ENDOSCOPIC BLADDER IMAGE REGISTRATION USING SPARSE GRAPH CUTS

Thomas Weibel\textsuperscript{1,2}, Christian Daul\textsuperscript{1}, Didier Wolf\textsuperscript{1}, Ronald Rösch\textsuperscript{2}, Achraf Ben-Hamadou\textsuperscript{1}

\textsuperscript{1} Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039, Nancy-University, CRNS)
2, avenue de la forêt de Haye, 54516 Vandœuvre-lès-Nancy, France
\textsuperscript{2} Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM)
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

ABSTRACT

Video endoscopy is one of the standard clinical procedures for visually detecting lesions on the internal wall of human bladders. In order to facilitate the diagnosis, it is helpful to build panoramic maps by registering consecutive images from the video sequence. We show how to efficiently reduce the computation time of graph cut based image registration by an order of magnitude. The number of nodes in a graph is greatly reduced using spatial image properties in order to minimize the loss of information. The set of edges in this sparse graph is obtained by applying a watershed transform on the set of nodes. This graph reduction has negligible negative effects on image registration quality compared to a dense graph cut, so that visually coherent panoramic maps of bladder walls can be built. Results demonstrate that the method improves the registration accuracy and reduces the computation time of other endoscopic bladder image registration methods. This work is an important step towards real time map construction.

Index Terms— Graph cuts, image registration and mosaicing, bladder cartography, endoscopy.

1. INTRODUCTION

One of the standard clinical procedures for visually examining the internal structure of human bladders is video endoscopy (see Fig. 1 for image examples). Lesions are usually spread over large areas, whereas the image area is smaller than 1cm\textsuperscript{2}. In order to detect lesions, it is helpful to build panoramic images of bladder walls to extend the field of view. These panoramic images can be used for a second diagnosis after the examination.

Building these panoramic maps requires the registration of consecutive image pairs in order to transform the image sequence into a global coordinate system. While the bladder surface prevents the robust detection of image keypoints\cite{1}, good registration results can be achieved with graph cuts, but the drawback is the high computational demand. In \cite{2}, the authors propose a method to solve a non-rigid image registration problem using multi-level graph cuts in order to deal with the large number of possible point to point correspondences and achieve a speed improvement of 50\% in comparison to a single level graph cut. In this contribution, we take a different approach to deal with the high computational demands of graph cut registration. A reduced set of nodes is selected based on spatial image properties to minimize the loss of information. The edges of this sparse graph are obtained by applying a watershed transform on the nodes. Results show that the negative effects of node reduction on image registration quality are negligible and visually coherent panoramic maps of bladder walls can be computed. The registration quality is superior to previously evaluated methods in \cite{3}, while the computation time is reduced by an order of magnitude. This speedup is a step towards computing panoramic images in real time.

2. IMAGE REGISTRATION USING GRAPH CUTS

2.1. Image Registration

There are only small instrument displacements between consecutive images $I_k$ and $I_{k+1}$ from an endoscopic video sequence. Therefore it is possible to successfully register image pairs $(I_k, I_{k+1})$ using 2D perspective transformations. These can be used to place the sequence into a global coordinate system to create a panoramic image. The transformation corresponds to the homogeneous matrix $\mathcal{H}$ given in (1). It maps a pixel $p' = (x', y') = \left( \frac{x}{w}, \frac{y}{w} \right)$ from the target image $I_{k+1}$ to

\begin{align*}
\mathcal{H} &= \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} \\
\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} &= \mathcal{H} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11}x' + a_{12}y' + a_{13} \\
a_{21}x' + a_{22}y' + a_{23} \\
a_{31}x' + a_{32}y' + a_{33}
\end{pmatrix}
\end{align*}
a pixel \( p = (x, y) \) in the source image \( I_k \):

\[
\begin{pmatrix}
    u \\
    v \\
    w
\end{pmatrix}
= \begin{bmatrix}
    f \cos \phi & -S_x \sin \phi & t_x \\
    S_y \sin \phi & f \cos \phi & t_y \\
    h_1 & h_2 & 1
\end{bmatrix}
\begin{pmatrix}
    x \\
    y \\
    1
\end{pmatrix},
\]

where \( t_x, t_y \) are translation parameters, \( \phi \) is an in-plane rotation, \( f \) is the scale factor, \( S_x, S_y \) are shearing parameters and \( h_1, h_2 \) are perspective parameters [4].

### 2.2. Energy minimization using graph cuts

Graph cuts provide a powerful and intuitive framework for labeling problems such as image segmentation, image denoising and computing visual correspondences between two images (stereo matching, optical flow, registration) [5]. Those applications have in common that they can be formulated as an energy function decomposable into a sum of functions on nodes and edges:

\[
E(\chi) = D(\chi) + \lambda V(\chi),
\]

where \( \chi = \{x_1, \ldots, x_p, \ldots, x_N\} \) is a labeling that assigns a label from a set of labels \( L \) to each node. \( D(\chi) \) and \( V(\chi) \) are usually referred to as data and interaction terms and \( \lambda \) controls the smoothness of the resulting labeling.

When dealing with binary labeling problems, such as image segmentation, an exact solution can be obtained with a single graph cut. In the case of multi-label problems, such as optical flow computation or stereo matching, where \( L \) contains all possible pixel displacements for an image point between two images \( I_k \) and \( I_{k+1} \), an approximate solution can be obtained by iteratively computing binary graph cuts over the label set. The most commonly used iterative algorithm is the alpha expansion [5]. From a theoretical point of view, this approximation ensures a labeling that is within a known factor of the global optimum [5]. From the practical point of view, the information that is used to build the graph is crucial in order to produce satisfying results. For correspondence problems, the data term takes the form

\[
D(\chi) = \sum_{p \in I_k} D_p(x^*_p),
\]

and is based on a similarity measurement between pixel \( p \) in image \( I_k \) and pixel \( p + \vec{x}_p \) in image \( I_{k+1} \), where \( \vec{x}_p \in L \) is a two dimensional vector representing the disparity of pixel \( p \). Examples are the sum of squared differences, correlation, hamming distance or the robust but computationally expensive SIFT or SURF features. The interaction term takes the form

\[
V(\chi) = \sum_{p,q \in N} V_{p,q}(\vec{x}_p, \vec{x}_q),
\]

and penalizes assigning different labels to neighboring pixels \( p,q \) in \( I_k \). Examples are the common Potts model \((V_{p,q} = 1 \cdot (\vec{x}_p \neq \vec{x}_q))\), or metric functions in the label space.

---

1 All computations that are presented in this paper have been obtained with an Intel(R) Core(TM)2 CPU with 2 GHz and 2 GB RAM.
We incorporate two additional constraints, namely a minimum Euclidean distance between two selected pixels and a (minimum) number of pixels to be found. In this way it is ensured that selected pixels are spread over the entire image in order to get a good estimate of the transformation (see Fig. 4). Note that textured pixels do not have to be detected accurately and systematically. An approximate detection is enough to reduce the number of nodes while minimizing the loss of information. Fig. 3 compares the results obtained by selecting 100 pixels equally rastered (left) and with the modified Harris corner detector (right). Both results have been computed with $\mathcal{V}(\chi) = 0$ in order to better visualize the influence of the choice of nodes on $\mathcal{P}(\chi)$. The equally rastered graph produces more outliers than the Harris based selection. When choosing the Harris based node selection, one has to face the problem of selecting an adequate neighborhood system due to the nodes no longer being in a fixed relationship to each other. Straight forward approaches could be to look for other selected nodes in a window centered around each selected node, or use k-nearest neighbors and set up edges between every such pair, weighted by their distance. The drawback of the windowed method is that the significance of the interaction term varies for each pixel depending on the number of neighbors within the window. Consider for example a node that is more likely to be labeled incorrectly, then the interaction penalty for assigning a false label will be less expensive when it has only a few distant neighbors and zero when it is not close enough to other pixels. The k-nearest neighbor approach on the other hand can result in an edge system with holes. We therefore propose a neighborhood system that takes the spatial distribution of the selected nodes into account. Using the selected nodes as basins, the watershed transform [6] computes the influence zone of each node, which can be used to generate a natural neighborhood system for the nodes. The k-nearest neighbor system. An edge between two nodes will be added to the graph if their influence zones are neighbors.

Another time critical issue is the large set of labels that needs to be considered to allow rotations of high angles or even large translations, because the alpha expansion’s computation time grows linearly with the size of the label set. While there are approaches to reduce the label set using coarse-to-fine strategies, such as [2], the performance varies with the transformations involved and therefore we have chosen to present the results in this paper using a fixed label set.

3. METHOD COMPARISON

Three reference images, extracted from real endoscopic exams (Fig. 1), have been used to evaluate the robustness of the sparse graph cut in comparison to a standard dense graph cut as well as to an equally rastered graph cut with the same number of nodes. In order to simulate endoscopic displacements, typically occurring transformations (translation, scaling and rotation) denoted by $H_{true}$, have been applied to the three images. After registering, the resulting labellings have been used to compute the transformation matrix $H_{est}$ (estimation of $H_{true}$). Note that there were no visible outliers in any of the labelings, therefore the estimation quality of $H_{est}$ for the three methods is comparable, although a robust fitting using RANSAC is generally advised to deal with outliers. As a measurement for the registration error, we use the mean Euclidean pixel deviation $\epsilon$ (the mean distance between homologous pixels of the target image and the source image transformed by $H_{est}$).

Best results are achieved with the dense graph cut which takes about 154 seconds to compute. Both the sparse graph cut and the equally rastered graph cut take 5 seconds to compute. At the same time, the results of the proposed sparse version are 4 up to 10 times more accurate. Table 1 shows the resulting error $\epsilon$ for three exemplary transformations.

Next, we put our algorithm in perspective by comparing its performance to the other methods and also to the mutual information based registration.

![Dense, Sparse, Harris](image1.png)

![Sparse, equally](image2.png)

![Sparse, equally](image3.png)

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Dense</th>
<th>Sparse, Harris</th>
<th>Sparse, equally</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 10^{5}$, $f = 1.2$</td>
<td>0.06</td>
<td>0.26</td>
<td>2.91</td>
</tr>
<tr>
<td>$\phi = 5^{5}$, $f = 1.1$</td>
<td>0.02</td>
<td>0.26</td>
<td>0.94</td>
</tr>
<tr>
<td>$(t_x, t_y) = (-4, 3)$</td>
<td>0.09</td>
<td>0.24</td>
<td>1.51</td>
</tr>
</tbody>
</table>

Table 1. Comparison of mean Euclidean pixel deviations for dense and sparse graph cuts on exemplary image transformations.

![Fig. 4. Selected nodes and watershed based neighborhood system. An edge between two nodes will be added to the graph if their influence zones are neighbors.](image4.png)

![Fig. 5. Pig bladder phantom. (a) indicates the camera movement and the imaged area for each acquisition. (b) result obtained with the mutual information based registration. (c) result obtained using sparse graph cuts. Note the missing visible image borders and the properly reconstructed structures on the left in (c).](image5.png)
registration quality with two other methods evaluated in [3], namely a mutual information based registration and a Baker and Matthews optical flow algorithm. Fig. 5(a) shows a pig bladder photograph overlayed with the simulated acquisition path and imaged area (for a detailed description of the evaluation setup, please refer to [3]). Fig. 5(b) shows the panoramic image built with the mutual information algorithm. Fig. 5(c) shows the panoramic image built with our sparse graph cut. The image built by the optical flow algorithm is virtually the same as the one built by mutual information. Note how in the image obtained with the graph cut version, the image borders at the top are not visible and the structures on the left are reconstructed properly, in contrast to the map obtained with mutual information. Fig. 6 quantifies these observations: during the registration process, the $\epsilon$ error remains constantly weak (below 0.3 pixels) for the graph cut based registration algorithm independent of the type of transformation, while the other two algorithms are less accurate (above 3 and 5 pixels respectively) when transformations other than translation are involved. Computation time of the graph cut method is slightly higher than that of the optical flow method and significantly lower than that of the mutual information method. As registration quality is the most critical part in the mosaicing process, the graph cut method provides the best results.

Finally, we have tested our algorithm on real (human) endoscopic video sequences. Fig. 7 shows a translation along the bladder wall while zooming in. While image borders between two registrations are mostly visible due to illumination changes, the structure of the bladder remains clearly visible and coherent.

4. DISCUSSION

We have shown that for image registration, the sparse graph cut method yields a computation speedup by an order of magnitude in comparison to a standard dense graph cut. At the same time, the negative effects caused by the node and edge reduction based on spatial image properties are negligible. One could argue that computation time is the least important criteria in the mosaicing process because bladder mosaics are used for a second diagnosis standardly performed after the examination [1]. Nevertheless, it is desirable to build panoramic images in a few minutes or even in real time. The sparse graph cut registration algorithm is a step towards this goal and produces constantly weak errors independent of the type of perspective transformation involved. Therefore, it is more suitable for building panoramic images from endoscopic video sequences than previously evaluated methods. The 2D mosaicing process is an important basis for 3D image cartography of the bladder wall on which our attention will be focused in future research.

5. REFERENCES


