Skin image mosaicing: a comparative study of optical flow methods

Khuram FARAZ[1,2], Sharib ALI[1,2], Walter BLONDEL[1,2], Ernest GALBRUN[1,2], Marine AMOUROUX[1,2], Christian DAUL[1,2]

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

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

{firstname.lastname}@univ-lorraine.fr

Résumé – Le mosaïquage d'images constitue une solution d'intérêt pour construire automatiquement des panoramas (champs de vue étendus) de surfaces cutanées, particulièrement bien adaptée au diagnostic en télédermatoscopie. Cette étude exploratoire de faisabilité compare les performances de trois méthodes de flot optique (une méthode locale, une méthode globale basée sur les coupes de graphes et une méthode basée sur une approche variationnelle totale) appliquées au mosaïquage d'images de la peau. Des résultats quantitatifs et qualitatifs sont donnés, respectivement pour des fantômes avec une vérité terrain connue et pour des données patients.

Abstract – Skin image mosaicing is an important task, but only few work has been published in this domain. Large field of view mosaics of cutaneous surfaces facilitate diagnosis and is unavoidable in the context of tele-medicine. This explatory (and feasibility) study compares the performances of three optical flow methods (a local method, a global graph-cut method and a total variational approach) in the frame of skin image mosaicing. Both quantitative and qualitative results are given, respectively on phantom data with known ground truth and on patient data.

1 Introduction

Skin lesion diagnosis and follow-up are based on visual inspection by dermatologists. The latter are mainly located in urban areas. Therefore, telemedical approaches are developed to solve the economical and health problems of people with reduced mobility and/or living in remote areas. The interest and feasibility of tele-dermoscopy has been highlighted in recent studies [1]. Dermatologists need high resolution and quality images to perform reliable diagnosis. To facilitate scene interpretation and lesion follow-up, extended FOV images with high resolution can be obtained by superimposing the common parts of the limited FOV images using mosaicing algorithms. In practice, two strategies can be employed : 1) either the large FOVs are computed at the patient's home and the mosaic is tele-transmitted, or 2) the complete video-sequence is transmitted and the mosaic is build in the dermatologist's office. In both approaches, homologous image regions with few texture information have to be robustly superimposed.

Mosaics are built by registering consecutive image pairs of video-sequences. Only few work relating to skin image mosaicing have been published. Loewke et al. [2] used a local optical flow approach with a final error minimization with cross-correlation. This two step approach was used to mosaic skin images acquired under confocal microscopy where mainly 2D translations were dominant. The major drawback of this approach is that it is suitable to estimate only small 2D translations and in-plane rotations. Holmberg et al. [3] have shown the feasibility of skin image mosaicing based on textures at macro-scale. However, only simple transformations (2D translations, in-plane rotation and scale factor) are determined with their approach. In practice, more complete transformation (including viewpoint changes) have to be computed since hand-held camera displacements cannot be controlled.

This exploratory work compares the performance of three registration methods for building large FOV mosaics of skin surfaces. Challenging features of skin images lie in its variability (colour, hue, reflectence) and poor texture. This contribution analyses the feasibility of skin image registration with different optical flow (OF) algorithms. Section 2 first justifies the choice of the transformation linking geometrically two images, and then presents three OF methods (sections 2.1, 2.2 and 2.3) used to determine the parameters of this transformation. Sections 3 quantitatively compares the performance of the methods on human skin images and video sequences.

2 Algorithms for Image Registration

The aim of the registration algorithm is to superimpose the common parts of the consecutive image pairs (I_i, $I_i + 1$) of a sequence. I_i and I_{i+1} are target and source images respectively. The aim of the registration algorithm is to find the parameters of transformation $T_{i,i+1}$ superimposing homologous pixels p_{i+1} and p_i with coordinates $(x_{i+1}, y_{i+1})^T$ and $(x_i, y_i)^T$ in I_{i+1} and I_i respectively. Considering that the skin surfaces imaged are quasi-planar, a homography can be used to closely approximate the real transformation linking geometrically two consecutive images [4]. In Eq. (1), 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 parameters, 2D translation and perspective changes respectively. The value of parameter α is entirely defi-

ned by the perspective parameters h_1 and h_2.

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

Also, let $\mathbf{u} = (u, v)$ be the optical flow vector estimated at a point $\mathbf{x} = (x, y)$ for an image pair and $\mathbf{u}^0$ is the initial value of the flow vector field. All of these notations will be used in the latter sections.

2.1 Inverse compositional method

The inverse compositional algorithm described in [5] was used to minimize following sum of squared differences (SSD) based on brightness constancy between images I_i and I_{i+1} :

$$SSD = \sum_{\mathbf{x} \in I_i \cap I_{i+1}} \left[I_i \left(\mathcal{W}_{i,i+1}(\mathbf{x}; \Delta\mathbf{M}) \right) - I_{i+1} \left(\mathcal{W}_{i,i+1}(\mathbf{x}; \mathbf{M}) \right) \right]^2 , \quad (2)$$

where $\mathcal{W}(\mathbf{x}; \mathbf{M})$ is a warping function which transforms coordinates $\mathbf{x}$ using the components of vector $\mathbf{M}$ which correspond to the parameters of the transformation matrix $T_{i,i+1}$. Eq. (2) is minimized with respect to $\Delta\mathbf{M}$ by iteratively updating the components of $\mathbf{M}$ after each warp of the source image to the target image :

$$\mathcal{W}_{i,i+1} \left(\mathbf{x}; \mathbf{M} \right) \leftarrow \mathcal{W}_{i,i+1} \left(\mathbf{x}; \mathbf{M} \right) \circ \mathcal{W}_{i,i+1} \left(\mathbf{x}; \Delta\mathbf{M} \right)^{-1} , \quad (3)$$

with '$\circ$' representing the product. The solution of Eq. (2) is found by using a first order Taylor series approximation and least-square estimate for $\Delta\mathbf{M}$. The initial transformation is taken to be an identity matrix $(\mathcal{W}_{i,i+1}(\mathbf{x}; 0))$. Since target I_i is a fixed image, its spatial derivative can be calculated once for all before the iteration process.

Yahir et al. [6] have successfully adapted this algorithm for bladder image mosaicing. However, they have done an initial warping using $\mathcal{W}_{i,i+1}(\mathbf{x}; \Delta\mathbf{M})$, with vectors $\Delta\mathbf{M}$ corresponding to translations (t_x and t_y in Eq. (1)) with initial values computed from the cross-correlation between the preprocessed [7] image pair (I_i, I_{i+1}). This reduced the number of iterations for the algorithm convergence.

2.2 RFLOW variational energy minimization

Local methods (section 2.1) are computationally less expensive but may not be robust enough to register sequences with large homogeneous regions. Global methods propagate the OF 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. The model was first proposed by Horn and Schunck [8] such as :

$$\min_{\mathbf{u}} \int_\Omega \underbrace{\| \nabla I_{i+1}.\Delta\mathbf{u} + I_{i+1}(\mathbf{x} + \mathbf{u}^0) - I_i(\mathbf{x}) \|^2}_{data-term} d\Omega + \underbrace{\int_\Omega \| \nabla\mathbf{u} \|^2}_{regularizer} d\Omega \quad (4)$$

The data-term of Eq. (4) is the first order Taylor series expansion of the SSD represented in Eq. (2) with $T_{i,i+1}$ as identity matrix. This data-term assumes that the intensity of the homologous pixels in the two images remains constant over a small time interval. However, this impose a strong constraint in large displacement cases and in cases with view-point changes of the camera. Moreover, l^2-norm in the regularizer oversmooths the gradient, thus leading to inaccurate flow fields along the texture boundaries. To address that issue, Ali et al. [9, 10] proposed (i) a complementary structure constancy assumption and illumination compensating function in the data-term and (ii) an l^1-norm based energy minimization instead of l^2-norm. This energy, E in Eq. (5), was minimized using primal-dual approach in convex optimization and is detailed in [10].

$$\begin{aligned} E = \min_{\mathbf{u}, L} \int_\Omega \{ (1-\phi) \mid \nabla I_{i+1}.\Delta\mathbf{u} + I_{i+1}(\mathbf{x} + \mathbf{u}^0) - I_i(\mathbf{x}) \mid^1 + \\ \phi \mid \nabla S_{i+1}.\Delta\mathbf{u} + S_{i+1}(\mathbf{x} + \mathbf{u}^0) - S_i(\mathbf{x}) \mid^1 + \mid \gamma L \mid^1 \} d\Omega + \\ \int_\Omega \{ \mid \nabla\mathbf{u} \mid^1 + \mid \nabla L \mid^1 \} d\Omega \end{aligned} \quad (5)$$

S_i and S_{i+1}, representing the structural information, were computed using the Eigenvalue information as detailed in [9, 10] and illumination difference function L was ponderated by the factor γ estimated from singular values as contrast information of the images. The weighting factor ϕ was empirically chosen to be 0.25. The robustness were justified for both intra and inter-patient cases which is a strong motivation of our choice for skin video mosaicing.

2.3 Graph-cut based method

Robust and accurate bladder image registration was achieved with a graph-cut method in [11, 12]. The data-term used both color and key-point information. Classical gradient based smoothness term was also used for consecutive image pairs. The energy to minimize is thus represented as :

$$E(\mathbf{u}) = E_{color}(\mathbf{u}) + \beta E_{keypoints}(\mathbf{u}) + \lambda_s E_{smooth}(\mathbf{u}) \quad (6)$$

Weibel et al. [11] use RGB triplets at each node of a 10×10 pixel grid to approximate the homography $T_{i,i+1}$ relating images I_i and I_{i+1}. Let $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ be the nodes forming a triangular image region $\Delta_{\mathbf{pqr}}$ with $\mid \Delta_{\mathbf{pqr}} \mid$ pixels and centered on $\mathbf{x}$, then :

$$E_{color}(\mathbf{x}) = \frac{1}{\mid \Delta_{\mathbf{pqr}} \mid} \sum_{\mathbf{x} \in \Delta_{\mathbf{pqr}}} \| I_i(\mathbf{x}) - I_{i,i+1}(T_{i,i+1}[(\mathbf{x} + \mathbf{u}) \;\; 1]^T) \|^2 . \quad (7)$$

$$E_{smooth} = \sum_{\mathbf{x} \in \mathbf{p},\mathbf{q}} \frac{1}{\| \mathbf{p} - \mathbf{q} \|^2} \| \mathbf{u_p} - \mathbf{u_q} \|_2^2 \quad (8)$$

The key-points are extracted using SURF [13] for an initialization to the minimization of energy E in Eq. (6). The vector field is regularized using the normalized squared l^2-norm of the flow field gradient between two nodes of triplet $\Delta_{\mathbf{pqr}}$ as formulated in Eq. (8). This allows for obtaining piece-wise smooth flow field for approximating homography $T_{i,i+1}$.

3 Results and discussion

3.1 Dataset

A high resolution image of the vertebral dorsal (type IV skin according to Fitzpatrick scale [14]) was taken as test "refe-

Method	TRE (in pixel)			FLE (in pixel)	$\bar{t}$ (in s)
	min	max	mean		
Inv. Comp. with FCC [6]	1.45	5.64	3.57	19.30	**0.1**
RFlow method [9]	0.20	**2.23**	**0.67**	**5.02**	3
Graph-cut method [11]	**0.18**	2.30	0.778	7.50	15

TABLE 1 – Method comparison on the image sequence with 20 image pairs simulating small displacements (protocol I). Panorama size of 839×433 pixels was obtained by registering images with a size of 400×400 pixels. $\bar{t}$ represents the average registration time for CPU implementation of the methods.

rence". Two sets of image sequences, under different protocols, were extracted with known ground truth homographies $T_{i,i+1}^{true}$ (see the quadrangles in Fig. 1 representing the first two sub-images, with known homography between them, extracted from a high resolution image). 20 image pairs extracted with protocol-I have small translations (upto 20 pixels) and in-plane rotation ($\pm 5^{\circ}$) between them. Protocol-II, introducing large translations of upto 50 pixels, strong in-plane rotation of $\pm 15^{\circ}$ and perspective changes of $\pm 10^{-5}$, was used to extract another 48 image pairs. Fudicial points placed on the parent image were used to evaluate registration errors.

3.2 Results

Two criteria were used for comparing the methods under study : 1) the target registration error (**TRE**) defined as :

$$TRE = \frac{1}{N} \sum_{p \in I_i \cap I_{i+1}} \| T_{i,i+1}^{true} p - T_{i,i+1} p \|^2, \qquad (9)$$

where N is the total number of pixels p in the overlapped image region and 2) the fudicial landmark error (**FLE**) computed as the Euclidean distance between the centroids of the true landmark position and that in the mosaic. The landmark closest to the last frame of the sequence is selected for this purpose. Thus, the TRE and FLE represent "local" and "global" registration errors repectively.

The inverse compositional approach [6] resulted in a large mean TRE of nearly 4 pixels and mosaicing error of approximately 20 pixels (refer Table 1). This led to a perceptible misalignment as shown in Fig. 1. Both the graph-cut and the RFLOW methods proved to be robust and accurate, giving FLE of only

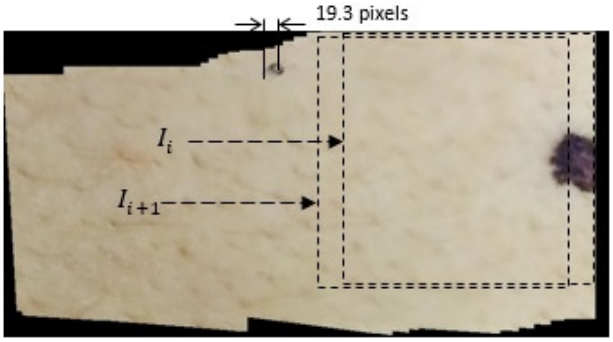

FIGURE 1 – Protocol I mosaic with inverse compositional algorithm using FCC in [6]. Large errors of 20 pixels are seen at the fudicial landmarks marked in the mosaic.

Method	TRE (in pixel)			FLE (in pixel)	$\bar{t}$ (in s)
	min	max	mean		
RFlow method [9]	**0.15**	**2.23**	**0.70**	**30**	**4**
Graph-cut method [11]	0.18	4.72	0.84	36	20

TABLE 2 – Comparison between the two most robust methods on 48 image pairs (protocol II). The mosaic was within a 691×911 pixels frame.

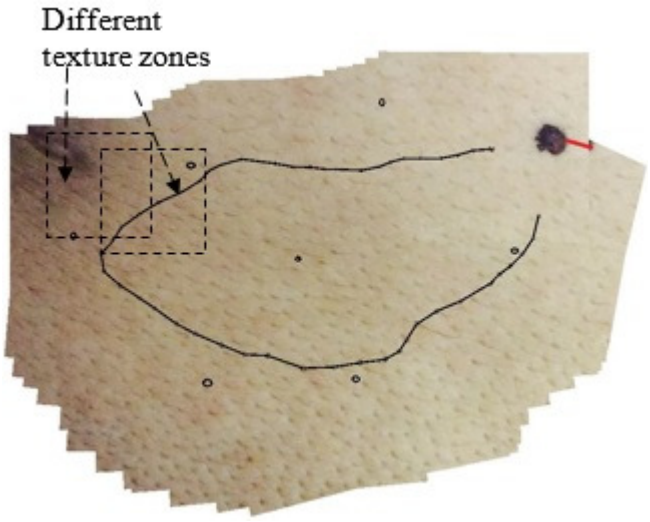

FIGURE 2 – Protocol II mosaic with the RFLOW method [9, 10] on a closed loop. The red mark represents the fudicial errors (FLE) between the start and end fudicial mark represented as large black spot.

7.50 and 5.02 pixels respectively. Mean TRE below 1 pixel was noted for both of these methods. However, the mean registration time $\bar{t}$ of the RFLOW method is five times smaller than that of the graph-cut method.

A more rigorous test was done with the images acquired under Protocol-II. The inverse compositional method was not robust enough to register all the image pairs (only the first 12 image pairs were registered with small TRE). Thus, the result with a complete sequence are only given for the graph-cut and the RFLOW methods in Table 2. The RFLOW method exhibited better local alignment accuracy (low TRE) than the graph-cut method and gave an aggregated global mosaicing error (FLE) of nearly 30 pixels. The misalignments in the mosaic were not visible at the small fudicial landmarks for both methods. However, a perceptible displacement of the large fudicial landmark between the end and the beginning of the mosaic (closing loop) can be seen as a short solid line in Fig. 2. In case of the graph-cut method, the TRE in the upper left region in Fig. 2 is relatively higher as a result of large texture variability and global intensity differences between image pairs in this region. This resulted in maximum TRE upto 5 pixels for graph-cut method while RFLOW gave more accurate result leading to only 2.2 pixels maximum error. A remarkable gain in average registration time is achieved with the RFLOW method in comparison to the graph-cut method under CPU implementation.

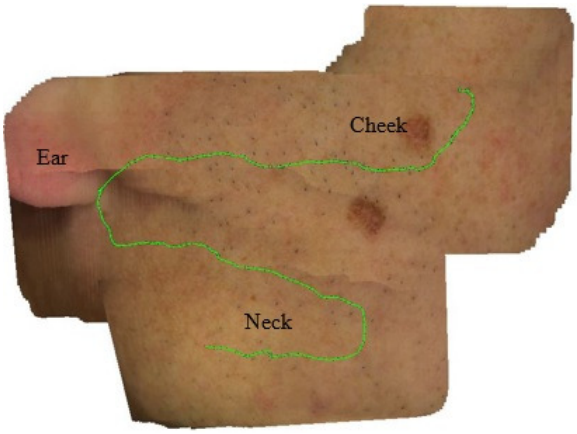

FIGURE 3 – Patient data mosaic of face and neck region. The green line represents the trajectory of the camera.

Fig. 3 gives another representative result in terms of skin mosaics obtained from a video-sequence acquired on the left facial part with some neck region on a patient. While no ground truth homographies are available, it allows for a qualitative evaluation of the RFLOW algorithm. As seen in Fig. 3, the RFLOW method is robust enough to lead to visually coherent mosaics even if the geometrical information visualized in the images is not quasi-planar as in the region including the border between the cheek and neck.

4 Conclusion and perspectives

This exploratory contribution presents a comparative study of three state-of-the-art methods initially developed for bladder mosaicing and adapted to the registration of cutaneous surfaces. The initial results of the feasibility tests of such mosaicing are promising. This analysis will help us develop dedicated algorithms for skin image mosaicing. The principle observation of this study is that the computation of OF with a total variational l^1 approach provides the best compromise between registration robustness, accuracy and time. Graph-cut based methods, on the other hand, are robust and accurate but too slow. In contrast, local OF methods are fast but suffer from lack of robustness for registration of skin images with high texture variability.

Acknowledgement

PhD grant of K. Faraz is co-funded by the European Regional Development Funds (FEDER) and the Conseil Régional de Lorraine (Regional Council of Lorraine) in the framework of project InnovaTICs-Dépendance. PhD grant of S. Ali is co-funded by the Agence Nationale de la Recherche (National Research Agency) and the Conseil Régional de Lorraine (Regional Council of Lorraine) in the framework of project CyPaM2 ANR-11-TECS-001

References

[1] C. Massone, A.G.G. Brunasso, T.M. Campbell, and H.P. Soyer, "Mobile teledermoscopy–melanoma diagnosis by one click ?," *Semi. in Cutananeous Medicine and Surgery*, vol. 28, no. 3, pp. 203–5, 2009.

[2] K. Loewke, D. Camarillo, W. Piyawattanametha, D. Breeden, and K. Salisbury, "Real-time image mosaicing with a hand-held dual-axes confocal microscope," in *Proc. SPIE*, 2008, vol. 6851.

[3] B. Holmberg and H. Lanshammar, "Possibilities of texture based motion analysis," *Comp. Methods and Programs in Biomed.*, vol. 84, no. 1, pp. 1 – 10, 2006.

[4] 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 Int. Conf. on Im. Proc., (ICIP)*, 2013, pp. 1291–95.

[5] S. Baker, R. Gross, T. Ishikawa, and I. Matthews, "Lucas-kanade 20 years on : A unifying framework : Part 2," *Int. Journal of Comp. Vis. (IJCV)*, vol. 56, pp. 221–255, 2003.

[6] Y. Hernandez-Mier, W. Blondel, C. Daul, D. Wolf, and F. Guillemin, "Fast construction of panoramic images for cystoscopic exploration," *Comp. Med. Imag. and Graph.*, vol. 34, no. 7, pp. 579–592, 2010.

[7] R. Miranda-Luna, Y. Hernandez-Mier, C. Daul, W. Blondel, and D. Wolf, "Mosaicing of medical video-endoscopic images : data quality improvement and algorithm testing," in *Elec. and Elect. Eng., 2004. (ICEEE). 1st Int. Conf. on*, 2004, pp. 530–535.

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

[9] S. Ali, C. Daul, and W. Blondel, "Robust and accurate optical flow estimation for weak texture and varying illumination conditions : Application to cystoscopy," in *Int. Conf. on Im. Proc. Theory, Tools and Appli. (IPTA)*, 2014.

[10] S. Ali, C. Daul, E. Galbrun, M. Amouroux W. Blondel, and F. Guillemin, "Robust bladder image registration by redefining data-term in total variational approach," in *Medical Imaging : Image Proc., SPIE*, 2015.

[11] 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.

[12] T. Weibel, C. Daul, D. Wolf, and R. Rösch, "Contrast-enhancing seam detection and blending using graph cuts," in *Int. Conf. on Pat. Rec. (ICPR)*, 2012, pp. 2732–2735.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comp. Vis. and Image Under. (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[14] T.B. Fitzpatrick, "The validity and practicality of sun reactive skin types I through VI," *Arch Dermatol*, vol. 124, no. 6, pp. 869–871, 1988.