Performance evaluation of simultaneous RGB analysis for feature detection and tracking in endoscopic images

F. Selka[12]
selka.faical@gmail.com
S. Nicolau[1]
stephane.nicolau@ircad.u-strasbg.fr
A. Bessaid[2]
a.bessaid@gmail.com
L. Soler, J. Marescaux[1]

[1] IRCAD Strasbourg, France
[2] GBM
Abou Bekr Belkaid University
Tlemcen, Algeria

Abstract

In laparoscopic surgery, soft tissue motion tracking and 3D structure reconstruction are crucial to provide an augmented reality view in navigation systems. Performing an accurate real-time surface 3D reconstruction requires an efficient detection of interest points. In this paper, we propose an approach to increase the number of good features to track and to improve their tracking robustness in endoscopic images based on simultaneous RGB analysis, instead of gray level only. The proposed method has been evaluated on human and pig endoscopic images, using Shi-Tomasi, SURF and SIFT feature detector and Lucas-Kanade tracking algorithm. Results confirm that our approach increases the number of detected features up to 40% and avoids wrong tracking of about 17% of points, in comparison with gray level channel.

1 Introduction

Minimally invasive surgical (MIS) procedures are gaining popularity in the medical community for their ability to reduce patient recovery time, patient morbidity and patient trauma. The main tool used by surgeons is an endoscopic camera, which is inserted through an orifice (natural or artificial) into the human body. However, this technique has also drawbacks for surgeons such as field of view limitation and 3D vision loss, which lengthen intervention duration. These limitations encouraged several areas of research in the field of the computer vision. Most of these researches try to provide real-time information to surgeons to decrease the limitations mentioned above [9]. Recovering in real-time the 3D geometry of the abdominal cavity could, for example, allow piloting automated systems, with the aim of assisting surgeons and increasing intervention safety. These systems are usually based on real-time organ surface reconstruction, which relies on tracking and matching of feature points. The surface reconstruction quality thus highly depends on this first tracking step.
In this paper, we highlight that the standard approaches for feature tracking in endoscopic

images are performed on gray level images, and show that considering all RGB information increases the number of good features and improves the robustness of the tracking step in endoscopic images.

Generally, methods proposed for 3D surface reconstruction in MIS [9, 10, 12], are based on tracking regions of interest in successive image sequences. The problem of locating a region of interest in one image and finding the corresponding region in another one is difficult in MIS since images can be low in contrast, noisy and poorly illuminated[9]. To solve this problem, an algorithm must be used to identify the most robust points for the preliminary step. Feature point detectors based on cornerness measures like Harris[5] and Shi-Tomasi[11] are still famous and used to extract points[10]. More recently, descriptors like SIFT[8], SURF[3], have been widely used to describe and match a region of interest. Both approaches were integrated for tracking deformable soft tissues in MIS[10, 12]. To reduce algorithm complexity and time computation, most works in endoscopic surgery use gray level images to identify features instead of using the full RGB information. Mountney et.al.[10] suggest that color does not seem to bring significant improvement. However, no quantitative evaluation has been provided. Although we agree that computational time is important, it seems also important to provide as much robust and spread points as possible. Point robustness is important to avoid relying too much on a supplementary detection step of wrong matches using RANSAC [4] or removal outliers based on epipolar constraints [6], which can be computationally expensive. In this paper, we show that a simulatenous analysis of RGB channels allows for better identification and robustness of features in endoscopic images. Our approach has been motivated by recent works related to narrow (NBI) or multi (MBI) band imaging techniques [7]: using a specific linear combination of R, G, B channels can highlight the visualisation of different tissues which do not absorb identical wavelengths. The remaining part of this paper is structured as follow: In section 2, we firstly argue that considering simultaneous RGB analysis allows to increase the number of robust features compared to the use of gray-level image only. Then, we explain how we use the 3 RGB channels to identify and track features. In section 3, we show on human and pig *in-vivo* data that our approach increases the number of good features up to 40% and that we track more points than in gray level channel with a better robustness.

2 Simultaneous detection and tracking on RGB channels

The use of the RGB space is very common in image processing since it is provided by most acquisition devices. In RGB space, each signal corresponds to a different wavelength band of the visible spectrum. The gray level (BW) image is a linear combination of the 3 signals: $BW = 0.299 \times R + 0.587 \times G + 0.114 \times B$.

These weights depend on the exact choice of the RGB primaries, the ones we provide in previous equation are typical [1]. Usually, good features are selected using a detector based on gradient intensity. A basic analysis highlights the 2 following drawbacks. On the one hand, if one point is detected on B channel for instance, and not on the others, it is then likely that this point will not be detected in BW image, since the gradient intensity in BW will be weighted on all channels and B contribution is too low. Thus, a point that could be a good feature to track in an independent channel will not be tracked. On the another hand, a point detected in BW image may be tracked more efficiently in the channel in which its gradient properties are stronger. For these reasons, we propose to perform a simultaneous analysis of RGB channels. Our method can be divided in 2 main steps, the selection of good

features to track in R, G, B channels, and the tracking strategy.

2.1 Selection of features to track

In this subsection we firstly describe how we select the features that will be tracked along the endoscopic video sequence. We choose to detect features using Shi-Tomasi algorithm [11]. Let S_x be the feature response of the channel x : S_x is a set of 2D point coordinates. At the very beginning of the video i.e. frame 0, we firstly compute S_R, S_G and S_B and consider the union of these point sets $S_R \cup S_G \cup S_B$. Secondly, we merge points in $S_R \cup S_G \cup S_B$ which are very close. Typically, a point $M(x,y)$ and a point $U(u,v)$ are considered identical if $(x+\tau > u > x-\tau)$ and $(y+\tau > v > y-\tau)$, and the new point is the average of M and U. τ is chosen considering the resolution of the image and the size of the observed scene. In our case, we use a HD camera and the scene has a rough size of 30 cm, it is then reasonable to choose $\tau = 5$ pixels since it corresponds to 0.7 mm. Finally, the new set after the point merging process is called S^0_{all}. Note that the points in S^0_{all} are no longer associated to a specific color channel. In fact, we consider that all of them can be a good feature to track in all channels. This choice may seem odd, but we have experimentally noticed that due to illumination change a good feature in R channel only, for instance, can become a good feature in B or/and G channels after several frames.

Points in S^0_{all} will be tracked along frames, the updated point set in frame i will be denoted S^i_{all}. After many frames, several points in S^i_{all} are no more visible in the video sequence due to the camera movement. Once 10 points have moved out of the endoscopic image, we decide to launch again Shi-Tomasi detector on R, G and B channels, to compute $S_R \cup S_G \cup S_B \cup S^i_{all}$ and to merge points which are too close, obtaining an updated S^i_{all} which now includes new points from the simultaneous detection on RGB channels. This process is performed each time that 10 points have left the endoscope field of view to reduce computational local. The next subsection explains how points in S^{i+1}_{all} are estimated from their position in the previous frame S^i_{all}.

2.2 Tracking strategy

Let P^i be a point in S^i_{all} in frame i. In this subsection we explain how we compute the position P^{i+1} in frame $i+1$ from its estimated position P^i in the previous frame. Firstly, we estimate on R, G and B channels the motion of P^i in frame $i+1$ using Lucas Kanade algorithm [2]: $P^{i+1}_R, P^{i+1}_G, P^{i+1}_B$. Ideally, if the tracking was perfect we should observe $P^{i+1}_R \simeq P^{i+1}_G \simeq P^{i+1}_B$. Moreover, since we assume that the endoscope motion varies slowly, the motion of P^i (i.e: $\|P^i\ P^{i+1}\|$) should have a magnitude close to the average motion of $P^{i-2}, P^{i-3}, \ldots, P^{i-n}$ on the previous frames. Mathematically, this means that the distance m^{i+1} of P^{i+1}_R, P^{i+1}_G and P^{i+1}_B to their gravity center O^{i+1} : $m^{i+1} = \frac{O^{i+1}P^{i+1}_R + O^{i+1}P^{i+1}_G + O^{i+1}P^{i+1}_B}{3}$ should be extremely small (cf. Fig 1 a). Practically, $P^{i+1}_R, P^{i+1}_G, P^{i+1}_B$ are not identical, and their motions can be inconsistent.

In the best case, the distance between these 3 estimations is very small and motion estimation is consistent, it is then likely that they are all reasonable estimations and we then decide that our estimation of P^{i+1} in frame $i+1$ will be equal to $\frac{P^{i+1}_R + P^{i+1}_G + P^{i+1}_B}{3}$. More generally, the computation of P^{i+1} will depend on the scattering of $P^{i+1}_R, P^{i+1}_G, P^{i+1}_B$ (represented by m^{i+1}) and on their motion $d^{i+1}_R = \|P^{i+1}_R P^i\|$, d^{i+1}_G and d^{i+1}_B.

Briefly, we discard in a first step P^{i+1}_x if d^{i+1}_x is above a threshold $\lambda_{\overline{d}}$. In a second step, we compute the scattering magnitude m^{i+1} with the remaining points and estimate P^{i+1} depending on m^{i+1}. Fig.1 a-f describes all possible cases, the yellow circle representing the

threshold of tolerated scattering $\lambda_{\overline{m}}$. Note that Lucas Kanade algorithm sometimes fails at providing an estimation (in that case, the feature is called *lost* point in Sec. 3.2). Practically, it can happen that P^i motion in frame $i+1$ is not estimated on channel x. In our algorithm, this case is identical to $d_x^{i+1} > \lambda_{\overline{d}}$.

The two thresholds $\lambda_{\overline{d}}$, $\lambda_{\overline{m}}$ are defined as follow. Let $\overline{d}$ be the average displacement of several points P selected in ROI located in the image center over the last 20 frames $\overline{d} = \sum_{j=0}^{20} \frac{\|P^{i-j}\ P^{i+1-j)}\|}{20}$ and $\sigma_{\overline{d}}$ the standard deviation of these distances. We suppose there is no tracking failure over these 20 frames since the illumination condition does not change in the image center. In a similar way, $\overline{m}$ is the average of m^i on the 20 previous frames, and $\sigma_{\overline{m}}$ its standard deviation. We choose $\lambda_{\overline{d}} = \overline{d} + 3\sigma_{\overline{d}}$ and $\lambda_{\overline{m}} = \overline{m} + 3\sigma_{\overline{m}}$. $\overline{d}$, $\overline{m}$ and their standard deviation will be re-estimated every 20 frames.

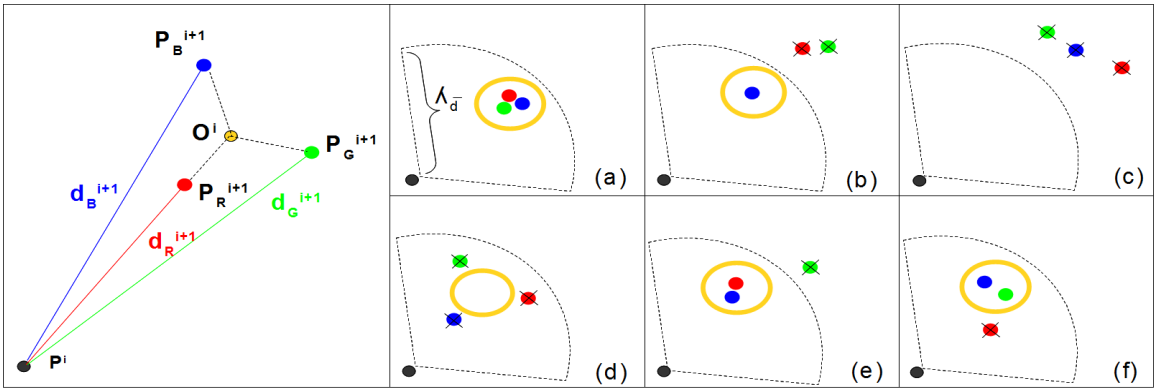

Figure 1: Left: $P_R^{i+1}, P_G^{i+1}, P_B^{i+1}$ are the estimated position of P^i in frame i+1, and O^{i+1} their gravity center. Right: these 6 figures sketch possible tracking scenario during simultaneous tracking in RGB channels. In each case, we firstly discard estimation in channel x if d_x^{i+1} is above $\lambda_{\overline{d}}$ (case b,c,e). Then, if m_i is below $\lambda_{\overline{m}}$, P^{i+1} is the average of the remaining points (case a,b,e). If $m_i > \lambda_{\overline{m}}$ (case d,f), we check if the distance between 2 points is below $\lambda_{\overline{m}}$ and average them (case f). If not, the tracking is stopped (case d).

3 In-vivo evaluation of RGB simultaneous analysis

We evaluate our approach on 4 *in-vivo* HD video sequences corresponding to an abdominal exploration. Two of them contain a human liver and gallbladder and the other two contain pig bowels (cf Fig.2). From left to right, the sequence contains respectively 550, 675, 625, 875 frames. We used a 3 CCD HD endoscope Storz Image1 Hub with resolution of 1920×1080 and 0° degree endoscopic lens. In this evaluation, we firstly evaluate the supplementary amount of features using all RGB channels with Shi-Tomasi, SURF and SIFT detectors, in comparison with gray level only. Secondly, we show that the feature set is more robust when tracked with our approach than in gray level channel only.

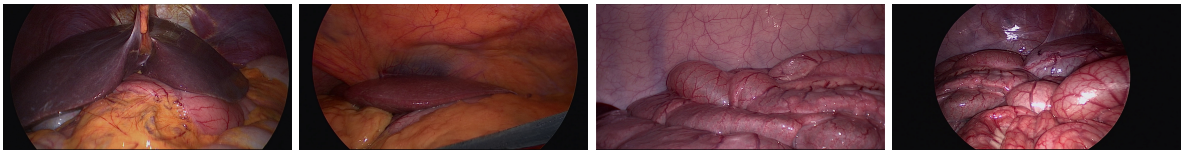

Figure 2: Image sample of the different sequences used in our analysis.

3.1 RGB simultaneous analysis to increase the amount of features

We evaluate our feature selection method on all frames of the 4 endoscope sequences using Shi-Tomasi, SIFT and SURF algorithm and compare it to feature selection on gray level

image only. The Opencv library was used with the following parameters, which were used on all sequences: for Shi-Tomasi the threshold was defined as 0.02 with minimum distance of 35. The derivatives were calculated using generalized Sobel with aperture size = 3. A smoothed window of a Gaussian ($\sigma = 1$) with (3×3) size was used to average the derivatives. For SIFT: octave = 4, threshold 0.05 and edge threshold = 5 was taken. For SURF: hessian threshold = 800. Tab.1 shows the average percentage of new features compared to features detected in gray level image for the liver and the bowel sequences. One can see that considering simultaneous detection in all channels can provide more than 40% new features for Shi-Tomasi and SURF and more than 30% for SIFT. Fig.3 provides an example of our feature detection method. It is worthy to note that 99% of points detected in BW image are also selected with our approach.

Detector	Shi-Tomasi		SIFT		SURF	
Sequence	Bowels	Liver	Bowels	Liver	Bowels	Liver
Average	46.63%	44.62%	34.84%	36.50%	42.27%	40.45%

Table 1: Average percentage of new features using simultaneous RGB analysis, compared to gray level analysis only.

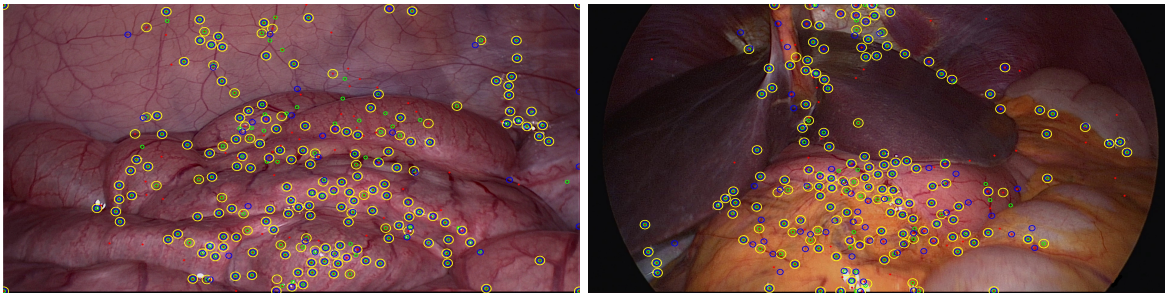

Figure 3: Example of feature detection using Shi-Tomasi. Red/green/blue/yellow circles correspond to features detected in R/G/B/BW channels.

3.2 Evaluation of feature robustness using simultaneous RGB tracking

To compare the tracking robustness of our method with the standard tracking based on gray level image, we propose the following. We select features in each video sequence using our approach. Then, we track all points along sequences and visually report features, which are *lost* by Lucas-Kanade algorithm or *wrong* (tracked feature does no longer represent the original feature). We also report the percentage of frames during which the feature is *lost* or *wrong*. The Lucas Kanade algorithm was used with a (25×25) window size for block matching and a pyramid level set to 3. Tab.2 provides the average results for the feature set tracked on *BW* image and using our method (*RGB*). One can see that our approach allows to avoid losing almost all points, which were *lost* in *BW* image. This means that each tracked feature was always found in at least one of the 3 RGB channels. Results also show that all selected features have been consistently tracked along all sequences, whereas we report an average of about 7 wrong points tracked in *BW* image. On average, our approach increases up to 17% the number of points which are properly tracked.

Liver	BW	RGB	Bowels	BW	RGB	Liver	BW	RGB	Bowels	BW	RGB
Lost point	33	0	Lost point	52	0	Wrong point	8	1	Wrong point	7	0
Lost frame	0.97%	0%	Lost frame	1.67%	0%	Wrong frame	1.07%	0.2%	Wrong frame	3.06%	0%

Table 2: The average result for the two sequences. liver/(bowel). tracking on 302 points over 404 frames /(tracking on 282 points over 360 frames.)

4 Conclusion

This paper presents a method for feature detection and tracking in endoscopic images based on simultaneous RGB analysis, which allows to provide more features and robustness than using gray-level images. This method was evaluated on patient and pig data and results show that up to 40% (resp. 40%, 30%) of supplementary points can be detected using Shi-Tomasi method (resp. SURF, SIFT descriptor). We also show that the proposed method for feature tracking increases the number of robust points up to 17%. We believe this work has highlighted that using all RGB information for detection and tracking of features can nicely decrease the number of outliers for the usual next step: shape reconstruction from motion. In the future, we will strengthen our evaluation on more organs (stomach, pancreas, kidney). We also plan to investigate several linear and non-linear combinations of RGB to enhance specific tissues or organs using wavelength response of each structure.

References

[1] OpenCV. In *http://opencv.willowgarage.com/documentation*, 2010.

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[3] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.

[4] O.G. Grasa, J. Civera, and JMM Montiel. Ekf monocular slam with relocalization for laparoscopic sequences. In *ICRA 2011*, pages 4816–4821. IEEE.

[5] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proceedings of The Fourth Alvey Vision Conference (1988)*, pages 147–152, 1988.

[6] M. Hu et. al. Reconstruction of a 3d surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes. *Medical Image Analysis*, 2010.

[7] S. Kodashima and M. Fujishiro. Novel image-enhanced endoscopy with i-scan technology. *World Journal of Gastroenterology: WJG*, 16(9):1043, 2010.

[8] D.G. Lowe. Object recognition from local scale-invariant features. In *ICV 1999*, volume 2, pages 1150–1157. Ieee.

[9] P. Mountney, D. Stoyanov, and G.Z. Yang. Three-dimensional tissue deformation recovery and tracking. *Signal Processing Magazine, IEEE*, pages 14–24, 2010.

[10] P. Mountney et. al. A probabilistic framework for tracking deformable in minimally invasive surgery. *MICCAI 2007*, pages 34–41.

[11] J. Shi and C. Tomasi. Good features to track. In *Proceedings CVPR'94*, pages 593–600. IEEE, 1994.

[12] D. Stoyanov et. al. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. *MICCAI 2005*, pages 139–146, 2005.