Customizing Graph Cuts for Image Registration problems

Thomas Weibel[1,2], Christian Daul[1], Didier Wolf[1], Ronald Rösch[2]

[1]Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039, Nancy-University, CNRS)
2, avenue de la forêt de Haye, 54516 Vandœuvre-lès-Nancy, France

[2]Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM)
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

thomas.weibel@itwm.fraunhofer.de, christian.daul@ensem.inpl-nancy.fr,
didier.wolf@ensem.inpl-nancy.fr, ronald.roesch@itwm.fraunhofer.de

Résumé – Le recalage d'images peut être réalisé par des méthodes robustes et précises de minimisation d'énergie formulées sous la forme de “graphs cuts”. Ce papier donne un aperçu sur deux modifications de la méthode standard des graphs cuts. Ces modifications permettent de traiter des images faiblement texturées ou fortement bruitées. Dans le cas d'images texturées, nous montrons comment il est possible de diminuer les temps de calcul d'au moins un facteur dix (avec des “sparse graph cuts”) tout en préservant la précision de la méthode standard des graph cuts. Dans le cas d'images texturées fortement bruitées, nous montrons que l'incorporation d'un terme d'ordre élevé dans les graphs cuts permet de préserver la précision du recalage.

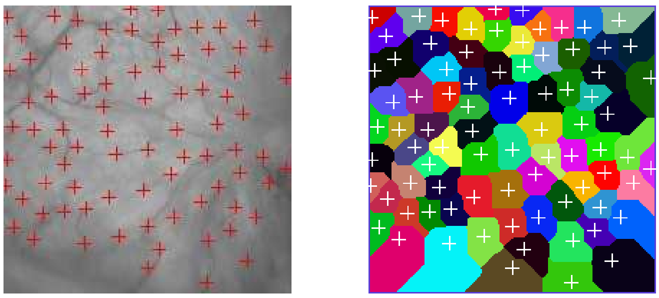
Abstract – Image registration can be formulated in terms of energy minimization, which can be efficiently and robustly solved using graph cuts. This paper gives an overview on how the standard graph cut approach can be modified to deal with weakly textured images and/or images affected by strong noise. First, we show how nodes in textured images can be selected sparsely so that the computation time decreases by at least one order of magnitude while maintaining the high registration accuracy. Secondly, we will explain how higher order terms can be incorporated into the graph cut formulation when significant noise levels prevent an accurate registration otherwise.

1 Introduction

Accurately estimating a perspective transformation between two overlapping images I and J of a planar scene usually involves detecting keypoints (such as SIFT[1] or SURF[2]) in both frames, which are then robustly matched (e.g. with RANSAC[3]). The task becomes difficult however when image primitives cannot be robustly detected, which is the case for weakly textured images and/or when strong noise is present in the images. The registration can then be formulated as an energy minimization problem and solved efficiently and robustly using a graph cut framework (see [4] for a detailed introduction). The energy function to be minimized is usually written as

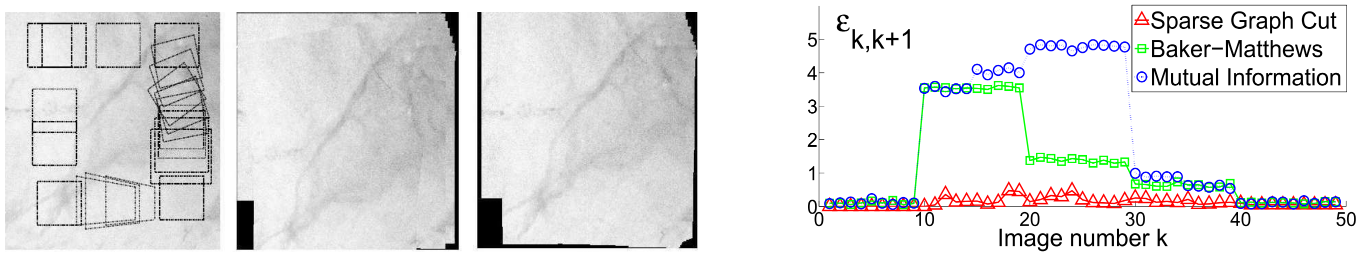
$$E(X) = \sum_{p\in I} D_p(x_p) + \lambda \sum_{(p,q\in N_p)\in I} V_{pq}(x_p, x_q), \qquad (1)$$

where $X : I \to L$ is a solution that maps each pixel $p\in I$ to a pixel $p'\in J$ by assigning a 2D displacement vector $x_p\in L$ (a fixed set of possible vectors), N_p is a neighborhood system for p and λ is a regularization parameter to adjust the smoothness of the solution. The data term D_p can be any similarity measurement between pixels, such as windowed Euclidean RGB difference. The interaction term V_{pq} penalizes assigning different displacement vectors to neighboring nodes. For registration problems, a linear or quadratic term gives good results. This paper describes two modifications of the standard graph cut formulation. First, we show how a significant computational speedup can be achieved while maintaining the registration quality. This is done by selecting a smaller subset $V \subset I$ based on local texture distribution. Then we show how in the presence of noise, higher order terms are able to recover the perspective transformation when the standard approach fails due to its simple pairwise nature.



a) b)

Fig. 1: **a)** Selected nodes using our algorithm for a frame of an endoscopic examination of the internal wall of the human bladder. **b)** the adjacency system based on the watershed transform.



a) b)

Fig. 2: Method comparison. **a)** From left to right: Pig bladder phantom with simulated endoscope movements indicated, mosaic by maximizing Mutual Information [5], mosaic using sparse graph cuts. **b)** Mean registration error for the three methods, as indicated in the box.

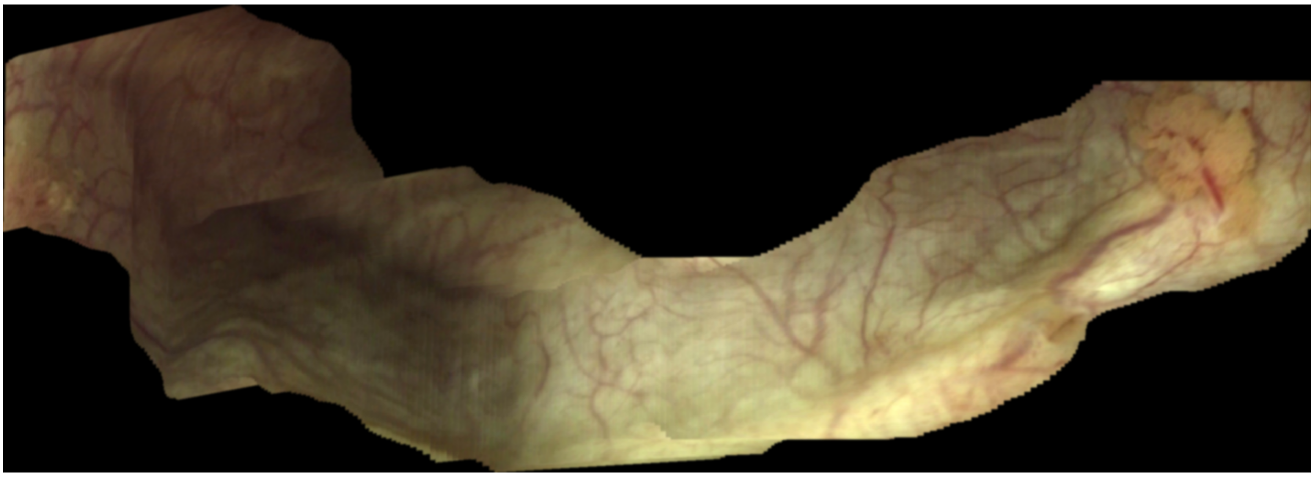
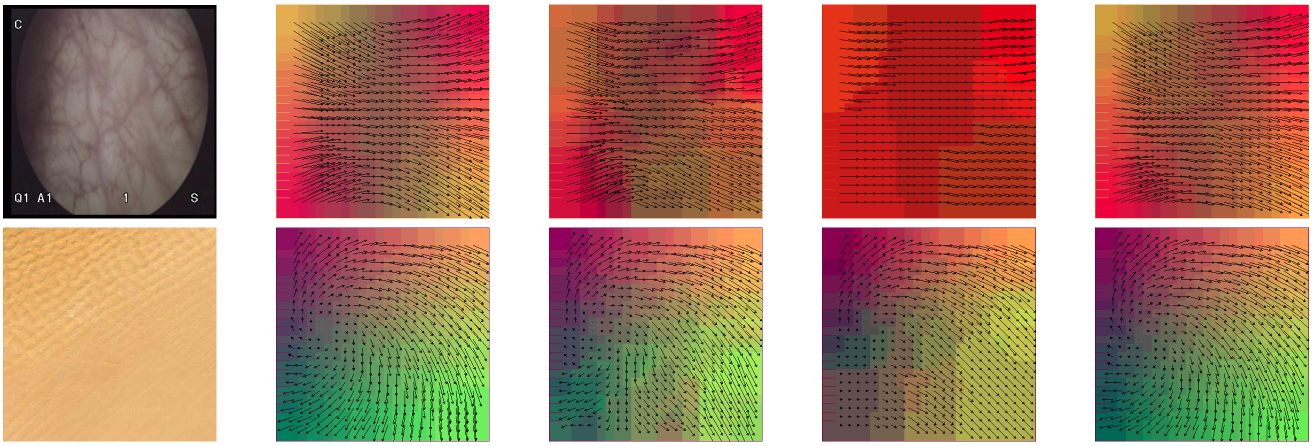


Fig. 3: Mosaic created from a real cystoscopic video sequence consisting of 450 images (such large field of views facilitate bladder cancer diagnosis).

2 Bladder Image Mosaicing using Sparse Graph Cuts

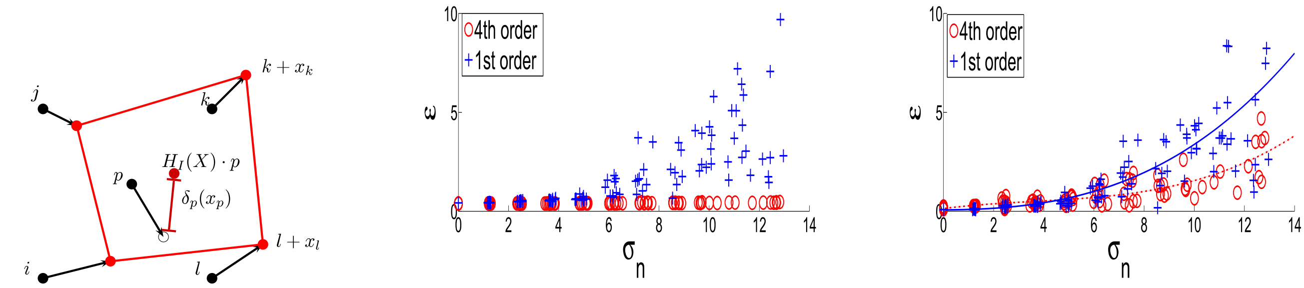
In [6], we have shown that graph cut based registration is qualitatively superior to previously evaluated methods (Mutual Information [5], Baker-Matthews optical flow [7, 8] for creating panoramic images (or mosaics) of the internal wall of the human bladder. It is however significantly slower than the fast Baker-Matthews algorithm. The key idea is that even though keypoints cannot be robustly detected in overlapping images to do the matching, they can be used to select regions with significant textural information. This allows to reduce the number of nodes used in the graph cut substantially, while the loss of information is minimized. We modify the Harris corner detector [9] to select pixels/nodes somewhat uniformly spread in I while assuring a minimum distance between them. This results in a well spread node selection in I, as can be seen in Fig. 1 a). For each node, its neighbors are found using the watershed transform. In the graph, an edge will be added between two nodes if their corresponding basins are neighbors. Fig. 1 b) shows this "natural" neighborhood system for the nodes in a).

A comparison of the registration accuracy of our sparse graph cut method with that of the two other methods is given in Fig. 2. A high resolution photograph of a pig bladder has been used to simulate known cystoscopic displacements typically occurring in bladder examinations (the perspective transformations are indicated by the black frames). The sparse graph cut method performs well in terms of the mean Euclidean pixel deviation ϵ from the ground truth (remains below 0.5 pixels throughout the sequence) independent of the type of transformation. In Fig. 2 b), $\epsilon_{k,k+1}$ is the mean error when registering image I_k on image I_{k+1}. The reader is referred to [6] for details. Fig. 3 shows a mosaic of a real examination computed with the sparse graph cut algorithm. Each registration is computed in about a second, whereas a dense approach needs several minutes, depending on the size of L.



a) b) c) d) e)

FIG. 4: Qualitative effects of the higher order term. **a)** One of two input images. **b)** Solution obtained with parameter $\lambda = 0.1$ of equation (1) and Gaussian noise ($\sigma_n = 1.5$). The pairwise terms are able to recover the true transformation accurately. **c)** As more noise ($\sigma_n = 7.5$) is added, the obtained solution becomes distorted **d)** $\lambda = 0.2$ leads to an over smoothing of the solution. **e)** Using equation (3) with parameters $\lambda = 0.1, \gamma = 2$ and $\sigma_n = 7.5$, a smooth and accurate solution close to the true transformation is obtained.



a) b) c)

FIG. 5: Higher order term: **a)** The Euclidean distance between $p + x_p$ and $H_I(X) \cdot p$ defines the cost $\delta_p(x_p)$ of the higher order term. $H_I(X)$ is the perspective transformation that maps the four corner pixels $i, j, k, l \in I$ to their positions indicated by the current solution X. **b)** Deviations ϵ from a fitted perspective transformation using least median of squares for increasing Gaussian noise. **c)** Deviations from the known perspective transformation for increasing Gaussian noise.

3 Registering noisy images with Higher Order Graph Cuts

As was shown in the previous section, graph cuts can be used to robustly estimate the perspective transformation between two weakly textured images. In the presence of strong noise however, the pairwise nature of (1) has its limits in accurately recovering perspective transformations. Fig. 4 a) shows two examples of weakly textured images (aerial footage of a desert, frame from a cystoscopic examination), representing I. J is generated by applying a ground truth perspective transformation to I. When adding weak Gaussian noise (with standard deviation $\sigma_n = 1.5$) on I and J, minimizing equation (1) accurately recovers the applied perspective transformation (the vector field in Fig. 4 b) is visually equivalent to the ground truth). With more noise added ($\sigma_n = 7.5$), the pairwise terms of equation (1) prevent the recovery of a globally smooth vector field. The result in Fig. 4 c) is locally smooth, but with visible boundaries and globally distorted. Increasing the regularization parameter λ of equation (1) leads to a heavy over smoothing, as can be seen in Fig. 4 d). Both Fig. 4 c) and d) do not conform to a perspective transformation. It is well known that pairwise terms are unable to appropriately represent the rich statistics present in natural images, and thanks to recent contributions [10, 11], higher order terms (involving more than two nodes) can be incorporated into the graph cut formulation. Using a $4th$ order term $\delta_p(x_p)$, we can enforce the result to be close to a perspective transformation:

$$\delta_p(x_p) = ||H_I(X) \cdot p - (p + x_p)||, \qquad (2)$$

where $H_I(X)$ is the perspective transformation that maps four non-coplanar reference pixels in image I to their positions indicated by the solution X. In our experiments, the four corner pixels in I are used as the reference pixels, Fig. 5 a) illustrates the geometry of the term (2). Note however that any four non-coplanar points can be chosen, as the effect of the term is due to the interaction between all nodes via these four. The higher order energy function is then defined as

$$E(X)=\sum_{p\in I} D_p(x_p)+\lambda \sum_{(p,q\in N_p)\in I} V_{pq}(x_p,x_q)+\gamma \sum_{p\in I}\delta_p(x_p), \quad (3)$$

solvable using the methods presented in [10] and [11]. Fig. 4 e) shows the result after minimizing this energy, for the same noisy image as in c) and d). The transformation is visually close to result obtained in the low noise case b). Fig. 5 b) shows the mean Euclidean pixel deviation ϵ from a perspective transformation fitted using least median of squares, with and without the perspective term of equation (3). The resulting solution using the perspective term is always conform to a perspective transformation. In Fig. 5 c), ϵ from the ground truth transformation is plotted. The perspective term attenuates the negative effects of increasing noise σ_n on registration accuracy.

4 Conclusions

In this paper, we have given an overview on how graph cuts can be customized to be used in estimating perspective transformations when image primitives cannot be robustly detected in overlapping images. When image quality is acceptable, a speedup of an order of magnitude without decreasing the accuracy can be achieved and potentially be used in real time applications. When image quality is severely affected by noise, higher order terms can be formulated to overcome the limitations of standard first order terms. Currently we extend these methods to be used in estimating 3D mosaics of endoscopic video sequences taken in cystoscopic examinations.

References

[1] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

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

[3] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] V. Kolmogorov and R. Zabih. What energy functions can be minimizedvia graph cuts? *IEEE transactions on pattern analysis and machine intelligence*, pages 147–159, 2004.

[5] 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. *Biomedical Engineering, IEEE Transactions on*, 55(2):541–553, 2008.

[6] T. Weibel, C. Daul, D. Wolf, R. Rösch, and A. Ben-Hamadou. Endoscopic bladder image registration using sparse graph cuts. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 157–160. IEEE, 2010.

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

[8] Y. Hernández-Mier, W. Blondel, C. Daul, D. Wolf, and F. Guillemin. Fast construction of panoramic images for cystoscopic exploration. *Computerized Medical Imaging and Graphics*, 34(7):579–592, 2010.

[9] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.

[10] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, 2007.

[11] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2993–3000. IEEE, 2009.