

TV- $\mathcal{L}1$ based fast and robust mosaicing of cystoscopic images

Sharib ALI^{1,2}, Walter BLONDEL^{1,2}, Christian DAUL^{1,2}

¹Université de Lorraine, CRAN, UMR 7039, 2 avenue de la Forêt de Haye, 54516 Vandœuvre-lès-Nancy cedex, France

²CNRS, CRAN, UMR 7039, 54516 Vandœuvre-lès-Nancy, France

{sharib.ali, walter.blondel, christian.daul}@univ-lorraine.fr

Résumé – L’augmentation du champ de vue par mosaïquage d’images facilite le diagnostic et le suivi de lésions vues dans une séquence cystoscopique. Alors que les méthodes de flot optique ont été utilisées pour des scènes très différentes, elles ont rarement été testées sur des images endoscopiques pauvres en textures et floutées. Dans cette contribution est présentée une méthode variationnelle totale (TV- $\mathcal{L}1$) d’estimation de flot optique pour ce type d’images. Cet algorithme est comparé à d’autres méthodes de flot optique à l’aide d’images reproduisant de façon très réaliste des données patient. Cette comparaison, ainsi que des résultats obtenus pour des données cliniques, mettent en lumière l’avantage de la méthode TV- $\mathcal{L}1$ en termes de précision, de robustesse et de vitesse de recalage d’images.

Abstract – Image mosaicing helps in diagnosis and follow-up of lesions in cystoscopic bladder image sequences by increasing the field of view. Optical flow methods have been used for various scenes, whereas they were barely experimented for data with poor texture and blurry images like that of endoscopic data. In this paper, we present the use of total variational optical flow method (TV- $\mathcal{L}1$) for such high texture variability and blur. This method is compared to other optical flow methods on phantom images simulating realistic patient data. This comparison and the result on patient data show the advantage of the TV- $\mathcal{L}1$ method over the other methods in terms of registration accuracy, robustness and time.

1 Introduction

Cystoscopy is the standard clinical examination allowing the urologist for searching lesions on the bladder internal wall and assessing their evolution. The small field of view (FOV) of the video-sequence images complicates lesion diagnosis and follow-up. Increasing the FOV by computing two-dimensional (2D) panoramic images (see fig. 2(d)) helps urologists to analyse the state of the overall surface of the bladder. The mosaicing process registers consecutive images I_{i+1} (source) and I_i (target) and places them in a common coordinate system (usually that of the first image I_0). It was shown that both in the white light (WL) modality [1, 2] and in the fluorescence modality [3] a perspective transformation is an appropriate model to register bladder images. In Eq. (1), $T_{i,i+1}$ has to be computed so that it superimposes homologous pixels $p_{i+1} = (x_{i+1}, y_{i+1})^T$ of I_{i+1} and $p_i = (x_i, y_i)^T$ of I_i . The parameters f , ϕ , (s_x, s_y) , (t_x, t_y) and $\{h_1, h_2\}$ denote the scale factor, in-plane rotation, shearing, 2D translation and perspective changes respectively. The value of parameter α is entirely defined by the perspective parameters h_1 and h_2 .

$$\begin{pmatrix} \alpha x_i \\ \alpha y_i \\ \alpha \end{pmatrix} = \begin{pmatrix} f \cos \phi & -s_x \sin \phi & t_x \\ s_y \sin \phi & f \cos \phi & t_y \\ h_1 & h_2 & 1 \end{pmatrix} \begin{pmatrix} x_{i+1} \\ y_{i+1} \\ 1 \end{pmatrix} = T_{i,i+1} \begin{pmatrix} x_{i+1} \\ y_{i+1} \\ 1 \end{pmatrix} \quad (1)$$

Feature descriptors computed for strong texture points can be used in the fluorescence modality to register bladder images [3]. However, in the standard WL modality, such an approach fails due to the strong intra- and inter-patient texture variability and/or due to the lack of texture. To solve this problem, some

authors used only iconic data to register the images by either maximizing the grey-level mutual information [1] or by using graph-cut techniques for minimizing the color differences of the superimposed pixels [2]. These methods are robust and accurate but very slow.

For many scenes, differential methods are often the best compromise in terms of registration accuracy and speed. This was also shown in [4] for bladder images in the WL modality. Two techniques exist depending on local or global energy minimization scheme. Local methods [5] are based on brightness consistency assumption in a given neighborhood of the pixels and are robust against noise. But, such methods lead to rank deficient matrices during optical flow calculations in homogeneous or low texture regions that often arise in cystoscopic image sequences. In contrary, global methods propagate the optical flow field from image regions with rich textures to image regions with poor textures. It guarantees a dense flow field since the whole image data is used. It consists of a data term and a regularization term. The optical flow calculation model proposed by Horn and Schunck [6] can be formulated as,

$$\min_{(u,v)} \left\{ \underbrace{\int_{\Omega} (I_x \cdot u + I_y \cdot v + I_t)^2}_{\text{data term}} + \lambda^2 \underbrace{\int_{\Omega} \|u\|^2 + \|v\|^2}_{\text{smoothness term}} \right\}, \quad (2)$$

where (I_x, I_y) are image derivatives and (u, v) are flow vectors w.r.t. x and y image axes respectively. I_t is the grey-level difference of the consecutive image pairs. λ defines the trade-off between the smoothness (regularization) and the data term.

The major drawback of this method is that, due to the regu-

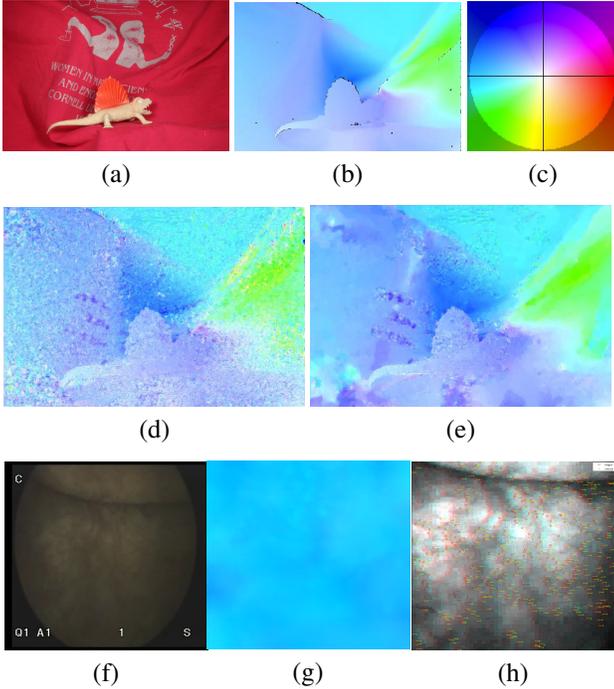


FIGURE 1 – Middlebury dataset for optical flow [8] is used for TV- $\mathcal{L}1$ optical flow method evaluation (a-e). TV- $\mathcal{L}1$ practiced on real cystoscopic data (f-g). The experiments were run for 7 warps using 150 iterations at each warp in a multi-resolution approach. λ was set to 30. a) Dimetrodon sequence. b) Ground Truth. c) Flow color-code representing the flow vectors in pixels. d) TV- $\mathcal{L}1$ without illumination term (*i.e.* $\gamma = 0$). e) TV- $\mathcal{L}1$ with ($\gamma = 0.02$) in the illumination term. f) Cystoscopic image. g) Flow plot using the flow color code with illumination term ($\gamma = 0.02$). h) Correspondences obtained from the flow vectors after RANSAC[9]. Red circles and + symbol in green represent the set of correspondence matches.

larization term, it penalizes high gradients of u and v resulting in over-smoothing effect. To overcome this problem, variational method with $\mathcal{L}1$ -norm was proposed and solved by using duality approach for faster convergence of the flow field [7]. In this paper, it is shown that the total variational method is robust for the cystoscopic images and hence can be used for fast, accurate and robust mosaicing of the bladder scene. In section 2, TV- $\mathcal{L}1$ method is explained with reference to middlebury optical flow dataset [8] and application on cystoscopic image sequence is shown. Section 3 deals with stitching of the images with the obtained local homographies. Section 4 presents the result of this method and compares it with the other existing optical flow methods using a dataset with known ground truth transformations.

2 TV- $\mathcal{L}1$ optical flow

The TV- $\mathcal{L}1$ optical flow estimation model is based on minimizing the function containing a data term using the $\mathcal{L}1$ -norm and

a regularization term using the total variation for the flow calculations. The major advantage of this method is that it allows discontinuities in the flow field and is less sensitive to homologous pixel intensity variations between images. Let $\mathbf{u} = \{u, v\}$ be the 2D displacement field at pixel coordinate $\mathbf{x} = \{x, y\}$. The optical flow problem linearized using Taylor’s expansion and with an additional function w for varying illumination (refer [7]) can be modeled as,

$$\rho(\mathbf{u}) = \nabla I_{i+1}(\mathbf{x} + \mathbf{u}^0) \cdot (\mathbf{u} - \mathbf{u}^0) + I_{i+1}(\mathbf{x} + \mathbf{u}^0) - I_i(\mathbf{x}) + \gamma w, \quad (3)$$

where I_{i+1} and I_i are the two consecutive source and target images respectively, \mathbf{u}^0 a close approximation to \mathbf{u} and γ is the weight. The $\mathcal{L}1$ penalization for both the regularization term and the data term can be understood as the minimization of the following energy function :

$$E = \min_{(u,v)} \left\{ \lambda \int_{\Omega} \|\rho(u,v)\|_1 + \int_{\Omega} \|u\|_1 + \|v\|_1 \right\}, \quad (4)$$

where λ gives the trade-off between data term and regularization term. A convex relaxation term is introduced to find the global minimum solution of the non-trivial Eq. (4) as,

$$E = \min_{(\mathbf{u}, \mathbf{p})} \left\{ \lambda \int_{\Omega} \rho(\mathbf{p}) + \int_{\Omega} \|u\| + \|v\| + \frac{1}{2\theta} \|\mathbf{u} - \mathbf{p}\|^2 \right\}, \quad (5)$$

where \mathbf{p} be another auxiliary variable as \mathbf{u} and θ be a small constant forcing to minimize the function.

Total variational ($\mathcal{L}1$ - norm) method using traditional optical flow constraint without illumination term γw in Eq. 3 was used in [10] for CT-lung and brain MRI image registration. Our experiments with the publically available “dimetrodon” sequence in Middlebury dataset [8] showed improved flow accuracy and faster convergence with the illumination term in Eq. (3) as shown in Fig.1(a-e). Experiments with the clinical data shown in Fig.1(f-h) also shows the efficiency of this algorithm. Geometrically consistent features were extracted from the calculated flow vectors. A multi-resolution approach was used for the experiments. The flow vectors were used to formulate the feature matches between the image pairs ($x_{i+1} \simeq x_i + u$, $y_{i+1} \simeq y_i + v$). The matched features were used to compute the perspective transformation matrix $T_{i,i+1}$ with an over determined system.

3 Homography Estimation

A set of local homographies $T_{i,i+1}^{2D}$ were obtained as explained in [11] from the matched features between the consecutive image pairs. The first image was defined as the reference frame I_0 . The global transformation $T_{0,i}^{2D}$ established from local transformations as in Eq. (6) was used to build the mosaiced images shown in Figs. 2(d), 3 and 4 by placing I_i in the coordinate system of I_0 .

$$T_{0,i}^{2D} = \prod_{k=0}^{i-1} T_{i-k-1, i-k}. \quad (6)$$

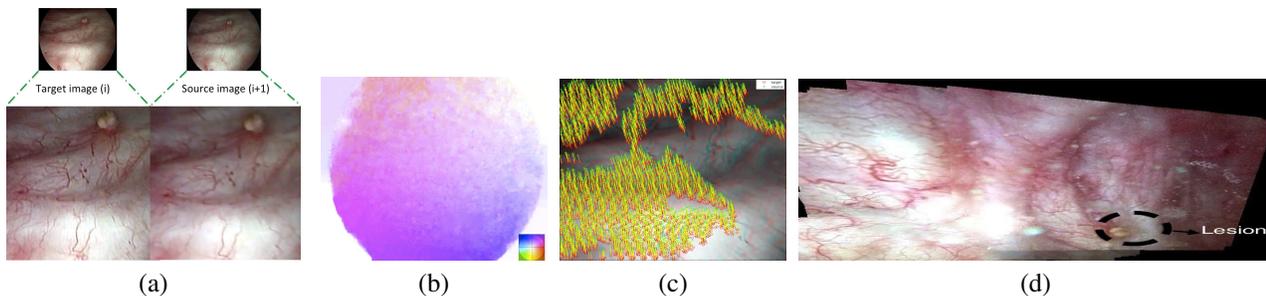


FIGURE 2 – Dense feature extraction using the total variational method with $\mathcal{L}1$ – *norm*. a) Cystoscopic image pair (I_i, I_{i+1}) with texture variability and blur. There is less texture in the bottom of the image than around the image center. b) Flow field representation using the color code given Fig.1(c). c) Geometrically consistent dense feature matches (yellow vectors) established from the flow vectors for the image pair of (a). d) Mosaicing obtained from a patient bladder video-sequence. The mosaic center is less textured than the remaining panoramic image parts.

4 Experimental Results

An image pair with varying texture, illumination and deformation is shown in Fig.2 (a). A flow field is computed using the method explained in section 2. Fig. 1(c) shows the dense correspondence established after removal of geometrically inconsistent flow. A video sequence with 250 similar images were selected to get a mosaic shown in Fig. 2 (d). In this section, we have presented and discussed the method used and compare it with some existing optical flow methods in the literature using the dataset explained below.

Dataset : A pig bladder (having very similar texture to the human bladder for urologist) was first excised, opened and then a high resolution image was acquired. A sub-image I_0 was chosen as a reference coordinate and 49 sub-images with known transformations $T_{i,i+1}^{true}$ were computed. Thus, a realistic cystoscopic medical examination was simulated with both blurry and poor texture conditions along with strong displacements of about 50 pixels and a strong perspective change. The mean registration error $\bar{\epsilon}$ obtained from registration error between the consecutive image pairs $\epsilon_{i,i+1}$ is calculated as :

$$\bar{\epsilon} = \frac{\sum_{i=0}^{i=49} \epsilon_{i,i+1}}{50}, \text{ with } \epsilon_{i,i+1} = \frac{1}{|I_i|} \sum_{p \in I_i} \| T_{i,i+1}^{true} p - T_{i,i+1}^{est} p \|. \quad (7)$$

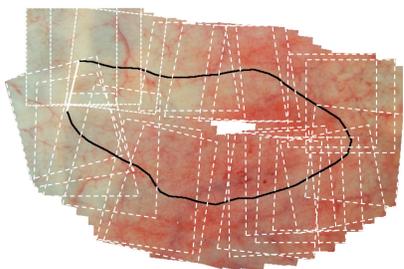


FIGURE 3 – Phantom bladder mosaicing using TV- $\mathcal{L}1$ method. The quadrangles represent the images and black line presents trajectory of the simulated video-sequence.

Eq. (7) places the pixels p from image I_{i+1} into the coordinate system of image I_i using both the estimated homography $T_{i,i+1}^{est}$ and ground truth values $T_{i,i+1}^{true}$. This error determined for $|I_i|$ number of points is ideally null for each I_i, I_{i+1} pair.

Method comparison : Baker and Mathews method gave a mean registration error ($\bar{\epsilon}$) of nearly 6 pixels. Global methods like Horn-Schunck method and Brox et. al. were not robust (registration failed for some images) and their accuracy was highly dependent on weighing parameters when the registration was possible. The over-smoothing of the flow vectors were persistent in most of the image pairs giving high mean registration error of nearly 20 pixels. TV- $\mathcal{L}1$ method with a function able to handle illumination changes as explained in section 2 gave robust, fast and accurate registration of the image pairs under different conditions of both texture and illumination. An $\bar{\epsilon}$ of 1.2 pixels was achieved with relatively far less computational time (8s) than the methods in comparison (refer Table. 1). The experiment was run for 3 warps and 150 iterations at each warp in multi-resolution approach. λ was set to 50 and γ as 0.01. A complete mosaic of the phantom image using TV- $\mathcal{L}1$ method is shown in Fig. 3. A visually coherent mosaic was built by minimizing texture discontinuities and color gradients with the method described in [12].

TABLE 1 – Method comparison. $\bar{\epsilon}$ and \bar{t} are mean values computed for the phantom data of Fig. 3. All algorithms were implemented in MATLAB and tested on a 2.7 GHz i7 processor.

Method	$\bar{\epsilon}$ (in pixels)	\bar{t} (in s)
Baker and Mathews [13]	6.2	60
Horn-Schunck [6]	17	30
Brox et. al. [14]	20	40
TV- $\mathcal{L}1$ [7]	1.2	8

5 Conclusion

Cystoscopic sequences under white light modality have strong intra- and inter- patient texture variability, illumination changes and blur which makes the registration more challenging than

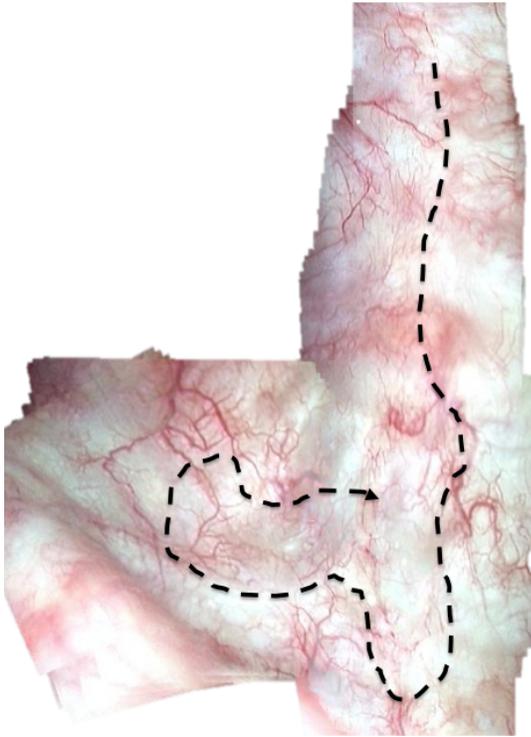


FIGURE 4 – TV- $\mathcal{L}1$ method used for mosaicing 500 images of a cystoscopic real patient data. The black dashed line defines the endoscope trajectory with an arrow pointing the direction of the loop completion.

the classical scenes usually employed to test optical flow methods. We experimentally verified that in presence of such high texture variability and illumination changes, some existing optical flow methods failed to register the image pairs robustly and accurately while total variational method was adapted to give robust and accurate registration of consecutive image pairs (I_i, I_{i+1}) for panorama building from such sequence.

Acknowledgement

The Ph.D. grant of S. Ali is financed by the French Research National Agency (ANR) under the framework of project “Cy-PaM2” (ANR-11-TECS-001).

Références

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

[2] T. Weibel, C. Daul, D. Wolf, R. Rösch, and F. Guillemin, “Graph based construction of textured large field of view mosaics for bladder cancer diagnosis,” *Pattern Recognition*, vol. 45, no. 12, pp. 4138–4150, 2012.

[3] A. Behrens, T. Stehle, S. Gross, and T. Aach, “Local and global panoramic imaging for fluorescence bladder endoscopy,” in *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Minneapolis, MN, USA, September 2009, pp. 6690–6693.

[4] Y. Hernandez-Mier, W. Blondel, C. Daul, D. Wolf, and F. Guillemin, “Fast construction of panoramic images for cystoscopic exploration,” *Computerized Medical Imaging and Graphics*, vol. 34, no. 7, pp. 579–592, 2010.

[5] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of the 7th international joint conference on Artificial intelligence - Vol. 2*, 1981, IJCAI’81, pp. 674–679.

[6] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[7] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[8] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[9] 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*, vol. 24, no. 6, pp. 381–395, 1981.

[10] T. Pock, M. Urschler, C. Zach, R. Beichel, and H. Bischof, “A duality based algorithm for TV- $\mathcal{L}1$ optical-flow image registration,” in *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention (MICCAI)*, 2007, pp. 511–518.

[11] S. Ali, C. Daul, T. Weibel, and W. Blondel, “Fast mosaicing of cystoscopic images from dense correspondence : combined surf and TV-L1 optical flow method,” in *IEEE International Conference on Image Processing, (ICIP)*, Melbourne, Australia, September 2013.

[12] T. Weibel, C. Daul, D. Wolf, and R. Rösch, “Contrast-enhancing seam detection and blending using graph cuts,” in *21st International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, November 2012, pp. 2732–2735.

[13] S. Baker, R. Gross, T. Ishikawa, and I. Matthews, “Lucaskanade 20 years on : A unifying framework : Part 2,” *International Journal of Computer Vision (IJCV)*, vol. 56, pp. 221–255, 2003.

[14] T. Brox, A. Bruhn, N. Papenber, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in *ECCV*, May 2004, vol. 3024 of *Lecture Notes in Computer Science*, pp. 25–36.