33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Gap Detection in Endoscopic Video Sequences Using Graphs

Alexander Behrens, Masato Takami, Sebastian Gross, and Til Aach

***Abstract*— In minimal invasive surgery (MIS) a complete and seamless inspection of organs, e.g. the urinary bladder, using video endoscopes is often required for diagnostics. Since the endoscope is usually guided by free–hand, it is difficult to ensure a sequence of seamless frame transitions. Also 2–D panoramic images showing an extended field of view (FOV) do not provide always reliable results, since their interpretations are limited by potentially strong geometric distortions. To overcome these limitations and provide a direct verification method, we develop a gap detection algorithm using graphs. Exploiting the motion information of the applied zig–zag scan, we construct a graph representation of the video sequence. Without any explicit global image visualization our graph search algorithm identifies reliably frame discontinuities, which would lead to holes and slit artifacts in a panoramic view. The algorithm shows high detection rates and provides a fast method to verify frame discontinuities in the whole video sequence. Missed regions are highlighted by local image compositions which can be displayed during the intervention for assistance and inspection control.**

I. Introduction

Free–hand video sequences are often acquired during medical minimal invasive interventions, where the whole surgical field is scanned by endoscopes, e.g. to inspect the interior surface of hollow organs. Consequently, the acquisition of sufficient image sequences, which fully cover the whole scene or at least a specified scan region without any gaps or missing views is required. This become difficult, if the camera is moved arbitrarily by free–hand, or the hand–eye–coordination is disturbed, like in a video endoscopy setup. In particular for medical diagnostics, a gap–free image sequence is essential, since only a seamless observation of the whole organ prevents missing tumors and makes an analysis of cancer distributions feasible.

The visualization of a scene using an extended field of view (FOV), constructed by successive images, is commonly provided by panoramic images. Several image mosaicking algorithms have been proposed for video endoscopy [1]–[5]. Missing views become directly visible in an image composition, since they result in image gaps or slit artifacts. Schematic examples of incomplete panoramas composed from a spiral and zig–zag scan are shown in Fig. 1. Although real–time capable mosaicking algorithms are available [1], [2], [5], panoramas are usually constructed after the image acquisition process. This impedes the user to provide additional pan shots of the scene to fill out image gaps interactively. Furthermore, adequate visualizations and interpretations of overview images are not always feasible, since they can show strong geometric distortions induced systematically by the projection of a 3–D object onto a 2–D plane. Thus, gaps may be missed or covered by distorted images, which lead to wrong interpretations. Only local panorama images, which show an extended but still limited FOV allow feasible and less distorted visualizations [6], [7].

The authors are with the Institute of Imaging & Computer Vision, Faculty of Electrical Engineering and Information Technology, RWTH Aachen University, Germany

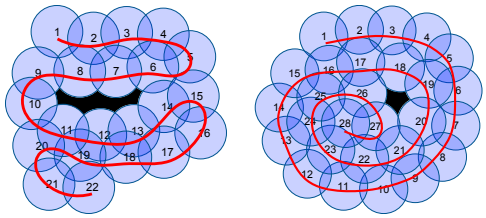

Fig. 1. Discontinuous frame transitions during a spiral (left) and a zig–zag scan (right) of an endoscope (trajectories are highlighted by solid red lines) result in gaps within the final panoramic image compositions.

To overcome these limitations, we discuss in this paper an algorithm to verify continuous and seamless frame transitions of a given input video sequence without composing any global panorama image. Using an extended frame matching method adaptive to the applied scan scheme, a graph representation is constructed to describe the transitions and spatial relations between all video frames. The characteristics of the graph are then analyzed to detect discontinuities in the data, which would result in gaps within a panorama image. Again, instead of constructing one global overview image, the verification process is described by a graph search algorithm, which is computationally less complex and independent from mapping and distortion effects.

II. Graph Construction

First, an undirected graph $G = (V, E)$ is constructed from a given video sequence $\mathcal{I}$ of N frames. Each of its nodes V represents an image, labeled by its frame number. The edges E of a node describe its spatial neighborhood. If two images overlap and share a common transition region, an edge connecting their graph nodes is constructed. If then the whole border of an image is overlapped by neighbor images, its graph node is characterized as a *center node* and highlighted by a bold circle line. An example of aligned video frames and their graph representation is given in Fig. 2.

A. Image Matching

To identify image overlaps our feature–based matching [1] is applied using SURF [8]. Based on point correspondences $p_i \leftrightarrow p_j$ an affine transformation matrix $H_{j,i}$ between two images I_i, I_j is estimated including RANSAC outlier rejection [9]. Commonly, a transition zone is identified by

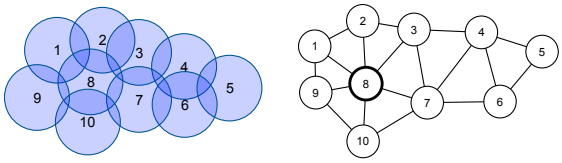

Fig. 2. Video endoscope images aligned on a plane (left) and its graph G with nodes V = {1,2,...,10}, and edges $E = \{\{1,2\},\{1,8\},\{1,9\},\{2,3\},\ldots,\{9,10\}\}$ (right).

warping the whole image I_j pixel–wise into the coordinate system of image I_i using $H_{j,i}$. Here, we analyze the overlap region based on the intersection points of the image borders only. Although our algorithm is generally independent on the image shape, we focus on the processing of elliptically shaped endoscopic images, as shown in Fig. 3.

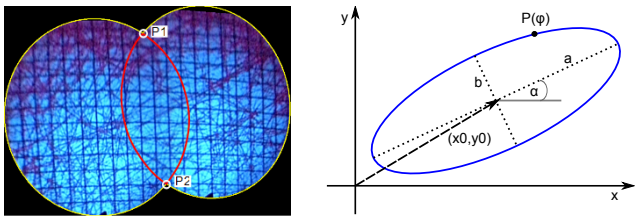

Fig. 3. Overlap region of two endoscopic images with intersection points P_1 and P_2 of the elliptical shaped image borders (left). Ellipse described in parameter form with center point (x_0, y_0), major axis a, minor axis b, and rotation angle α (right).

In a first step, the image border of each frame is detected and described by points $P(\varphi)$, fulfilling the ellipse equation

$$P(\varphi) = \begin{pmatrix} x_0 + a\cos(\varphi)\cos(\alpha) - b\sin(\varphi)\sin(\alpha) \\ y_0 + a\cos(\varphi)\sin(\alpha) - b\sin(\varphi)\cos(\alpha) \end{pmatrix}. \quad (1)$$

Using ellipse parameters (cf. Fig. 3) the intersection points of two overlapping ellipses E_1, E_2 are identified by border points, which lie on both ellipses. Without loss of generality solutions of none to four points and the identity case ($E_1 = E_2$) are possible. Although intersection points can also be determined by numerical algorithms [10], we apply a fast and robust geometric method to calculate them in pixel coordinates. Geometrically, the sum of distances from any point $P(\varphi)$ on the ellipse to its foci F_1, F_2 is constant and equal to the major diameter, as described by

$$D(P(\varphi)) = \overline{P(\varphi)F_1} + \overline{P(\varphi)F_2} - 2a = 0. \quad (2)$$

Thus, the position of any point $P_2(\varphi)$ on E_2 related to E_1 can directly be determined. Figure 4 shows, that the sign

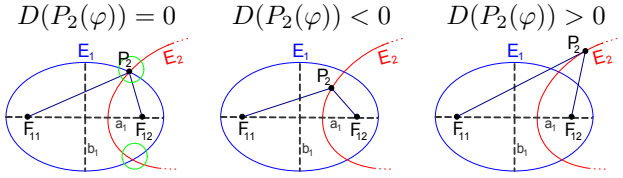

Fig. 4. Distance function $D(P_2(\varphi))$ between point $P_2(\varphi)$ and ellipse E_1 dependent on its position.

of (2) identifies, whether $P_2(\varphi)$ lies inside, outside or on (intersects) E_1. After also the points $P_1(\varphi)$ are analyzed with respect to E_2 in the same way, the boundary

$$\mathcal{B} = \{P_2(\varphi), P_1(\varphi) \mid D(P_2(\varphi)) \leq 0,\ D(P_1(\varphi)) \leq 0\}, \quad (3)$$

of the common overlap region can be defined as a set of points, including the intersections and all border points, which lie inside the opposite ellipse. In an iterative process all other neighbor images of I_i are tested for overlaps in the same way. Finally, the angles φ of all boundary points, which are covered by transition zones with neighbor images, are sampled and saved in the graph node V_i.

Since a feature–based image registration generally requires a certain size of the overlap area, and free–hand video sequences may lead to only small frame transitions, we developed an extended matching method to detect even very small overlap areas. Instead of transforming only I_j into the coordinate system of I_i using $H_{j,i}$, also all neighbor images of I_j and I_i are aligned, as illustrated in Fig. 5. Small overlap regions are then identified by detecting the

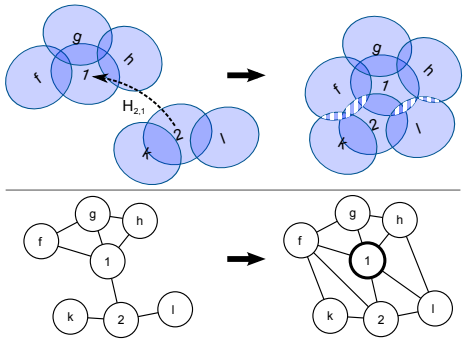

Fig. 5. All neighbor images f, g, h of image I_1, and k, l of I_2, respectively, are transformed into the same coordinate system to detect additional minor overlap regions (visualized by striped patterns) (top). Related graph representations before and after the extended matching process (bottom).

boundaries (cf. (3)) between all image pairs. New transitions result in additional edges between the related graph nodes, as shown in Fig. 5. In conclusion, the following matching steps are carried out:

1) Estimate a homography $H_{j,i}$ between two images I_i, I_j, using SURF features.
2) Transform I_j together with all neighbor images of I_j and I_i into the coordinate system of I_i.
3) Determine all overlap regions with intersection points using (2), and calculate their boundaries using (3).
4) Construct for each overlap region a graph edge, and mark in both nodes (V_i, V_j) the angles φ of the image boundary $\mathcal{B}$ as covered.
5) Select next image pair and go back to step 1, until all image pairs are processed.

B. Matching Strategies

To construct a complete graph of the video sequence, each image I_n can be matched with all previous images $\{I_m \in \mathcal{I} \mid m = 1\ldots(n-1)\}$ in a sequential manner. The brute–force search leads then to a total number of

$\frac{1}{2}n \cdot (n-1)$ comparisons. To reduce the overall matching costs, we exploit the image order of a specified scan scheme such as a zig–zag scan, as shown in Fig. 6.

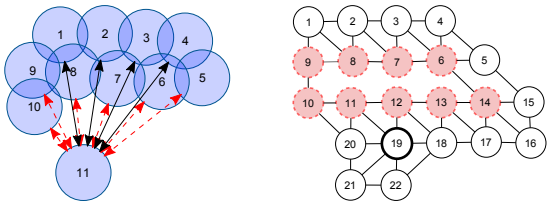

Fig. 6. Image I_{11} is matched ten times using brute–force matching, whereas the number of matches can be limited by the number of images in the previous row (in this example: six, highlighted by red dashed arrows) (left). Labeled graph of the zig–zag scan in Fig. 1, showing *border nodes*, *center nodes* (bold circle lines), and *gap nodes* (red dotted circles) (right).

III. Gap Detection

The verification of seamless transitions in the video sequence is performed by analyzing the graph characteristics. Using the information about the applied scan strategy, e.g. a zig–zag scan, each graph node is categorized and labeled as a *border*, *center*, or *gap node*. A node is classified as a *center nodes*, if all boundary pixels and angles φ, respectively, are covered by overlap regions (cf. (3)). In the case of a zig–zag scan, nodes in the first and two last scan rows, as well as all start and end nodes of each row are selected as border elements. For an inward–moving spiral scan, border elements are specified by nodes of the first complete circulation. A *border node* is defined as a border element, having only one single boundary segment (φ intervall), which has not been covered by any image transitions. In contrast, *gap nodes* may provide more than one uncovered boundary segment. In example the graph in Fig. 6 determined by the zig–zag scan of Fig. 1 shows one center node (19), border nodes (1..5, 15..18, and 20..22), and gap nodes (6..14).

With the definition of a chordless cycle [11] as a repeated sequence of nodes in G, connected by edges E between two adjacent nodes and not connected by any edge between non–adjacent nodes, we specify a graph gap by following statement:

> "A *gap* is a chordless cycle of G, which only consists of *gap nodes*."

With our definition the gap detection in the video sequence can now be formulated as a graph search, looking for chordless cycles in G consisting of gap nodes. In practice this becomes difficult, since a full search is computationally expensive and requires a complete and correct graph representing all image transitions by edges. Missed edges or misclassified nodes will immediately result in false gap detections. Thus, we exploit again the characteristics of the applied scan strategy to limit the gap search in single subgraphs of G and provide a more robust gap detection.

Assuming that gap nodes of a zig–zag scan must be adjacent to nodes of the subsequent scan row, we limit the gap detection to a subgraph G_s consisting of two rows of nodes only. A gap is then identified, if at least two sequential gap nodes are present in each row of G_s.

A. Visualization

After gaps in the graph and their related discontinuous image transitions in the video sequence are detected, local panorama images are composed for visualization. For this purpose, the minimal path between all selected gap nodes of G_s is calculated. The path may include gap, center and border nodes. Applying our iterative mosaicking algorithms [1], a local panorama image is composed from the images of the minimal path. Only at this stage, the gap which results from discontinuous image transitions within the video sequence is visualized and displayed for verification.

B. Refinement

In a postprocessing step an image based refinement is applied to reject false gap detections. Applying a connected component analysis the number of contours c in the panorama is determined. If $c > 1$ the panorama contains a hole, whereas $c = 1$ identifies seamless transitions. In addition to this two classes the zig–zag scans may provide also slit artifacts due to discontinuities between two subsequent rows. Thus, the number of contour pixels along vertical and horizontal profiles are analyzed, as illustrated in Fig. 7. If

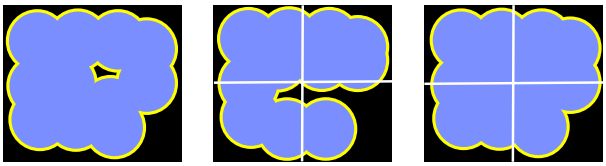

Fig. 7. Panorama with a hole provides two contours (left). In the case of $c = 1$, a number of contour pixels along the image profiles (white lines) bigger than two indicates a slit (center), otherwise the image is seamless (right).

more than two contour lines along the profiles are detected, the image is classified as having a slit.

IV. Results

For evaluation 10 video sequences of a 2–D test pattern, as well as 32 sequences of a spherical urinary bladder phantom are acquired from a free–hand zig–zag guided endoscope. Representative results of one video sequence are given in Fig. 8. The characteristics show that the extended matching, which exploits additional overlaps between local neighbor images, results in significantly higher numbers of detected transitions (graph edges) compared to the standard feature based method. The computing time of the brute–force matching increases with $\frac{1}{2}n \cdot (n-1)$, whereas the zig–zag strategy results in a linear characteristic (cf. Fig. 8). Although the zig–zag matching leads to a faster computation it provides a comparable number of overlap regions as the full matching.

The results of one of the 2–D test pattern sequences are given in Fig. 9. It shows the constructed graph and three detected gaps with their related panorama images. The first and last image visualization show holes (true positives). In the refinement step the second image is re–classified as

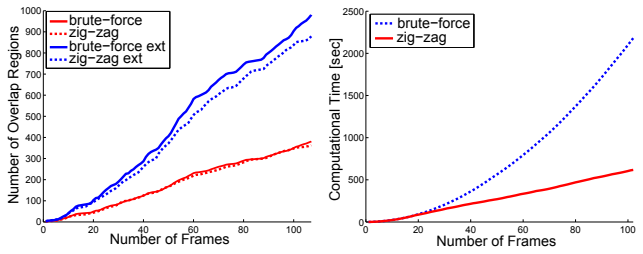

Fig. 8. Number of detected transitions of the extended matching compared to the standard method for a brute–force and zig–zag strategy (left). Computation time for brute–force and zig–zag matching over the total number of frames (right).

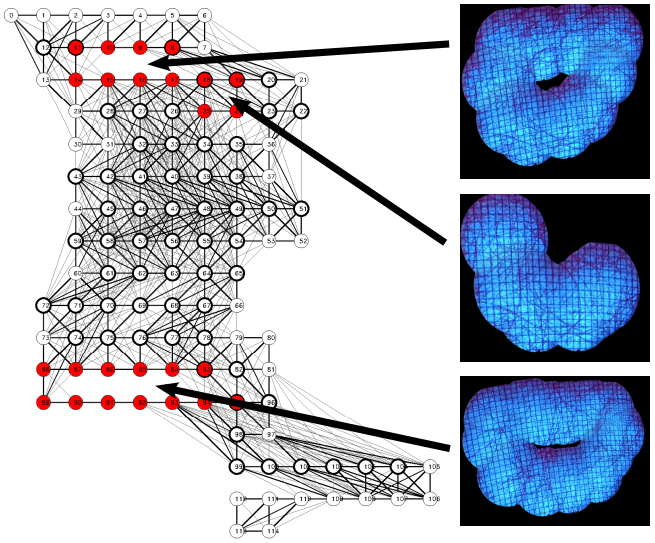

Fig. 9. Graph of a zig–zag video sequence with three detected gaps and their image compositions.

seamless, which is rated as a true negative detection. An example of a detected slit artifact is given in Fig. 10.

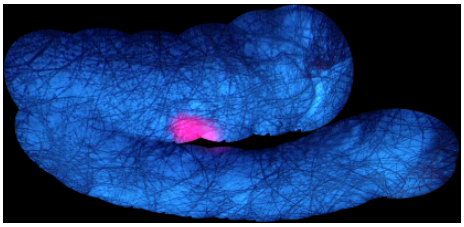

Fig. 10. Image composition showing a slit at the center.

Quantitative evaluation results are listed in Table I. It shows the classification results of all sequences. Based on visual verifications ground truth data are determined for assessment. Detected gaps, which image compositions show holes or slit artifacts are rated as true positives (TP), and rejected seamless ones become true negatives (TN). The high TN rates indicate that the image based refinement method is very robust against false gap detections, generated by incomplete graph representations. Also the high TP rates represent a reliable detection of transition discontinuities in the video sequence. Only if the endoscope deviates strongly

	# Sequences	# Gap Detections	Classification TP	TN
2–D test pattern	10	29	0.94	1.0
spherical phantom	32	38	1.0	1.0

TABLE I

from the zig–zag path, the algorithm may miss gaps due to wrong aligned nodes in the graph of the sequence, which explains the lower TP value of 0.94.

V. Conclusions

We showed, that a given endoscopic video sequences can be tested for seamless and continuous frame transitions using a graph representation. With the information about the scan scheme like a zig–zag, all video frames and their spatial alignment are represented by nodes and edges of the graph. Instead of using global panorama images our algorithm detects gaps in the video sequence, which will result in holes and slits in a panoramic view, by a graph search algorithm. The high classification results indicate a fast, robust and reliable detection. Thus, in the case of inspecting an organ for diagnostics, missed views can directly be identified from the video sequence. Gaps which indicate holes or slits are then visualized by local image compositions for control. With this information the surgeon is able to re–scan the by then unseen regions during the same intervention. In future work the algorithm will be optimized in speed, adapted to further scan strategies, and clinical studies will be carried out.

References

[1] A. Behrens, M. Bommes, T. Stehle, S. Gross, S. Leonhardt, and T. Aach, "Real-time image composition of bladder mosaics in fluorescence endoscopy," *Computer Science - Research and Development*, vol. 26, pp. 51–64, 2011.

[2] T. Bergen, S. Ruthotto, C. Munzenmayer, S. Rupp, D. Paulus, and C. Winter, "Feature-based real-time endoscopic mosaicking," in *Int. Symp. Image and Signal Processing and Analysis (ISPA)*, 2009, pp. 695 –700.

[3] A. Ben Hamadou, C. Soussen, W. Blondel, C. Daul, and D. Wolf, "Comparative study of image registration techniques for bladder video-endoscopy," in *Proc. SPIE Novel Optical Instrumentation for Biomedical Applications IV*, vol. 7371, 2009, pp. 737 118–1–7.

[4] R. Miranda-Luna, C. Daul, W. C. P. M. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, "Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm," *IEEE Trans. on Bio. Engineering*, vol. 55, no. 2, pp. 541–553, 2008.

[5] W. Konen, M. Naderi, and M. Scholz, "Endoscopic image mosaics for real-time color video sequences," in *Computer Assisted Radiology and Surgery (CARS)*, 2007.

[6] A. Behrens, I. Heisterklaus, Y. Müller, T. Stehle, S. Gross, and T. Aach, "2-d and 3-d visualization methods of endoscopic panoramic bladder images," in *Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, vol. 7964, no. 1, 2011, p. 796408.

[7] A. Behrens, T. Stehle, S. Gross, and T. Aach, "Local and global panoramic imaging for fluorescence bladder endoscopy," in *Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)*, 2009, pp. 6990–6993.

[8] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[9] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*. Springer Verlag, 2002.

[11] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, 2nd ed., ser. Annals of Discrete Mathematics, M. C. Golumbic, Ed. Elsevier, 2004, vol. 57.