Image Registration Algorithm based on Regular Sparse Correspondences and SIFT

Byeongyong Ahn, Hyung Il Koo† and Nam Ik Cho

* INMC, Department of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea.
E-mail: clannad@ispl.snu.ac.kr, nicho@snu.ac.kr Tel: +82-2-8808480
† Department of Electrical and Computer Engineering, Ajou University, Suwon, Republic of Korea.
E-mail: hikoo@ajou.ac.kr

Abstract—Image registration is to find the correspondences between different images of the same scene. It still remains a challenging problem especially when there is dramatic changes of object appearances. This paper presents a new image registration method for alleviating this problem, which first finds sparse correspondences and interpolates continuous dense motions from them. Unlike conventional registration methods, we regularly place control points to prevent biased distribution of feature points and find optimized correspondences by minimizing the cost function. The cost function is based on SIFT descriptors and considers the smoothness of motion and topological relations. Especially, the topological term prevents inconsistent solution like fold-over or duplication artifacts. For the optimization, we adopt a dual-layer belief propagation and coarse-to-fine scheme. Based on barycentric coordinates, we finally estimate dense motion from sparse correspondences. Experimental results show that the proposed method yields more plausible results and is computationally efficient.

I. INTRODUCTION

Image registration is the problem of estimating the geometric relation between two images of the same scene. This problem is considered an essential step for many computer vision applications such as motion estimation, stereo matching and image stitching, and has been studied for decades [1], [2]. However, it still remains a challenging problem due to the dramatic changes of the appearances of targets: targets experience geometric and photometric changes caused by pose variations, motion blurs, and surface deformations. Although robust object representation methods such as [3], [4] have been proposed to deal with these challenges, they are limited to salient and similar objects.

A. Related work

Conventional image registration methods can be classified into two categories: sparse methods and dense methods. Sparse methods [2], [5] detect sparse correspondences and derive a dense mapping from the detected sparse matches. They are efficient and can handle large displacements between the views. However, they may suffer from the biased distribution of feature points. That is, in low textured regions, the number of extracted feature points is not sufficient and the interpolated motions become unreliable (especially when the motion is a non-rigid one). Furthermore, the locality of feature description may yield lots of outliers, and the overall estimation process becomes complicated. On the other hand, dense methods [6], [7] directly find a dense mapping by optimizing a single cost function that considers all pixels in given images. These methods can overcome the limitation of sparse methods: they can handle non-continuous motions and can achieve better registration accuracy in low-textured regions. However, the complexity of such methods is usually very high, since all the pixels are considered in the optimization. Although adopting greedy algorithms for optimization may alleviate the complexity issues, they are prone to local minimas.

B. Our approach

In order to alleviate the limitations of conventional methods, we develop a new registration method combining the advantages of sparse and dense methods. To be precise, we first place control points regularly in order to handle low-textured regions with high accuracy while reducing the computational cost. Then, we develop a new cost function in order to find the correspondences of the control points. The function consists of data term and pairwise term: we adopt SIFT descriptor-based similarity measures for the data term and consider smoothness and topological constraint in the pairwise term. Actually, SIFT descriptors were already adopted as a similarity metric in the SIFT flow algorithm [8], which is one of the state-of-the-art dense registration methods. However, the algorithm suffers from the limitations of dense methods, i.e., high complexity and local minima. Therefore, we modify the algorithm into a sparse method.

The rest of paper is organized as follows. We present the proposed energy function whose minimization yields optimal sparse correspondences in Section II. In section III, we show how to interpolate dense correspondences from the sparse correspondences. Section IV presents the experimental result of our registration algorithm and we conclude this paper in Section V.

II. OPTIMIZED SPARSE CORRESPONDENCES

In conventional sparse methods, the registration performance highly depends on the distribution of the extracted feature points (i.e., they may fail to handle homogeneous regions, since only a small number of features are available). In order to alleviate this problem and obtain reliable results even on homogeneous regions, our method adopts regularly distributed control points. As shown in Fig. 1 (a), we place
control points with the horizontal and vertical spacing $\Delta$ on a source image $I_1$ and find corresponding points on the other image $I_2$ as shown in Fig. 1 (b).

In order to get reliable sparse correspondences, we develop a cost function considering interactions between neighboring control points as well as similarity of local regions. Let $P$ be a set of control points, $p = (x(p), y(p))$ be the coordinate of the point in $P$, and $w(p) = (u(p), v(p))$ be the motion vector of $p$. Then, the cost function is given by:

$$ E(w) = \sum_{p \in P} V_p(w(p)) + \sum_{p \in P} \sum_{q \sim p} (R_{p,q}(w(p), w(q))) + T_{p,q}(w(p), w(q)) $$

(1)

where the first and second term are the data and smoothness term respectively, which will be explained in the next subsections. Here, we adopt a first-order neighborhood system in order to consider the spatial regularity and $q \sim p$ indicates that $q$ and $p$ are adjacent (we adopt a 4 neighborhood system in this paper).

A. Data term

The data term measures the similarity between two local regions, i.e., $V_p(w(p))$ becomes small when $p$ in $I_1$ and $p + w(p)$ in $I_2$ are corresponding points, and vice versa. Since it is desirable to use robust metric that works in the presence of photometric and geometric changes, we adopt SIFT descriptors as [8]. Let us denote $s_1$ and $s_2$ as two SIFT images of $I_1$ and $I_2$ respectively. Then, the data term is given by

$$ V_p(w(p)) = \min(\|s_1(p) - s_2(p + w(p))\|_1, \tau_1) $$

(2)

where $\tau_1$ is a threshold that allows robust matching.

B. Smoothness term

In order to encourage smooth motions, we introduce the smoothness term:

$$ R_{p,q}(w(p), w(q)) = \min(|u(p) - u(q)|, \tau_2) + \min(|v(p) - v(q)|, \tau_2) $$

(3)

where we use a robust error function to allow motion discontinuities on the object boundaries. Moreover we separate the horizontal flow $u(p)$ from vertical flow $v(p)$ to adopt the dual-layer loopy belief propagation in the optimization [9].

C. Topology term

Although the smoothness term improves the overall registration performances, the cost function consisting of the data and smoothness terms may yield inconsistent results (e.g., one-to-many correspondences) as shown in Fig. 2. In order to prevent such results, we also introduce the topological term that imposes hard constrains. To express the constraint in a tractable way, we assume that the rotation is small and focus on the orders of control points:

$$ \begin{align*}
\text{if } x(p) > x(q), & \quad x(p) + u(p) > x(q) + u(q) \\
\text{if } y(p) > y(q), & \quad y(p) + v(p) > y(q) + v(q)
\end{align*} $$

(4)

Based on these conditions, the topological term is given by

$$ T_{p,q}(w(p), w(q)) = \begin{cases} 0 & \text{if two conditions are satisfied,} \\
\infty & \text{otherwise.}
\end{cases} $$

(6)

D. Optimization

For the efficient minimization of the cost function in (1), we adopt the dual-layer loopy belief propagation [9] and the complexity of the algorithm is reduced from $O(L^4)$ to $O(L^2)$ per iteration compared with the conventional belief propagation algorithm, where $L$ is the number of possible states for $u(p)$ and $v(p)$. We also adopt coarse-to-fine matching scheme for the further complexity reduction. However, unlike conventional $L1$ norm minimization problem, we could not adopt the distance transform function [10], since the topology term cannot be expressed in terms of the difference of motion. The minimization of (1) yields the optimal motion vectors $\hat{w}(p)$ for $p \in P$, and we can derive the optimal correspondences between $p$ on $I_1$ and $p + \hat{w}(p)$ on $I_2$.

III. DENSE CORRESPONDENCES ESTIMATION

Estimating dense correspondence from sparse correspondences is an essential step for sparse methods. Several algo-
algorithms have been proposed to solve this problem and two popular interpolation methods are compared in [11]: barycentric coordinates (BC)[12] and thin plate splines (TPS)[13]. TPS is inspired by the deformation of thin-plate and consists of two functions \( f(x, y), g(x, y) \) where each function is derived by minimizing the bending energy. On the other hand, BC provides efficient interpolation based on area-based coordinate system. The authors in [11] concluded that TPS provides higher accuracy but the gain is not so high when considering the computation cost of TPS. Based on this conclusion, we adopt barycentric coordinates to our algorithm. Since control points are distributed on a rectangular grid, we use the generalized barycentric transform [14] rather than conventional one.

Let \( p_i \) (\( i = 1, 2, 3, 4 \)) be the control points of a unit rectangle in a source image surrounding a point \( x \), \( y \) be the corresponding point to \( x \), and \( q_i = p_i + w(p_i) \) be the sparse correspondences of \( p_i \) as illustrated in Fig. 3. In a generalized barycentric transform, \( y \) inside the quadrangle \( \square q_1q_2q_3q_4 \) can be expressed by linear combination of the vertices \( q_i \).

\[
y = \sum_{i=1}^{4} w_i q_i
\]

(7)

where

\[
\omega_i = \frac{\cot(\gamma_i) + \cot(\delta_i)}{\|q_i - y\|^2}
\]

(8)

and \( \gamma_i \) are \( \delta_i \) are illustrated in Fig. 3. We can compute cotangent term using a ratio between dot and cross products.

\[
\cot(\gamma_i) = \frac{(y - q_i) \cdot (q_{i+1} - q_i)}{\|y - q_i\| \times \|q_{i+1} - q_i\|}
\]

(9)

\[
\cot(\delta_i) = \frac{(y - q_i) \times (q_{i-1} - q_i)}{\|y - q_i\| \times \|q_{i-1} - q_i\|}
\]

(10)

Finally, we get a pixel value at \( y = \sum_{i=1}^{4} \omega_i q_i \) in the warped image from a pixel value at \( x = \sum_{i=1}^{4} \omega_i p_i \) in the source image.

IV. EXPERIMENTAL RESULTS

We evaluate our registration method on several image pairs. In all experiments, we set \( \alpha = 4, \Delta = 4, \tau_1 = 4000, \) and \( \tau_2 = 200 \). In order to compare our method with the SIFT flow [8], we test our algorithm on a variety of images and some results are shown in Fig. 4. As shown in Fig. 4-(c) and (d), both methods can estimate the dense correspondence well in many test images. However, the SIFT flow algorithm sometimes suffers from the duplication artifacts as shown in metal, Mars, and toy images. On the other hand, our method yields more pleasing results. Moreover, our approach is more efficient. Our algorithm is implemented using Matlab MEX interface, and the computation time is about 2 seconds for \( 640 \times 480 \) images on a common FC with Intel(R) Core(TM) i5-3570 CPU @ 3.40GHz. On the other hand, conventional SIFT flow method takes about 10 seconds in the same environment. Table 1 shows computational time for several cases.

<table>
<thead>
<tr>
<th>Image size(pixel)</th>
<th>SIFT flow</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>600 \times 450</td>
<td>10.79s</td>
<td>2.44s</td>
</tr>
<tr>
<td>500 \times 500</td>
<td>9.72s</td>
<td>2.21s</td>
</tr>
<tr>
<td>900 \times 832</td>
<td>30.03s</td>
<td>8.859s</td>
</tr>
<tr>
<td>454 \times 306</td>
<td>5.318s</td>
<td>1.164s</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, we have proposed a new image registration method that incorporates the idea of SIFT flow algorithm into sparse methods. To be precise, we employ SIFT descriptor for robust local similarity metrics, and by applying regularly distributed control points, we could reduce the complexity of energy optimization problem while achieving high registration performances. For the efficient inference, we adopted dual-layer belief propagation algorithm and coarse-to-fine scheme. Experimental results showed that the proposed method is efficient and yields more pleasing results.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Ministry of Science, ICT and Future Planning (MSIP) (No. 2009-0083495).

REFERENCES

Fig. 4. Experimental results on Mars, metal, brain, toy images (from top to bottom) (a) Source image, (b) Target image, (c) Registered source image by conventional SIFT flow algorithm. (d) Registered source image by proposed algorithm. Red circles in SIFT results denote the regions with artifact.