LETTER

Kernel Based Image Registration Incorporating with Both Feature and Intensity Matching

Quan MIAO[†a)], ***Student Member***, **Guijin WANG**[†b)], ***Member, and*** **Xinggang LIN**[†c)], ***Nonmember***

SUMMARY Image sequence registration has attracted increasing attention due to its significance in image processing and computer vision. In this paper, we put forward a new kernel based image registration approach, combining both feature-based and intensity-based methods. The proposed algorithm consists of two steps. The first step utilizes feature points to roughly estimate a motion parameter between successive frames; the second step applies our kernel based idea to align all the frames to the reference frame (typically the first frame). Experimental results using both synthetic and real image sequences demonstrate that our approach can automatically register all the image frames and be robust against illumination change, occlusion and image noise.
key words: *image registration, kernel function, illumination change, occlusion, noise*

1. Introduction

The goals of image sequence registration are to estimate the motion parameter of each frame and align all the frames to the first frame. It is a fundamental and crucial module for various image applications, such as target tracking, panorama, image mosaicing and super-resolution. Thus it is a great challenge to design robust image registration methods that can cope with variations due to changes in illumination, partial occlusion and image noise. Recently, a wide repertoire of registration techniques [1] has been presented, which can be classified into two categories: feature-based and intensity-based methods.

The feature-based approach extracts reliable image features and employs feature correspondences to estimate motion parameters [2]–[4]. Local features based on invariant descriptors are most commonly used. The main advantage of the feature-based method is its relative robustness against changes in the object. However, coarse accuracy may be attained because establishing correspondences is non-trivial and only a small portion of available image intensity information is used [5].

As for the intensity-based technique, windows in high-variance areas are utilized [6], [7]. In contrast to feature-based approach, the intensity-based approach can achieve better registration accuracy. In [8], an intensity-based method is proposed by excluding pixels that have inadequately aligned images. However, the ROI must be chosen manually since the selection of ROI is not a trivial task; over-sparse mask will result in failing registration. In addition, fixed ROI might not capture large changes within frames.

To overcome the aforementioned problems, we propose a kernel based combined method. In the first step, instead of a dedicated ROI, feature points are applied using the intensity-based algorithm, to register successive frames. In the second step, we introduce the kernel function which penalizes the error between the coordinate of each feature point in the input frame and the sub-pixel location of the transformed correspondence, to refine image alignment. In direct contrast to the pixel selection method offered by [8], the improved performance is demonstrated on several challenging video sequences.

Manuscript received October 13, 2009.
Manuscript revised January 7, 2010.
[†]The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.
a) E-mail: miaoq07@mails.tsinghua.edu.cn
b) E-mail: wangguijin@tsinghua.edu.cn
c) E-mail: xglin@tsinghua.edu.cn
DOI: 10.1587/transinf.E93.D.1317

2. Motion Model

Suppose $\mathbf{x}$ and $\bar{\mathbf{x}}$ are corresponding points in an image pair, we have the following equation:

$$\bar{\mathbf{x}} \approx W(\mathbf{x}; h), \tag{1}$$

where $W(\mathbf{x}; h)$ is the mapping function which transforms one image to the other and $h = (h_1, \ldots h_n)^{\mathrm{T}}$ is the motion parameter. Any mapping function, such as translation, rotation or affine transform, is feasible as the motion model. In this paper, we present the planar projective transform as the motion model to handle image deformations; this is shown as Eq. (2):

$$W(\mathbf{x}; h) = \begin{pmatrix} \frac{(1+h_1)x+h_3y+h_5}{h_7x+h_8y+1} \\ \frac{h_2x+(1+h_4)y+h_6}{h_7x+h_8y+1} \end{pmatrix}. \tag{2}$$

3. Proposed Algorithm

In the beginning, our method sets the first image $I_1(\mathbf{x})$ of the sequence to be the reference image $I_{ref}(\mathbf{x})$. When the new image enters the system, we perform feature extraction and then execute inter-frame registration in the first step, as shown in Fig. 1. In the second step, we estimate the motion parameter between each input image and the reference image to generate the registered images (see Fig. 2). We propose to use feature points instead of a specific region because if the region covers a smooth area or experiences illumination change, noise or partial occlusion, incorrect registration is possible. In contrast, feature points are inherently

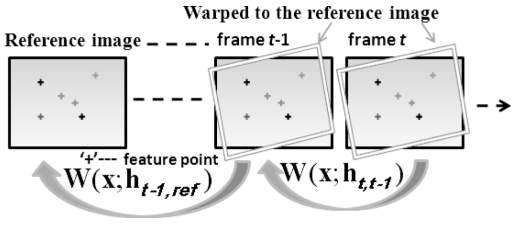

Fig. 1 The first step.

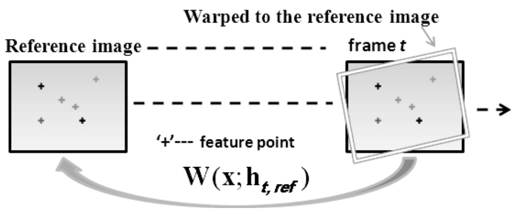

Fig. 2 The second step.

distinctive and more robust to those aforementioned negative effects. In this paper, we opted to use the Harris corner, because it is fast while not sacrificing the performance.

3.1 First Step

In this step, feature points are applied in intensity-based algorithm to register successive frames, thereby minimizing the following objective function:

$$\sum_{\mathbf{x}\in feature} \left| I_t(W(\mathbf{x}; h_{t,t-1})) - I_{t-1}(\mathbf{x}) \right|^2, \tag{3}$$

where $I_t(\mathbf{x})$ and $I_{t-1}(\mathbf{x})$ respectively denote the image in frame t and $t-1$; $\mathbf{x} = [x, y]^{\mathrm{T}}$ describes the pixel coordinate of each detected feature point in frame $t-1$. $h_{t,t-1}$ is the motion parameter used in $W(\mathbf{x}; h_{t,t-1})$ which maps the pixel $\mathbf{x}$ in frame $t-1$ to the sub-pixel location in frame t. We use bilinear interpolation to compute image brightness at subpixel locations.

To solve Eq. (3), the gradient descent method is applied, which consists of iteratively updating $h_{t,t-1}$. For further details on the method, see [9]. The identity warp $W(\mathbf{x}; 0)$ can be used as a very good initialization of $W(\mathbf{x}; h_{t,t-1})$, because the difference between two consecutive images is generally small.

At the end of this step, we solve $W(\mathbf{x}; h_{t,ref}) = W(W(\mathbf{x}; h_{t-1,ref}); h_{t,t-1})$ to recursively estimate $h_{t,ref}$, which is considered to be the initial motion parameter for input frame t in the second step.

3.2 Second Step

The relationship between frame t and the reference image is defined as:

$$I_{ref}(\mathbf{x}) \approx I_t(W(\mathbf{x}; h_{t,ref} + \Delta h_{t,ref})). \tag{4}$$

The expression in (4) is linearized by a first order Taylor expansion on $I_t(W(x; h_{t,ref} + \Delta h_{t,ref}))$:

$$I_{ref}(\mathbf{x}) \approx I_t(W(\mathbf{x}; h_{t,ref})) + \nabla I_t \frac{\partial W}{\partial h_{t,ref}} \Delta h_{t,ref}. \tag{5}$$

In this expression, $\nabla I_t = (\frac{\partial I_t}{\partial x}, \frac{\partial I_t}{\partial y})$ is the gradient of the input frame evaluated at $W(\mathbf{x}; h_{t,ref})$; i.e. ∇I_t is computed in frame t and then warped back onto the reference image using current $W(\mathbf{x}; h_{t,ref})$. The term $\frac{\partial W}{\partial h_{t,ref}}$ is the Jacobian of the warp.

As already stated, image sequence registration requires adaptive techniques for adjusting to changes in the target object over time. We use feature points with intensity-based algorithm, to attain a robust and accurate registration. Unlike the instance in the first step, the two frames $I_{ref}(\mathbf{x})$ and $I_t(\mathbf{x})$ are not successive and large displacements may appear. Some feature points between these two frames might have different illumination conditions or might not be in a planar object. Thus we employ the weighted least squares (WLS) [10] to handle this kind of issue in which the points have differing degrees of variability over the combinations of the predictor values. The weight is obtained by applying a kernel function on the distances of feature points across frames, to indicate how adequate each feature point is for aligning the images. Capturing this idea, we focus on solving the following optimization problem:

$$\min_{\Delta h_{t,ref}} \sum_{\mathbf{x}\in feature} \Bigg[I_{ref}(\mathbf{x}) - I_t(W(\mathbf{x}; h_{t,ref})) - \nabla I_t \frac{\partial W}{\partial h_{t,ref}} \Delta h_{t,ref} \Bigg]^2 K(\hat{\mathbf{x}} - W(\mathbf{x}; h_{t,ref})), \tag{6}$$

where $\hat{\mathbf{x}}$ is defined in frame t, denoting the corresponding point of $\mathbf{x}$. This paper uses the pyramidal Lucas-Kanade feature tracker [11] to find the location of $\hat{\mathbf{x}}$ in frame t. $K()$ is the 2-D realization of the kernel function, which is symmetric and attains its maximum at zero. There are numerous common kernel functions to choose from. In this paper we choose the Gaussian kernel to emphasize feature points that are well matched in luminance and location.

The motion parameter and the kernel weights are updated in successive steps iteratively. First the motion parameter is solved using a particular kernel weight; then the kernel weight is recalculated using the motion parameter in the prior step. To estimate $\Delta h_{t,ref}$ in Eq. (6), the gradient-based method is used. The motion parameter and the weight are estimated as follows:

(1) Set the iteration number $u = 0$. Establish the correspondence between feature points detected in frame t and the reference frame. Calculate the weight $K^{(0)}(\hat{\mathbf{x}} - W(\mathbf{x}; h_{t,ref}))$ using $h_{t,ref}$ estimated in the first step.

(2) Approximate $\Delta h^{(u)}{}_{t,ref}$ by solving Eq. (6).

(3) Update the motion parameter $h_{t,ref} \leftarrow h_{t,ref} + \Delta h^{(u)}{}_{t,ref}$.

(4) Compute $K^{(\mathrm{u}+1)}(\hat{\mathbf{x}} - W(x; h_{t,ref}))$ using current $h_{t,ref}$.

(5) Terminate if $\left\| \Delta h^{(u)}{}_{t,ref} \right\|_2 \le \tau$, otherwise set $u + 1 \to u$ and look back to step (2).

In addition, we also need to take into consideration that parts of the frame t might be new to the system. After $h_{t,ref}$ is obtained, the frame t is transformed to overlay over the reference image. When the overlap drops below a certain percentage (we use 70% in this paper), we update $I_t(\mathbf{x})$ as $I_{ref}(\mathbf{x})$ and record $h_{ref,1}$. When the subsequent frame t' arrives, we employ the proposed method to first compute $h_{t',\ ref}$ and then obtain the final motion parameter $h_{t',1}$ by solving $W(\mathbf{x}; h_{t',1}) = W(W(\mathbf{x}; h_{ref,1}); h_{t',\ ref})$.

3.3 Comparison with the Pixel Selection Method

The pixel selection method concentrates on picking up "good" pixels from selected ROI to align the images (i.e. it provides a hard threshold on how well each pixel in the ROI is matched). The registration accuracy can be increased using the selected pixels. However, the selection of ROI is not a trivial task; the ROI may not suffice for the large changes in the object. Meanwhile, setting the threshold is a problem. Large threshold may exclude too many pixels and the image intensity information left is inadequate for alignment while small threshold will include some "bad" pixels.

In contrast, the proposed method uses feature points to first attain an automatic registration. Also, most of the detected feature points are useful because they are distinctive and can withstand changes in the object. At the same time, kernel function is used to provide a soft threshold, making the method flexible and more robust. As a whole, the introduction of a kernel incrementally updates the contribution of each feature point and increases registration accuracy.

4. Experimental Results

We now present the experimental results of applying our algorithm on both real and synthetic image sequences. To better show the importance of the kernel, we construct a variation of our two-step approach without using kernel; we call it the combined method without kernel. We compare the pixel selection method [8], the combined method without kernel as well as the proposed method. Here we present two representative experiments to clearly show how the input frames are transformed to be superimposed over the first frame.

4.1 Moving Object with Occlusion

As a first attempt, we apply our method on a real sequence to demonstrate its robustness for illumination change and occlusion. The camera motion is non-linear while the object is stationary. A hand passes over the cards on the desk periodically. The hand shadow causes an illumination change in the image and partially occludes the card. Figure 3 (a) shows four example frames of the sequence. An occlusion occurs in the 19th frame; an illumination change exists between the 28th frame and the first frame. Figure 3 (b) shows the transformed version using the pixel selection method. The registration performance declines in the 19th frame: pixels

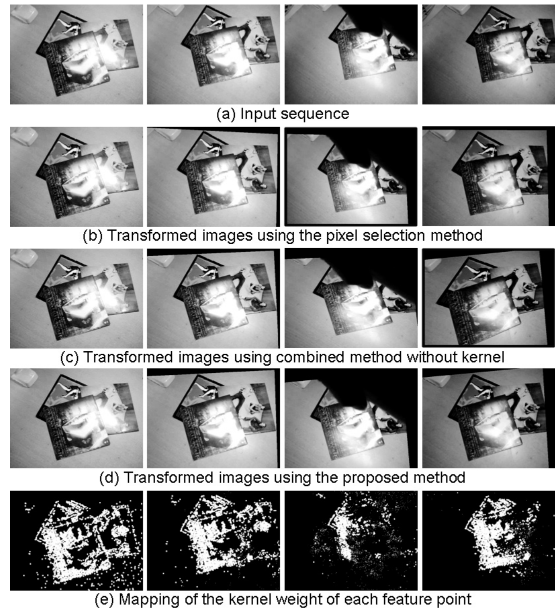

Fig. 3 Registration experimental results. From left to right column, the 1st, 8th, 19th, and 28th frame. Note the misalignments that occur in frame 19 of (b) due to occlusion and in frame 28 of (c) due to illumination change.

no longer accord well with each other in illumination and location. Figure 3 (c) shows the transformed version using the combined method without kernel. This method performs well up to the 8th frame, but the registered images exhibit a gross misalignment afterwards. The mis-registered part is dramatically improved by using our kernel-based approach, as seen in Fig. 3 (d). Figure 3 (e) shows the mapping of the kernel weight of each feature point used in the proposed approach. Some feature points are given less weights due to the shadow of the hand and the occlusion. Once the illumination condition returns to the initial state, the weights are increased again.

Given the final estimate of the motion parameters, we compute the destination of the four canonical points and compare them with the correct locations. Thus the RMS [9] error over the four points of the distance between their current and correct locations is computed, as shown in Fig. 4. The RMS error is low for all methods for the first few frames. However, error immediately increases in the pixel selection method due to occlusion (dotted line). As for the combined method without kernel (dashed line), the concatenation of the subsequent pairwise registration accumulates large error over time. In contrast, the accuracy of our approach (solid line) attains good registration at subpixel level; the negative effects of occlusion and illumination change are attenuated by the kernel.

4.2 Geometric Transformation with Image Noise

In this experiment, each frame is displaced from the previous by a projective transform model. In addition, we add to each frame some i.i.d. zero-mean Gaussian noise (with standard deviation $\sigma = 25$), to demonstrate the robustness

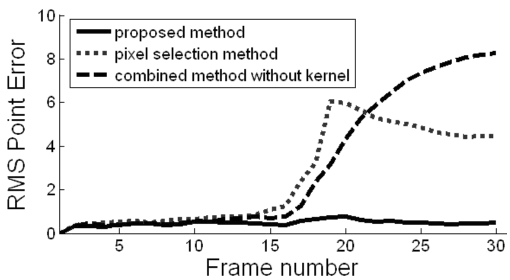

Fig. 4 RMS of the proposed algorithm versus the pixel selection algorithm and the combined method without kernel.

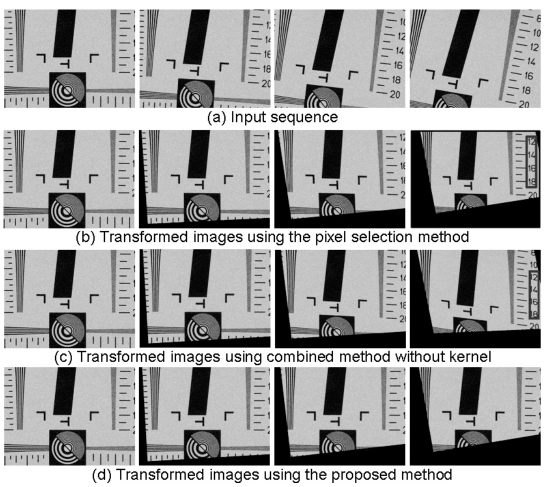

Fig. 5 Registration experimental results. From left to right column, the 1st, 14th, 27th, and 40th frame.

of our algorithm against noise.

The sequence consists of 40 frames with resolution of 640×480 pixels; four examples of them are shown in Fig. 5. Each row in Fig. 5 corresponds to that in Fig. 3. Regarding the subjective visual quality, both the pixel selection method and the combined method without kernel fail around the 27th and 40th frame. Note in frame 40 the appearance of details that do not exist in the first frame (show with blue rectangle). In contrast, our method does not exhibit this error.

5. Conclusion

This paper presented a new kernel-based image-registration algorithm. The proposed approach demonstrates that combining feature-based and intensity-based registration methods can efficiently combat against illumination change, occlusion and noise.

The proposed algorithm is composed of two steps: the first step estimates a motion parameter in the successive frames; the second step refines the motion parameter between the input and the reference frame. Feature correspondence, rather than region correspondence, is utilized. Moreover, we adapt and expand a kernel function to weigh the corresponding features in registered image pairs; features that are believed to be inadequate for image alignment are given lower weights. Experimental results verify that our approach outperforms the pixel selection method and the combined method without a kernel to achieve a subpixel accuracy level.

References

[1] B. Zitova and J. Flusser, "Image registration methods: A survey," Image Vis. Comput., vol.21, pp.977–1000, 2003.
[2] M. Brown and D. Lowe, "Recognising panoramas," ICCV, pp.1218–1225, 2003.
[3] A. Cervantes and E.Y. Kang, "Progressive multi-image registration based on feature tracking," ICIP, pp.633–639, 2006.
[4] B. Zitova, J. Flusser, and F. Sroubek, "Application of image processing for the conservation of the medieval mosaic," ICIP, pp.993–996, 2002.
[5] Y. Jianchao, "Image registration based on both feature and intensity matching," ICASSP, pp.1693–1696, 2001.
[6] M. Shimizu and M. Okutomi, "Multi-parameter simultaneous estimation on area-based matching," IJCV, vol.67, no.3, pp.327–342, 2006.
[7] D. Shen, "Image registration by local histogram matching," Pattern Recognit., vol.40, pp.1161–1172, 2007.
[8] M. Shimizu, S. Chang, and M. Okutomi, "Robust and accurate image registration with pixel selection," IEEE International Symposium on Signal Processing and Information Technology, pp.851–856, 2006.
[9] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," IJCV, vol.56, no.3, pp.221–255, 2004.
[10] S.J. Sheather, A modern approach for regression with R, Springer, New York, USA, 2009.
[11] J.Y. Bouguet, Pyramidal Implementation of the Lucas Kanade Feature Tracker, Intel Corporation, Microprocessor Reasearch Labs, 2000.