strong image features of the target. However, since the shape of an abnormality is deformable over successive images and the features of the abnormality are also weak, it is difficult to track such abnormalities precisely. The main cause of failure is that the target of an abnormality usually has no distinguishable image features. Nevertheless, Grabner proposed a tracking method [5] for an invisible target, which employs surrounding features such as large eigenvalues of a Hessian matrix to track a target. In addition, the conventional method can successfully track a target with strong features by using surrounding supporters, not only in the visible case, but shortly also for an occluded target.

In our application, capsule endoscopy images include a target and supporters with very weak features, as shown in Fig. 1. Abnormalities, without strong edges and color changes, are shown within the green circles. The supporters as well as the target are difficult to track using only their image features, because these are ambiguous. Therefore, the proposed method employs three mutual supporters with a triangular constraint whereby the triangular shape is maintained, but weak deformation is allowed between successive frames. The position of the target can be robustly determined from voting relative positions based on affine transformation of several triplet supporters.

2. SUPPORTERS WITH TRIANGULAR BASED TRACKING

2.1. Overview

The proposed method adopts a coarse-to-fine approach. For the coarse estimation, the target position is roughly estimated by the matching results of the supporters over successive frames. The supporters comprise several points around the target, which are detected by a point detection method. Then in the fine estimation, a precise position is estimated based on the features of the abnormality itself.

The three main characteristics of the proposed method are as follows.

(I) Between successive frames, the supporters are matched by considering the triangular constraint (Fig. 2).

(II) An abnormal position is voted using an affine matrix calculated from triplet supporters (Fig. 3).

(III) The precise position is estimated by features of the target itself.

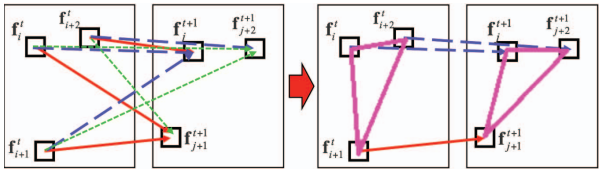


Fig. 2. Matching using the triangular constraint

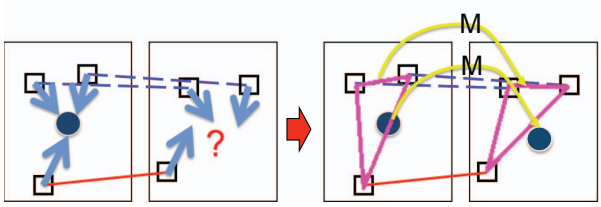


Fig. 3. Voting based affine transform

Regarding characteristic (I), supporters cannot be tracked correctly one-by-one, because a capsule endoscopic image sequence does not have clear features and besides the features are very similar in different regions. The proposed method adds a triangular constraint for matching supporters by maintaining the triangular shape between successive frames.

As for characteristic (II), the conventional method uses one-to-one relationships in the voting process for the target position. As such, it cannot cope with the changing posture or movement of the camera. On the other hand, the proposed method uses relationships between the triplets of supporters and the abnormal position when voting for the target position. In the next frame, by matching triangles, the relative position of the target is transformed according to that of the triangles, and then voting for the target position is carried out.

As for characteristic (III), an error in the position may result from the voting because of the deformable movement of the scene. Conventional approaches assume a model, such as a grid model, based on anchor points. However, as a capsule endoscopic image sequence consists of several regions that move differently, such a model is not applicable. Furthermore, to detect disappearances, the target information should be used. Thus, the proposed method uses the color information of the target, which is more reliable in situations where the target is deformable.

These coarse and fine estimations are explained in the following sections.

2.2. Coarse estimation using triangular constraint from supporters

First, supporter candidates $\mathbf{f}$ represented by Eq. 1 are extracted from an initial frame and a neighboring frame by the Kanade-Lucas-Tomasi (KLT) method [6]. Each frame is divided into four areas and points are detected in each area. The number of features is always smaller than the threshold value set in advance.

Assume $\mathbf{f}$ has position $\mathbf{x}$, its relationship with other supporters is $\mathbf{G}$, and the image features are $\mathbf{F}$.

$$\mathbf{f}_i^t = (\mathbf{x}_i^t, \mathbf{G}_i^t, \mathbf{F}_i^t) \tag{1}$$

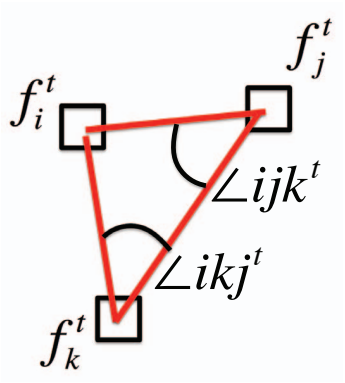


Fig. 4. Description of supporters

Image features $\mathbf{F}$ are calculated using a GI color histogram around KLT feature points. GI color was proposed by Hai Vu [7] and the GI color space is an efficient color space for describing color features in capsule endoscopic image sequences. The relationships between supporters $\mathbf{G}$ are expressed using Eqs. 2 and 3 with each supporter having two angles, where $\angle ijk$ denotes the angle consisting of the three points, i, j, k. In short, the combination of a supporter and two other supporters is described by $\mathbf{G}$. Therefore, one pair consists of three points, with each supporter having ${}_{n-1}C_2$ relations if the total number of supporters is n.

$$\mathbf{G}_i^t = \left(\mathbf{g}_{i,1,2}^t, \cdots, \mathbf{g}_{i,n-1,n}^t\right) \tag{2}$$

$$\mathbf{g}_{i,j,k}^t = \left(\angle \mathbf{ijk}^t, \angle \mathbf{ikj}^t\right) \tag{3}$$

Second, matching supporters in the initial frame and those in a neighboring frame is optimized by maximizing the energy function described by Eq. 4. This means choosing the pair of supporters that has maintained a positional relationship with another point and has a higher feature matching score.

$$\mathbf{E^t} = \sum_{i=1}^{N}\sum_{l=1}^{M} w_{i,l} S_{i,l} \tag{4}$$

$$w_{i,l} = \begin{cases} 1 \;\; if \; \mathbf{f_i^t} \; and \; \mathbf{f_l^{t+1}} \; is \; matched \\ 0 \;\; else \end{cases} \quad (5)$$

$$S_{i,l} = \begin{cases} B(F_i^t, F_l^{t;i}) \; T(G_i^t, G_l^{t+1}) \\ \quad if \; B(F_i^t, F_l^{t+i}) > Threshold \\ 0 \;\; else \end{cases} \quad (6)$$

$B(F_i^t, F_l^{t+i})$ is the Bhattacharyya distance between features $\mathbf{F_i^t}$ and $\mathbf{F_l^{t+1}}$ as given by Eq. 7. If the Bhattacharyya distance between features is smaller than the threshold, the pair is not accepted.

$$\mathbf{B(a,b)} = \sum_{p=1}^{N} \sqrt{a_p b_p}, \quad (7)$$

where considering samples a and b, N is the number of partitions, and a_p, b_i are the number of members of samples a and b, respectively, in the p-th partition of the histogram.

$\mathbf{T(G_i^t, G_l^{t+1})}$ is the coefficient of the triangular constraint, described by Eq. 8.

$$\mathbf{T(G_i^t, G_l^{t+1})} = \frac{\sum t}{(n-1)(n-2)} \quad (8)$$

$$t = \begin{cases} 1 \; if \; \angle ijk \simeq \angle lmn \cap \angle ikj \simeq \angle lnm \\ 0 \;\; else \end{cases} \quad (9)$$

Here, $\sum t$ is the number of triangles that include points i and l, and whose shape does not change between successive frames. If a triangle has one angle of almost 180, it is not included in the total number.

After creating supporter pairs between successive frames, a voting map is created. Locations of abnormalities are voted for on the voting map using an affine matrix that is calculated by each matching supporter triple.

2.3. Fine estimation using the abnormality features

In coarse estimation, a voting map is created, showing the region containing an abnormality. In this section, we explain how tracking the region is carried out.

In the initial frame, an abnormality color feature is described in the GI color space. Then, the most distinguishable component between the histogram of the target and the neighboring region is selected by choosing the minimum Bhattacharyya distance as given by Eq. 7.

After coarse estimation, a score is calculated by Eq. 10 for any position whose voting score is greater than the threshold.

$$\mathbf{V(x,y)} = C(x,y) B(Ab_{initial}, F), \quad (10)$$

where x, y is the position, $C(x,y)$ is the voting score calculated by the coarse estimation process, $B(Ab_{initial}, F)$ is the Bhattacharyya distance between a GI color histogram of the abnormality in the initial frame and that around position x, y in the current frame. The position of the maximum score is detected as the target. If maximum $B(Ab_{initial}, F)$ is lower than the threshold and the voting score in the image area is not greater than the threshold, the current frame does not include the target.

3. EXPERIMENTAL RESULTS

The proposed method was applied to a total of 120 randomly chosen sequences. Eight major types of abnormalities were present in these sequences as a shown in Fig. 1. The proposed method (with triangle) is compared with a method that does not use the triangular constraint or an affine matrix (non-triangle). The results are illustrated in Fig. 5, which shows a cumulative histogram depicting the average distance between the result and ground truth of each sequence. The ground truth is demarcated manually. With the proposed method, the distance between the result and ground truth of almost 70% of the sequence is within 20 pixels. From Fig. 5 we can see that the method with triangles is better than the non-triangle method.

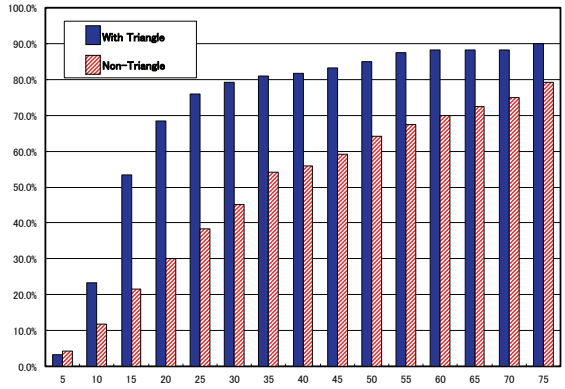


Fig. 5. Cumulative histogram of the distance between the result and ground truth

In an education system, frames where the abnormality appears and disappears are important. In this experiment, an initial frame, located somewhere in the image sequence where an abnormality has been detected, is selected manually. Next, the tracker searches back and forth to detect the frame in which the abnormality is resolved. Table 1 shows the results of detecting the disappearance of the abnormality. In the table, Lost denotes the case where the abnormality is lost by the tracker even though it still exists, while Stay denotes the case where the disappearance of the abnormality is not detected.

	with triangle	non-triangle
Correct	46	32
Stay	58	71
Lost	16	17

Table 1. Results of detecting the existence of an abnormality

Although the proposed method performs better than the non-triangle method, its performance is not satisfactory in terms of the requirements. The results show that the determination of abnormalities in the fine estimation is not that good.

The proposed method can track an abnormality even if its scene differences are large and its appearance is indistinguishable. Fig. 6 illustrates a correct case of tracking with the green rectangle denoting the ground truth and the yellow rectangle the result. Despite the movement in these sequences being irregular, the proposed method can successfully track the abnormality.

Fig. 7 illustrates incorrect tracking. The first row depicts the case that is missed in the coarse estimation. This frame has a lot of similar textures. Thus, in the supporter matching step, many incorrect pairs are created. The second row depicts the case where the

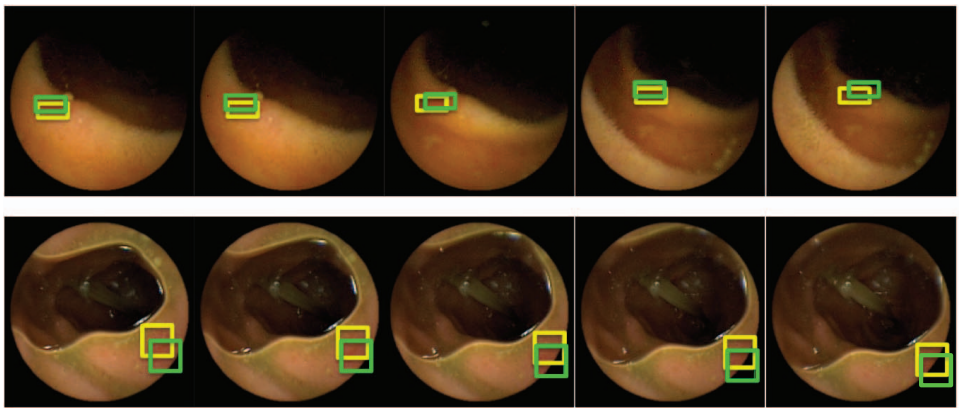


Fig. 6. Example of correct tracking

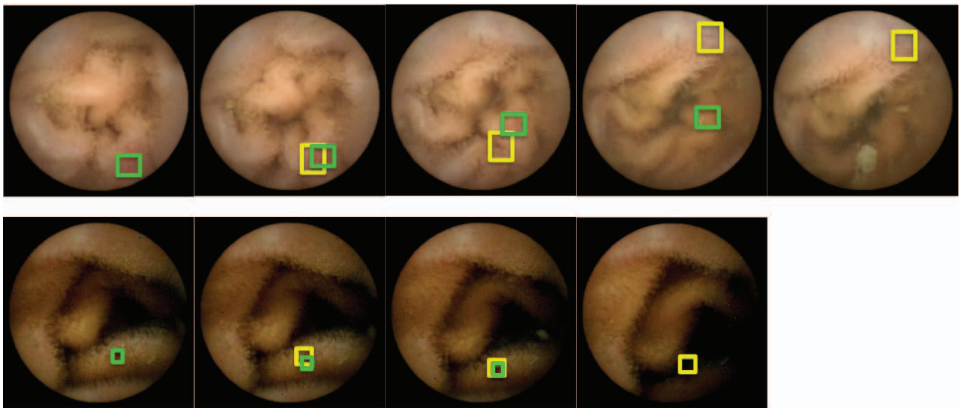


Fig. 7. Example of erroneous tracking

target remains, although the abnormality has disappeared. The failure in both cases is mostly due to the fine estimation process, which estimates the precise position based on color information.

4. CONCLUSION

In this paper we proposed a method for tracking areas with abnormalities in an image sequence from capsule endoscopy. It is not unusual that abnormality images in the small intestine do not have strong features such as sharp edges and distinct color changes. Consequently, the surrounding features are used to estimate the position. The position of the target can be robustly determined from voting relative positions based on an affine transformation of several triplets of supporters. The triangular constraint is also effective as shown in the experimental results, which means that the triangular shape is maintained, while still allowing weak deformation between successive frames. In conclusion, the proposed method is able to track an abnormality even if motion displacement is large and its appearance is indistinguishable. As a future work, we intend representing color components of both the abnormal area and the adjacent one using a Gaussian mixture model in the fine estimation, in order to determine the disappearance frame, that is, the end of the existence of the abnormality.

5. REFERENCES

[1] Son Tran and Larry S. Davis, "Robust object trackinng wvith regional affine invariant features," in *ICCV*. 2007, pp. 1–8, IEEE.

[2] Wei He, Takayoshi Yamashita, Hongtao Lu, and Shihong Lao, "Surf tracking," in *ICCV*. 2009, pp. 1586–1592, IEEE.

[3] Robert T. Collins, "Mean-shift blob tracking through scale space," in *CVPR (2)*. 2003, pp. 234–240, IEEE Computer Society.

[4] Huiyu Zhou, Yuan Yuan, and Chunmei Shi, "Object tracking using sift features and mean shift," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.

[5] Helmut Grabner, Jiri Matas, Luc J. Van Gool, and Philippe C. Cattin, "Tracking the invisible: Learning where the object might be," in *CVPR*. 2010, pp. 1285–1292, IEEE.

[6] Carlo Tomasi and Takeo Kanade, "Detection and tracking of point features," Tech. Rep., International Journal of Computer Vision, 1991.

[7] Hai Vu, Yasushi Yagi, Tomio Echigo, Masatsugu Shiba, Kazuhide Higuchi, Tetsuo Arakawa, and Keiko Yagi, "Color analysis for segmenting digestive organs in vce," in *ICPR*. 2010, pp. 2468–2471, IEEE.