A fast image registration approach based on SIFT key-points applied to super-resolution

M Amintoosi*[1,2], M Fathy[1] and N Mozayani[1]
[1]Department of Computer Engineering, Iran University of Science and Technology, Narmak, Tehran 16846-13114, Iran

[2]Faculty of Mathematics and Computer Science, Sabzevar Tarbiat Moallem University, Sabzevar, 397, Iran

Abstract: An accurate image registration is a fundamental stage in many image processing problems. In this paper, a new and fast registration approach based on scale invariant feature transform (SIFT) key-points, under Euclidean transformation model, is proposed. The core idea of the proposed method is estimation of rotation angle and vertical and horizontal shifts using averaging of differences of SIFT key-point pairs locations. The method is simple but requires some tuning modules for accurate estimation. Orientation modification and compensation and shift compensation are some of the proposed modules. The proposed method is fast, about ive times faster than RANSAC method for model parameters estimation. The accuracy of the proposed method is compared with some popular registration methods. Various comparisons have been done with LIVE database images with known motion vectors. The experimental results over two real video sequences show the high performance of the proposed algorithm in a super-resolution application.

Keywords: image registration, super-resolution, SIFT key-points

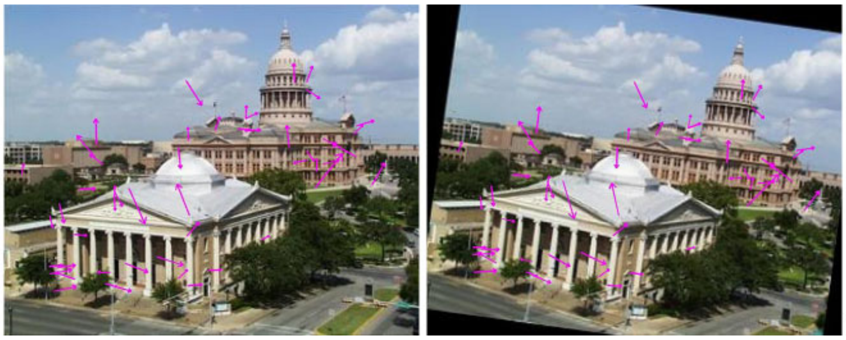
1 INTRODUCTION

One of the most critical aspects of many applications in image processing and computer vision, including super-resolution, is the image registration problem. Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns two images, the reference and sensed images.[1]

In image processing literatures, a variety of registration categories has been used. Regarding the transformation model among the images (such as translation, affine or projective), the registration method may be different. However, they can be categorised into two main approaches: area-based methods and feature-based methods. While the former uses the information from all pixels, the latter requires only a sparse set of feature correspondences to fit the motion model.[2]

The Lucas–Kanade registration algorithm,[3] is a famous area-based method, which is the basis of many other methods.[4,5] Their approach is based on a Taylor series approximation of the images. The motion parameters are the unknowns in the approximation, and they can be computed from the set of equations that can be derived from this approximation. As Taylor series only give a good approximation for small offsets, these registration methods are generally applied iteratively using a Gaussian pyramid. Vandewalle *et al.*[6] used a frequency-based registration method, where at first, the rotation parameters are estimated from a radial projection of the absolute values of the Fourier transform image. A simple one-dimensional correlation can be performed to compute the rotation angle from the projections for two images. Then, shifts are estimated from the linear phase difference between the rotation corrected images.

The MS was accepted for publication on 4 April 2011.
* *Corresponding author: Mahmood Amintoosi, Faculty of Mathematics and Computer Science, Sabzevar Tarbiat Moallem University, Sabzevar, 397, Iran; email: mamintoosi@iust.ac.ir*



1 Some of SIFT key-points of an image and its transformed version (of size 252×316): (a) 50 selected key-points from total 894 key-points; (b) 50 selected key-points from total 921 key-points. Key-points are displayed as vectors indicating scale, orientation and location. The signal-to-noise ratio (SNR)=70 dB and distance ratio=0.1 in the Lowe's algorithm. The transformation parameters of the right-side image are: $t_x=9$, $t_y=1$ and $\phi=8^\circ$

This method performs well if the images have some directionality.[6] Another fast image registration which is used in image stabilisation context is Gray-coded bit plane matching (GC-BPM).[7] This method is very computationally efficient since it uses binary boolean operations, but its performance is lower than popular methods such as Keren *et al.*.[4]

In many image processing applications, such as some remote sensing and super-resolution problems,[8] a global translational motion model is assumed, in which the low resolution input images have small rotation and translation differences with respect to each other. In this paper, a registration method based on the mentioned motion model, using scale invariant feature transform (SIFT) key-points, is proposed.

Among the various features used in feature-based image registration methods, SIFT key-points of Lowe[9] has gained a great attention in recent years. SIFT key-points are identified as the local maxima or minima of the difference-of-Gaussian filters across scales. To determine a key-point's orientation, a gradient orientation histogram is computed in the neighbourhood of the key-point. Peaks in the histogram correspond to the dominant orientations. Each key-point is denoted by a 128-element vector, named SIFT key-point descriptor. The location of each key-point in the image is specified by four floating point numbers giving subpixel row and column location, scale and orientation.

Following introducing SIFT by Lowe,[9,10] various applications of it, including matching and registration, were reported by some researchers. Mikolajczyk and Schmid[11] compared the performance of some descriptors computed for local interest regions and their results showed that the SIFT-based descriptors have the highest performance. Yi *et al.*[12] used SIFT key-points for multi-spectral remote images. They proposed a matching method and called it SR-SIFT algorithm (SIFT matching with scale restriction) to reduce the incorrect matches. The famous RANSAC (RANdom SAmple Consensus) algorithm[13] has been used many times for removing outliers (incorrect matches) from SIFT key-points pairs and estimating a homography matrix between two images.[14–17]

The original matching method proposed and implemented by Lowe[10] (the implementation is available online at: http://www.cs.ubc.ca/_lowe/keypoints/) consists of a nearest neighbour search and a heuristic criteria suggested by him, which is the ratio of closest to the second closest neighbour (named 'distance ratio'). The method is very powerful in finding correct matches among putative key-points. Figure 1 shows two versions of an instance image from LIVE dataset (as shown in Fig. 2) and some of their SIFT key-points matches with the Low's program. The rotation angle of the second image with respect to the first one was 8°. The key-points' orientations differ from ϕ, the image rotation angle with respect to the reference image.

Our goal is to estimate registration parameters (t_x,t_y,ϕ) between two images, directly from SIFT key-points' locations (x,y,θ). The angle resulting from difference of corresponding key-points' orientations is considered as an estimation of ϕ. The proper estimation of the rotation angle with this method has some limitations which is discussed in the next section. After rotation estimation, the key-points' rows and columns of the second image are rotated in a proper manner. Computing the registration parameters has been carried out by averaging after an outlier reduction stage.

The experimental results showed that the proposed method is about five times faster than RANSAC for

2 Some of LIVE database images[18] which are used in the experiments of this paper

obtaining registration parameters, while its estimation accuracy is competitive with RANSAC method. Also its precision is compared with some famous registration method like Keren *et al.*'s method.[4] The performance of the proposed method is compared with some other methods in a super-resolution application, while a high-resolution (HR) image is achieved from the motion-compensated low-resolution (LR) images with a super-resolution reconstruction method. Our implementation results show the high performance of the proposed method on read data as well as on synthesised data.

The rest of the paper organised as follows. Section 2 explains the proposed method. Section 3 provides the experimental results and Section 4 is dedicated to the concluding remarks.

2 THE PROPOSED METHOD

Assume we have a continuous two-dimensional reference signal $f_0(X)$ and its shifted and rotated version $f_1(X)=f_0[R(X+\Delta X)]$, when

$$X=\begin{pmatrix} x \\ y \end{pmatrix},\ \Delta X=\begin{pmatrix} t_x \\ t_y \end{pmatrix},\ R=\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \sin\phi \end{pmatrix} \quad (1)$$

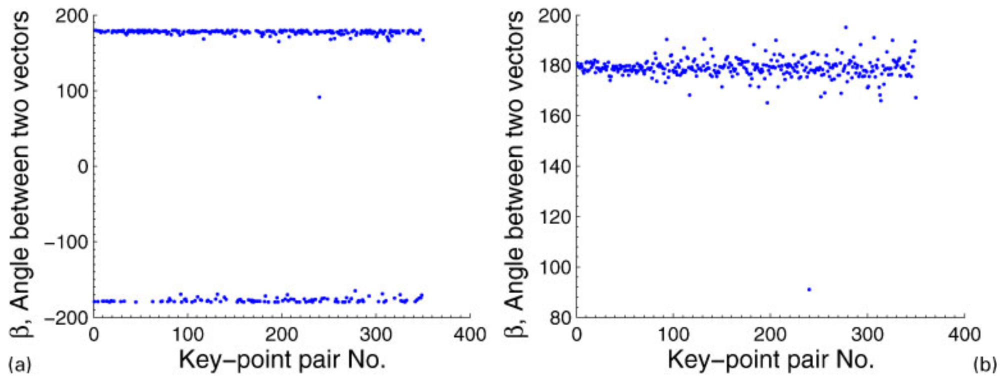
As mentioned in the previous section, each SIFT key point has a descriptor and a location vector. The descriptor is a vector of 128 values and location vector has four values for the key-point locations (row, column, scale and orientation). Key-points from a new image can be matched to those from previous images by simply looking for the descriptor vector with the closest Euclidean distance among all vectors from previous images (Section 7.1 in Ref. 10). Applying this method to two images such as those shown in Fig. 1, yields many correct matches between two images and possibly a few mismatches. In the following sections, how to deal with these mismatches is explained.

If (x_i,y_i), σ_i and θ_i denote location, scale and orientation for *i*th key-point, respectively, the goal here is to estimate (t_x,t_y,ϕ) between a pair of LR images, directly from these SIFT key-points' locations.

Suppose that $\left(x_i^1,y_i^1,\theta_i^1\right)$ and $\left(x_i^2,y_i^2,\theta_i^2\right)$ are the locations and orientations of *i*th key-point in image 1 (as the reference frame) and another image 2 (from total N matches found by Lowe's matching method). Let $\Delta x_i=x_i^1-x_i^2$, $\Delta y_i=y_i^1-y_i^2$ and $\Delta\theta_i=\theta_i^1-\theta_i^2$; the following notations for Δx, Δy and $\Delta\theta$ is used in the following subsections for explaining the proposed method

$$\begin{aligned} \Delta x &= \{\Delta x_1,\ \ldots,\ \Delta x_N\} \\ \Delta y &= \{\Delta y_1,\ \ldots,\ \Delta y_N\} \\ \Delta\theta &= \{\Delta\theta_1,\ \ldots,\ \Delta\theta_N\} \end{aligned} \quad (2)$$

Our approach for estimating t_x, t_y and ϕ from Δx, Δy and $\Delta\theta$ is simple. It is based on two assumptions: the normality of displacements and the separability of the shift and rotation estimations. At first, based on the normality of displacements, the incorrect matches, which may degrade the accuracy of estimation, are reduced by removing those data points which are far from the mean. Then based on the second assumption [the separability of the shift and rotation estimations was shown by Vandewalle[19] (Chapter 3)], ϕ, Δx and Δy are estimated. The rotation angle ϕ is approximated by averaging of $\Delta\theta$, in which some modifications are needed for accurate estimation. Then the key-points' locations of the second image are compensated according to the estimated rotation angle ϕ. Finally, t_x and t_y are approximated by averaging Δx and Δy. A normality test has been done for checking that the averaging and mismatch reduction method is meaningful. The details are discussed in the following sections.



3 Scatter plots for β in equation (3) over $\Delta\theta$ between two instance images with rotation angle equal 179°: (a) β values demonstrate two clusters, near to −180° and 180°; (b) resolving the problem by the proposed method in Section 2.1.3

2.1 Rotation estimation

The rotation angle ϕ may be approximated by averaging $\Delta\theta$, i.e. $\phi=\overline{\Delta\theta}$. Because this method is based on θ_s, the rotation angles of SIFT key-points, in the following, we name this rotation angle estimation method as T-SIFT. However, as we will see later, this method does not operate well for negative rotation angles. There are some questions, related to averaging $\Delta\theta$:

- Whether each delta follows a known distribution, like Normal distribution?
- If yes is the average of samples, has meaningful difference from the true value of parameters?

In the following, at first, the normality of the distribution of $\Delta\theta$ will be discussed. As we will see, $\Delta\theta$, based on T-SIFT, does not follow the normal distribution, when the rotation angle is negative, unless some modifications are applied.

2.1.1 *On the normality of $\Delta\theta$*

Suppose that we have M images that should be aligned with respect to a reference image in which their rotation angle are equal ϕ and their horizontal and vertical shifts are unknown. Let ϕ_m be the estimated rotation angle of mth image with respect to the reference image; $\Phi=\{\phi_1, \ldots, \phi_M\}$ which is considered as a random variable. Here we discuss whether we can consider the average value of Φ, $\overline{\Phi}$, as an estimation of rotation angle between two images (ϕ) or not.

A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. According to the central limit theorem, $\overline{\Phi}$ is approximately normally distributed for large samples, so the next step is to determine whether the expected value ϕ under the null hypothesis, has meaningful difference from its true value or not.

The experimental results over LIVE dataset,[18] verified that the estimated ϕ has not meaningful difference from true ϕ with the T-SIFT method when $0\leqslant\phi<180$, with 95% confidence; but it does not hold when $\phi<0$. Table 4 shows the results of Z-test for some methods which will be explained in more details later. The reason for these disappointing result is calculating of dot product for computing angle in Lowe's matching procedure, which is used in T-SIFT.

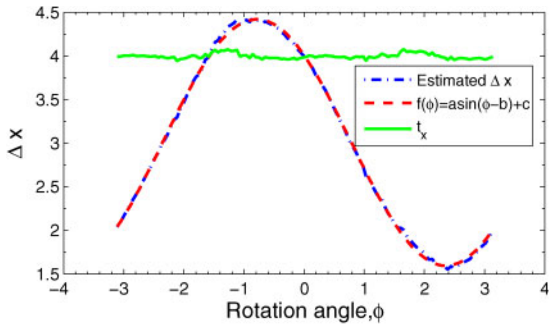
In the following, we first describe our solution for dealing with this problem and then explain our method for removing outliers, which leads to a better estimation of registration parameters.

2.1.2 *The orientation modification*

We used the implementation of SIFT key-points extraction and matching by Lowe[10] which computes dot products between unit vectors ($v_1 = \sin\theta_i^1, \cos\theta_i^1$)

Table 1 An example demonstrating a problem encountered in large angles

Key-point pair	θ_i^1	θ_i^2	$\alpha=\Delta\theta_i$	$\beta(\alpha)$
1	179	2	177	177
2	178	−3	181	−179



4 The fitting model for $\overline{\Delta x}$ over rotation angle ϕ $(-\pi \leqslant \phi \leqslant \pi)$

and $(v_2 = \sin\theta_i^2, \cos\theta_i^2)$ rather than Euclidean distances. It is computationally efficient, but would not specify whether v_1 is ahead or behind v_2. In most math libraries, a cos(.) will usually return a value between 0° and 180°. We used the following method for indicating whether v_1 is ahead or behind v_2.

The orientation of each key-point is in the range of [−180°, 180°]; hence, the difference between angles of two vectors will be in the range of [−360°, 360°]. Suppose that $\alpha = \Delta\theta_i$ is the angle between ith key-points pair, $\beta(\alpha)$, the modified orientation of α, is defined as follows

$$\beta(\alpha) = \begin{cases} \alpha & \alpha \in [-180^o, 180^o] \\ \alpha - 360 & \alpha \in (180^o, 360^o] \\ \alpha + 360 & \alpha \in [-360^o, -180^o) \end{cases} \tag{3}$$

Note that the modification is applied on the differences of SIFT key-points' orientations and the resulting angle may be positive or negative. The investigation of equation (3) is left to the reader. The advantage of this modification is that the difference angle between two key-points, indicates the direction of rotation angle.

2.1.3 *The problem of rotation angles near 180°*

It should be mentioned that with the above orientation modification, we have some problems for estimating of rotation angles, when it is close to 180° or −180°. Two key-point pairs shown in Table 1 are considered.

For both of these key-point pairs, the rotation angle (α) is close to 180°, but β is 177 and −179°. In this situation, some parts of the correspondences show positive angle (close to 180°) and some of them show negative angle (near −180°), which is not a suitable case for our algorithm that is based on averaging. Figure 3a shows this case for two images with the rotation angle of 179°. As can be seen, many of the angles of the corresponding key-points' pairs have been clustered in two groups: one close to 180° and the other close to −180°. For solving this problem, it is sufficient to replace each $\beta(\alpha)$ with its corresponding α for angles close to 180° or −180°, if the maximum of PDF of β is close enough to 180° or −180°. In this paper, 20° has been chosen as a closing threshold. The resulting scatter plot after this process has been shown in Fig. 3b.

Even after the above procedure, some outliers may exist. So those points which are far from the mean value more than $2.5\sigma_\phi$ are removed as outliers, where $\sigma_\phi = \text{std}(\Delta\theta)$. The new mean value of $\Delta\theta$ is our estimation of ϕ. The experimental results verified that the estimated rotation angle ϕ has not meaningful difference with true ϕ with 95% confidence (see Section 3, Table 4); hence, the estimated rotation angle is reliable.

2.2 Shift estimation

The overall method for estimating t_x and t_y is based on averaging of Δx and Δy, but there is some notes which is discussed in the following section.

2.2.1 *Orientation compensation*

It is obvious that using the average of Δx and Δy, the horizontal and vertical shifts have not been approximated correctly, unless the second image rotated based on ϕ and then t_x and t_y are estimated. If the image rotated by ϕ degree, we have to re-find the SIFT key-points and rerun the matching procedure. Instead of this time-consuming method, we just rotate the locations of key-points, which were found formerly, based on the estimated ϕ. The new Δx and Δy based on the rotated key-points of the second

Table 2 Estimated parameters of fitted function of $f(\phi)=a\sin(\phi-b)+c$ for horizontal shift estimation

t_x	a	b	c
2	−1.42	0.78	1
4	−1.42	0.78	3
6	−1.42	0.78	5

image are computed. The t_x and t_y are approximated by averaging of Δx and Δy after an outlier removal based on σ_x and σ_y from Δx, Δy.

2.2.2 *Shift compensation*

Our experimental results with fixed known t_x and variable t_y over various values of ϕ ($-2\pi<\phi\leqslant 2\pi$), show that the estimated t_x has a sinus shape function. It is the reason why the estimated value is not accurate as enough. The following sinusoidal function is fitted on the estimated Δx over ϕ

$$\overline{\Delta x}(\phi)\approx f(\phi)=a\sin(\phi-b)+c \tag{4}$$

The result of an experience with t_x=4 over 3480 images is shown in Fig. 4, The dashdot line in Fig. 4 is the estimated horizontal shift ($\overline{\Delta x}$). For each of 29 LIVE database images, 120 random image with t_x=4 pixel, random vertical shifts and 120 rotation angles ($-178^\circ:3:180^\circ$) were generated. Hence, each point in the Fig. 4 demonstrates the average of $\overline{\Delta x}$ over 29 images with an specified rotation angle. This process is repeated for t_x=2 and 6; the parameters of the fitted functions for t_x=2, 4 and 6 are shown in Table 2.

As can be seen in Table 2, a and b are fixed over various t_x and c=t_x-1, regardless of t_x. Substituting c=t_x-1 in equation (4) yields

$$\overline{\Delta x}(\phi)=a\sin(\phi-b)+t_x-1$$

$$\Rightarrow$$

$$t_x=\overline{\Delta x}(\phi)-a\sin(\phi-b)+1 \tag{5}$$

The dashed line in Fig. 4 is the fitting function, $f(\phi)=-1.42\sin(\phi-0.78)+3$, and the solid line is t_x, compensated based on equation (5). We performed a Z-test of the null hypothesis that the estimated t_xs are a random sample from a normal distribution with mean 4, against the alternative that the mean is not 4. The result indicated a failure to reject the null hypothesis at the 5% significance level.

Similar experiments was done for fixed t_y. Hence, the following functions are used for shifts compensation

$$t_x=\overline{\Delta x}(\phi)-a_x\sin(\phi-b_x)+1,\quad a_x=-1.42,\quad b_x=0.78 \tag{6}$$

$$t_y=\overline{\Delta y}(\phi)-a_y\sin(\phi-b_y)+1,\quad a_y=1.42,\quad b_y=-0.78 \tag{7}$$

2.3 The overall algorithm

The repetitive patterns in the images produce some mismatches in the matching stage of Lowe's algorithm. Here, those points which are far from the mean more than 2.5σ are removed as outliers. The overall framework based on the previous stages and outlier removal is shown in Algorithm 1. The experimental results showed better performance when ϕ was re-estimated after computing shift parameters. Hence, in Algorithm 1, we have a *for* loop.

Algorithm 1 Registration based on SIFT key-points' locations
Input: The pair of images: Image1 and Image2
Output: Registration parameters (t_x,t_y,ϕ).
1: Extract SIFT key-points
2: Find correspondence pairs of key-points, based on Lowe's suggestion
3: Compute the difference of key-point pairs locations: $\Delta\theta$, Δx and Δy according to equation (2)
4: Estimate rotation angle ϕ based on averaging of $\Delta\theta$ using the method described in Section 2.1
5: **for** i=1 to 2 **do**
6: Estimate horizontal and vertical shifts, t_x and t_y from Δx and Δy using the method described in Section 2.2.1
7: Remove outliers from $\Delta\theta$ and re-estimate ϕ
8: **end for**
9: Compute $\phi=\overline{\Delta\theta}$, compute t_x and t_y according to equations (6) and (7)

We named our proposed method OXYT-SIFT, where each letter is described in Table 3. Based on Table 3, other variations of the proposed method can be named easily, for example, OT-SIFT stands for the proposed method, where we have only orientation modification and rotation estimation based on $\Delta\theta$, without shift estimation.

Table 3 Describing the letters in OXYT-SIFT, the name of the proposed method

Letter	Explanation
O	Stands for orientation modification described in Section 2.1.2
X	Stands for shift estimation along X axis (Section 2.2)
Y	Stands for shift estimation along Y axis (Section 2.2)
T	Stands for estimation of rotation angles based on θ, without any modifications

Table 4 Z-test over some rotation angles (ϕ)

Image	$\phi=-178°$		$\phi=-127°$		$\phi=-76°$		$\phi=-25°$		$\phi=26°$		$\phi=77°$		$\phi=128°$		$\phi=179°$	
	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT
Bikes	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Building2	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Buildings	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Caps	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Carnivaldolls	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Cemetry	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Churchandcapitol	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Coinsinfountain	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Dancers	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Flowersonih35	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
House	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Lighthouse	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Lighthouse2	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Manfishing	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Monarch	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Ocean	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Paintedhouse	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Parrots	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Plane	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Rapids	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Sailing1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Sailing2	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Sailing3	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Sailing4	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Statue	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Stream	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0
Studentsculpture	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0
Woman	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
Womanhat	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0

In the next section, we will see the experimental results of the proposed method in image registration and an application to super-resolution.

3 EXPERIMENTAL RESULTS

Our experiments has been done over LIVE dataset images[18] in which some of them were shown in Fig. 2.

We mentioned in Section 2.1 that the estimated ϕ has not meaningful difference from true ϕ with the proposed method in Section 2.1.1 with 95% confidence. For each image of LIVE dataset, 120 image were generated with random t_x,t_y ($\in[-10, 10]$) and over various ϕ [$-180°$, $180°$]. Table 4 shows the result of Matlab Z-test function for $\Delta\theta$ with the mentioned T-SIFT method and its modified version based on orientation modification, named as OT-SIFT described in Section 2.1. As can be seen, the null hypothesis about normal distribution of $\Delta\theta$ can not rejected with T-SIFT method only for $\phi>0$, but the null hypothesis cannot rejected for all values of rotation angle by OT-SIFT method.

Also the effect of noise for normality of $\Delta\theta$ has been tested here. Table 5 shows the result of Z-test over various SNRs with fixed value of $\phi=-45°$. The SNR of 100 dB means without noise. These experiments ensures that the estimated rotation angle with the proposed method does not have significance difference from its true value.

3.1 Run times

Table 6 shows the average run times of different methods for estimation of registration parameters over 3480 images. The proposed OXYT-SIFT

Table 5 Z-test over various SNRs for $\phi=-45°$

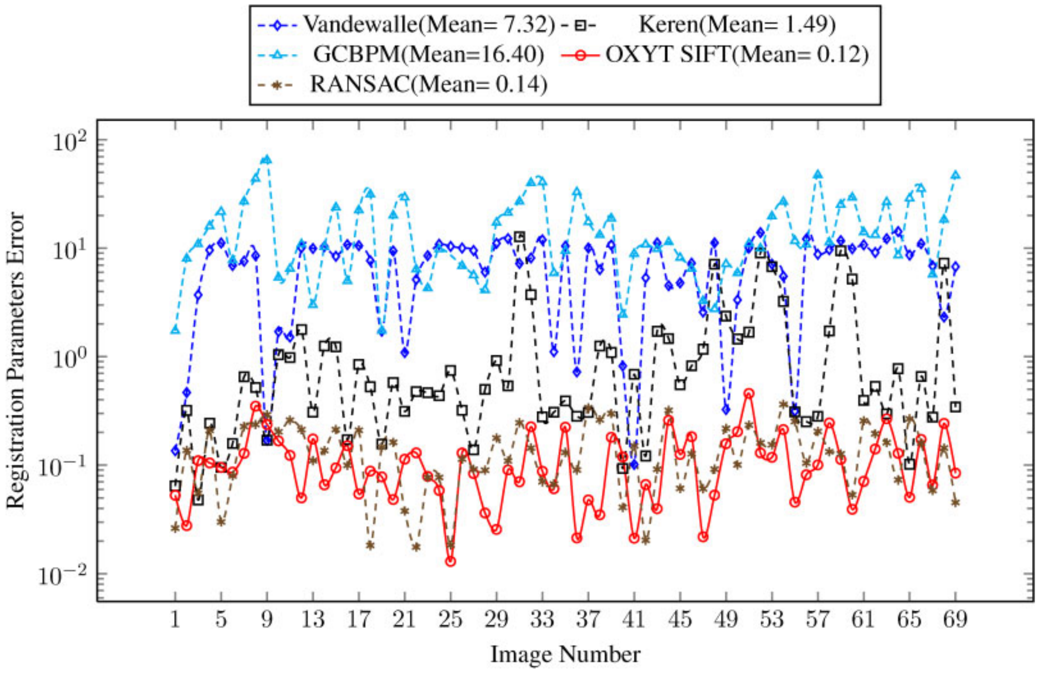
	SNR=20 dB		SNR=45 dB		SNR=70 dB		SNR=100 dB	
Image	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT	T-SIFT	OT-SIFT
Bikes	1	0	1	0	1	0	1	0
Building2	1	0	1	0	1	0	1	0
Buildings	1	0	1	0	1	0	1	0
Caps	1	0	1	0	1	0	1	0
Carnivaldolls	1	0	1	0	1	0	1	0
Cemetry	1	0	1	0	1	0	1	0
Churchandcapitol	1	0	1	0	1	0	1	0
Coinsinfountain	1	0	1	0	1	0	1	0
Dancers	1	0	1	0	1	0	1	0
Flowersonih35	1	0	1	0	1	0	1	0
House	1	0	1	0	1	0	1	0
Lighthouse	1	0	1	0	1	0	1	0
Lighthouse2	1	0	1	0	1	0	1	0
Manfishing	1	0	1	0	1	0	1	0
Monarch	1	0	1	0	1	0	1	0
Ocean	1	0	1	0	1	0	1	0
Paintedhouse	1	0	1	0	1	0	1	0
Parrots	1	0	1	0	1	0	1	0
Plane	1	0	1	0	1	0	1	0
Rapids	1	0	1	0	1	0	1	0
Sailing1	1	0	1	0	1	0	1	0
Sailing2	1	0	1	0	1	0	1	0
Sailing3	1	0	1	0	1	0	1	0
Sailing4	1	0	1	0	1	0	1	0
Statue	1	0	1	0	1	0	1	0
Stream	1	0	1	0	1	0	1	0
Studentsculpture	1	0	1	0	1	0	1	0
Woman	1	0	1	0	1	0	1	0
Womanhat	1	0	1	0	1	0	1	0

Table 6 Average run times over 3480 images for estimating of motion parameters

Method	T-SIFT	OXYT-SIFT	RANSAC
Time (ms)	2.81	3.50	16.78

Table 7 Estimated parameters of some images with various methods

Image	t_x					t_y					ϕ				
	Ground truth	Vandewalle	Keren	OXYT-SIFT	RANSAC	**Ground truth**	Vandewalle	Keren	OXYT-SIFT	RANSAC	**Ground truth**	Vandewalle	Keren	OXYT-SIFT	RANSAC
Bikes	**−2**	−1.95	−2.04	−2.01	−2.02	**−3**	−2.87	−3.04	−3.00	−2.99	**1**	1.00	0.97	0.95	0.99
Building2	**−7**	0.38	−7.02	−7.01	−6.87	**6**	−0.15	5.83	6.01	5.84	**−8**	−8.10	−7.82	−8.10	−8.00
Buildings	**−3**	0.72	−2.86	−2.98	−3.18	**7**	0.38	6.40	6.96	7.14	**10**	9.80	10.22	10.12	10.01
Caps	**−3**	−1.88	−3.09	−2.95	−3.09	**2**	1.09	1.02	2.00	2.18	**4**	3.10	3.66	3.84	3.96
Carnivaldolls	**−1**	0.03	−0.91	−1.15	−1.08	**−9**	0.89	−8.82	−8.97	−8.93	**2**	1.70	1.77	1.91	2.03
Cemetry	**0**	−0.01	−0.09	0.00	0.07	**10**	−0.79	9.93	10.03	9.93	**−4**	−4.00	−3.87	−4.15	−4.01
Churchandcapitol	**−3**	−2.30	−2.86	−3.08	−3.15	**−3**	−1.65	−2.97	−2.99	−3.01	**1**	0.20	1.07	0.98	0.99
Coinsinfountain	**5**	1.02	4.93	4.98	4.93	**4**	0.74	4.16	4.01	4.03	**3**	2.90	2.56	2.87	3.02
Dancers	**−5**	−0.13	−5.20	−5.01	−5.01	**9**	−0.18	8.40	9.01	9.02	**0**	−0.10	0.39	−0.01	−0.01
Flowersonih35	**−7**	−1.53	−6.79	−6.99	−7.08	**2**	−0.47	1.57	2.00	2.04	**2**	1.40	1.86	2.04	1.99
House	**−2**	0.69	3.53	−2.00	−2.20	**7**	0.30	−2.63	6.99	7.14	**10**	9.60	3.77	9.93	10.00
Lighthouse2	**−1**	−0.69	−1.12	−1.05	−1.02	**−4**	−3.01	−4.22	−4.03	−3.94	**3**	2.60	2.82	3.01	2.99
Manfishing	**−5**	0.54	−4.94	−5.04	−5.23	**−8**	0.39	−8.27	−8.01	−7.76	**10**	10.00	9.88	10.03	9.97
Monarch	**−2**	−1.54	−1.98	−2.00	−2.08	**−4**	−3.35	−3.97	−3.97	−3.96	**−1**	−1.20	−0.92	−0.88	−0.99
Ocean	**−10**	−0.36	−8.36	−10.02	−10.08	**−5**	0.70	−5.45	−5.03	−4.95	**6**	5.80	5.81	6.00	5.98
Paintedhouse	**−1**	0.44	−1.69	−1.03	−1.07	**−7**	0.14	−7.37	−7.01	−6.89	**5**	4.90	4.75	4.82	5.01
Parrots	**−1**	−0.99	−1.58	−1.03	−1.17	**−1**	−0.87	−2.31	−1.04	−0.86	**7**	6.70	5.12	6.85	7.03
Plane	**−10**	−0.17	−7.39	−10.09	−10.11	**10**	0.13	1.35	10.01	9.92	**0**	−0.10	0.52	0.09	−0.08
Rapids	**−1**	−0.97	−1.15	−1.03	−1.19	**−3**	−2.91	−3.13	−2.99	−2.84	**8**	7.70	7.76	8.03	8.00
Sailing1	**−4**	1.00	−3.99	−3.89	−4.04	**−8**	0.30	−9.57	−7.79	−7.88	**7**	6.70	6.28	7.07	7.03
Sailing4	**−4**	0.62	−4.02	−3.98	−3.85	**−9**	0.70	−8.80	−9.03	−9.21	**−9**	−8.70	−8.66	−9.06	−9.03
Stream	**−10**	−0.51	−9.66	−10.00	−10.05	**10**	−0.19	9.37	10.00	10.06	**3**	0.00	2.70	3.13	2.99
Studentsculpture	**−5**	0.15	−4.75	−5.01	−5.04	**5**	0.42	4.89	5.00	5.06	**2**	1.90	1.98	1.94	2.01



5 Values of mean square error between estimated registration parameters and their ground truth values, over 69 distorted images

method is about five times faster than RANSAC in average. For each of 29 LIVE images, 120 random images with known motion vectors was created. Their SIFT key-points were extracted and saved before computing running times.

The initial SIFT matching time which is common in the methods is discarded. The mean run time for RANSAC was 16.78 ms, against to 3.5 ms for OXYT-SIFT, i.e. the OXYT-SIFT is faster than RANSAC about five times.

Distance ratio for matching SIFT key-points was set to 0.8, and distance threshold for deciding outliers in RANSAC homography was set to 0.01.

3.2 Registration comparison

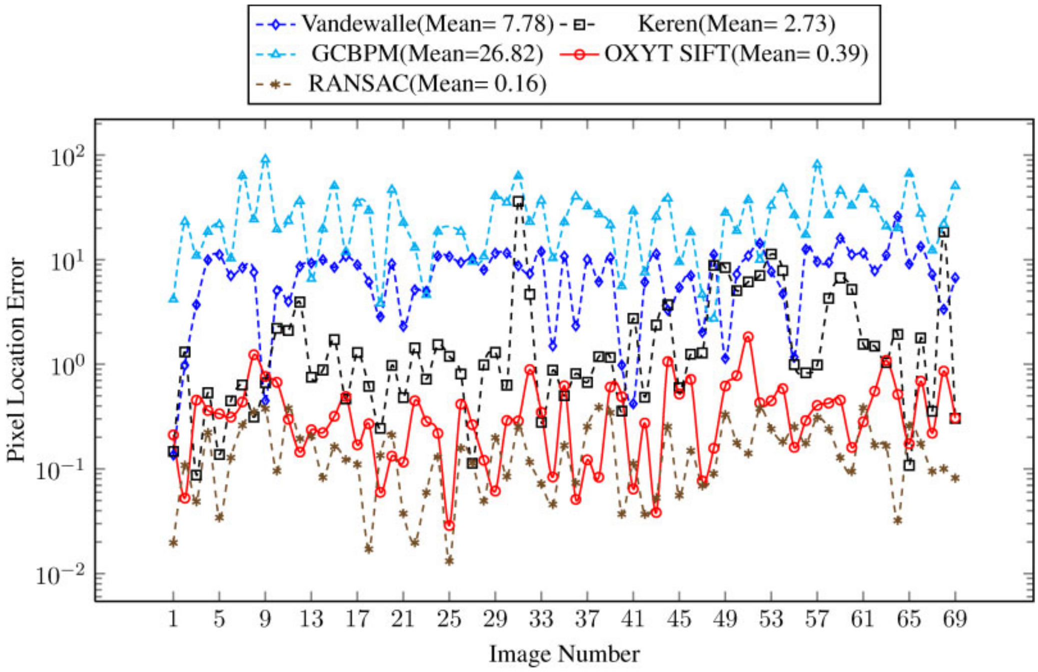
In addition to RANSAC, we applied three registration methods: frequency method,[6] GC-BPM[7] and Keren *et al.*,[4] for comparing the proposed registration approaches.

For every image in LIVE dataset, four distorted image was generated by resizing each image by a factor of 0.5, then the image was shifted by known values among the X and/or Y axis (in the range of $[-10, 10]$ pixels), and rotated by a specified angle (in the range of $[-10^\circ, 10^\circ]$). Noise is added to the image, so that the SNR of the produced image was 70 dB and JPEG compression were the last steps. It should be mentioned that the motion parameters for the first distorted image were set to zero as it is a reference frame. The images were generated in a manner to be used in registration comparison and in a super-resolution application.

Because of the image size restriction in GC-BPM method, six images of dataset were dropped and the comparisons were done on remaining 23 images. Since for each reference image we had three distorted images, the total number of tested images is 69. Table 7 shows the estimated parameters (t_x,t_y,ϕ) for 23 of 69 distorted images, along with their ground truth values, with various methods (instead of GC-BPM for table size limitation).

For better comparison, in Fig. 5, the mean square error (MSE) between motion parameters estimated by all mentioned registration methods for all 69 images is demonstrated. The average value of each method is demonstrated beside its legend. As can be seen, the proposed method produced better results against the others in average.

Since three registration parameters have different units, the following error measure is a better criterion than the errors in parameters. For each estimated model, the MSE is computed over 30 random points in the image coordinate frame of the distance between their current and correct transformed locations. Figure 6 shows comparison of values of the MSE

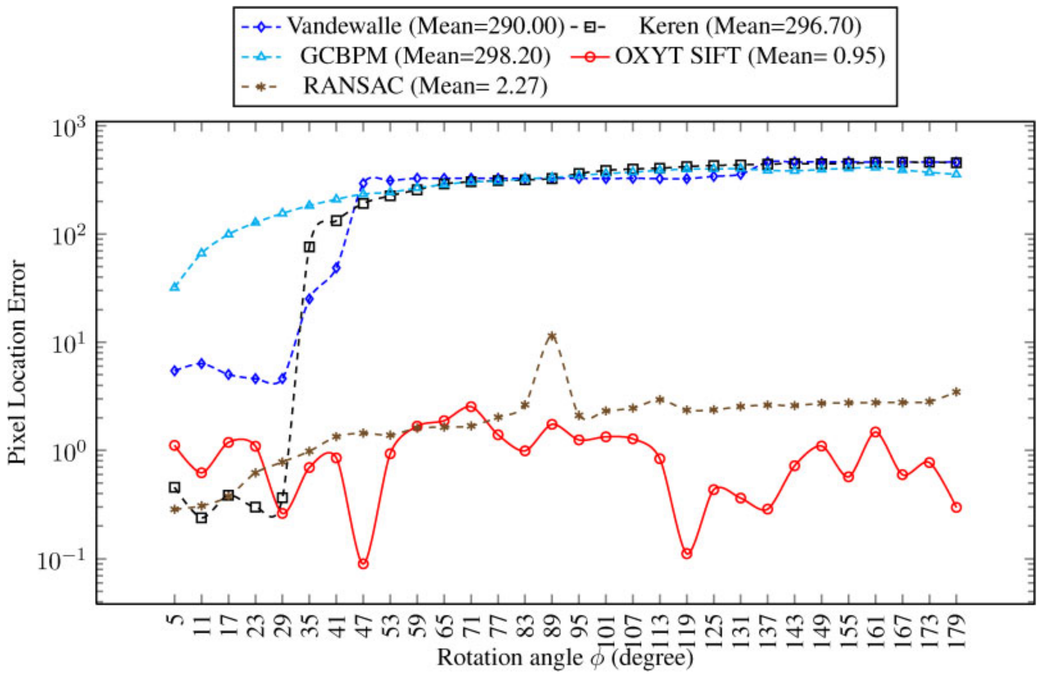


6 Values of mean square error between estimated registration pixel location error and their grand truth values, over 69 distorted images

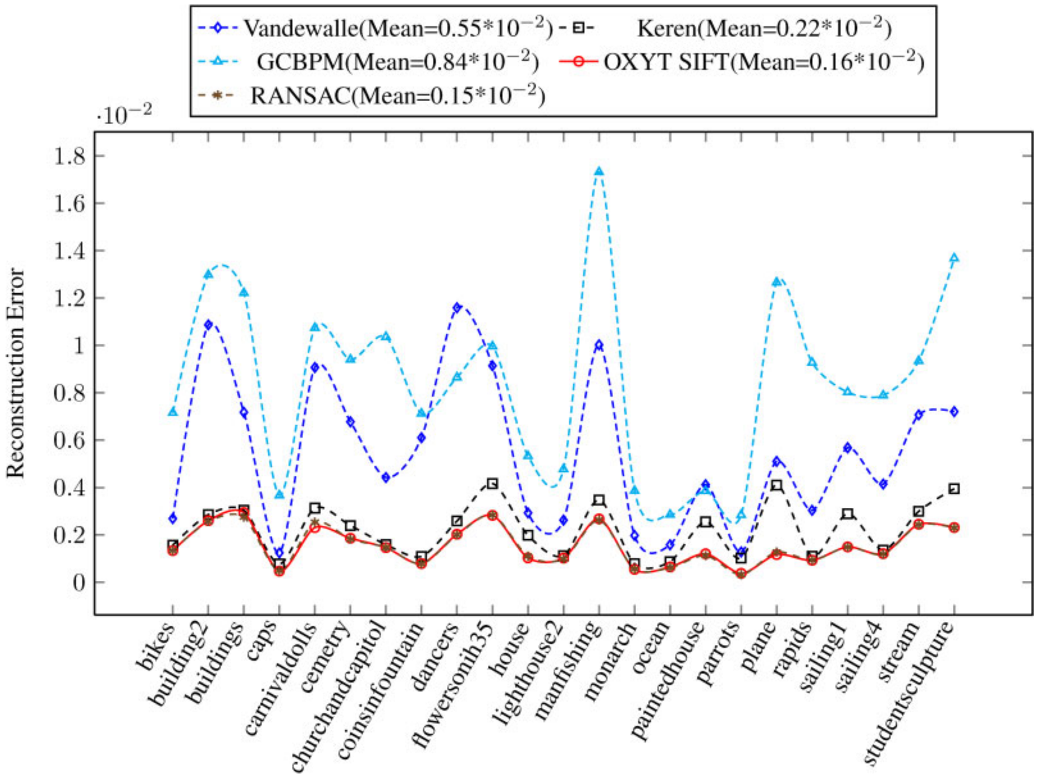
between the estimated locations and true locations by mentioned registration methods. As can be seen, the proposed method after RANSAC method produced better results against the others in average.

3.3 Dealing with large angles

In the above experiments, the range of rotation angles was [$-10°$, $10°$], but the proposed method can



7 Registration pixel location error for 'Bulidings' image of size 384×256 over various values of rotation angles



8 MSE between reconstruction HR image with the mentioned registration methods as the first stage of super-resolution, over four LR images corresponding to each LIVE image

be used to a wider range ([−180°, 180°]). Figure 7 shows the result of running the mentioned methods on an image rotated by various angles in the range of [0°, 180°]. According to the aforementioned criterion, the MSE between the estimated locations and true locations of 30 pixels is chosen for comparison.

As can be seen, the proposed method produced the best result in average. Its performance does not decrease with increasing rotation angle, in contrast to those of the other methods. The first three registration methods in Fig. 7 are not suitable for estimation of large rotation angles. The performance of RANSAC method decreases when ϕ increases.

3.4 Experimental results for super-resolution problems

The super-resolution techniques fuse a sequence of low-resolution images to produce a higher-resolution image. The LR images may be noisy, blurred and have some displacements with each other. The origin of the classic from of super-resolution comes back to the work of Tsai and Huang[8] motivated by the need to improve the resolution of images acquired by the Landsat 4 satellite. These methods utilise information from multiple observed images to achieve restoration at resolutions higher than that of the original data. The super-resolution restoration methods register the observed images to a common reference frame in order to reconstruct the HR image. Since Euclidean transformation model is a common assumption in multi-frame super-resolution literatures, here it is chosen as an application of the proposed method.

In this section, two experiments have been carried out, the first on the previous synthesised data set and the other on real data. As we will see, the proposed approach has a good performance in both cases.

3.4.1 Super-resolution experiment on synthesised data

As mentioned earlier, for each image in the dataset, four LR images were generated with random motion vectors. Hence, for each image of LIVE, as a HR image, we have four LR images, in which the first LR image is considered as the reference image. The motion parameters are estimated with various registration methods to produce a HR image with a magnifying factor of 2. Among the super-resolution reconstruction methods, the interpolation approach is used here.

Figure 8 show the MSE between the produced HR images with each registration method against the real HR image.

a b c d

9 Cemetry LR images, with the motion parameters (t_x,t_y,ϕ) of each LR image: (a) reference LR frame; (b) (0,10,−4); (c) (5,−9,8); (d) (8,−2,1)

As can be seen, the proposed method and RANSAC method produced the better results. Note that in this experiment, the rotation angles was belong to [−10°, 10°]; based on the results of the previous section, the better result of the proposed method in larger rotation angles is expected.

3.4.2 *Super-resolution experiment on real data*

We compared the proposed registration method with Keren *et al.*'s method[4]and RANSAC methods on two real videos, in which consequence frames can be transformed to each other with an Euclidean transformation. This happens for example because of hand shaking during video capturing. From each video, another LR video is created by blurring and down-sampling by a factor of 2. Table 8 shows the specifications of these two videos.

Since these are real data, the ground truth registration parameters is not available; hence, the registration comparison is not applicable. Instead, a super-resolution reconstruction method is applied for comparison proposes. A common video super-resolution method is

Table 8 Description of our test sequences. The differences between HR and LR frames are clear with zooming on the electronic version of the paper

Sequence name	Seq. 1	Seq. 2
First HR frame		
Frames	330	69
Resolution	640 × 480	720 × 576
Device:	Sony HDR-SR12E	Panasonic NV-GS75
First LR frame		
Resolution	320 × 240	360 × 288
A middle LR frame		

10 Close-up views of the original HR image, replication (nearest) resized version of the first LR image and super-resolution results, when Vandewalle, Keren, RANSAC and the proposed 'OXYT-SIFT' methods have been used as registration stage: (a) original HR frame; (b) close-up of text area of HR frame; (c) resized of the first LR frame (nearest); (d) Vandewalle; (e) Keren; (f) RANSAC; (g) OXYT-SIFT

using a 'sliding window' technique. The window is moved forward in time to produce successive super-resolved frames in the output sequence.[20] These window frames are registered related to the the first frame in the window and a reconstruction method is applied to the registered frames. Here the window size

Table 9 Values of average MSE between reconstructed HR frames of two videos over various frame intervals and the ground truth HR frames

MSE		Method		
Seq.	Frame interval	Keren *et al.*	OXY-TSIFT	RANSAC
1	2	145.40	**118.70**	118.80
	4	186.39	**144.78**	149.44
	10	371.18	**225.53**	234.63
2	2	298.00	295.74	**273.69**
	4	355.44	305.60	**296.37**
	10	532.82	**332.50**	332.55

11 Error map image between Figs. 10d–g with related HR image (Fig. 10b) (brighter pixel means higher error): (a) Vandewalle; (b) Keren; (c) RANSAC; (d) OXYT-SIFT

is set to four frames and the interpolation method is used for reconstruction stage.

The sliding window is applied to each LR video. More accurate registration leads to more similarity of the reconstructed frame to original ground truth HR frame. The experiments has been carried out over various frame intervals (FI) for each video. FI=n means frame numbers 1, $1+n$, $1+2n$, … are considered for experiment. Larger FI leads to larger displacements between frames.

Table 9 shows values of the average MSE over various frame intervals for sequences 1 and 2. Each cell in the table represent the average MSE of all frames used in the experiment. The bold letters in each row indicate the lower MSE in that row. As can be seen, the proposed OXYT-SIFT method and RANSAC method produced the good results, which is the same as our experiments on synthesised data.

3.5 Visual comparison

For visual comparison the super-resolution results, when various methods has been used as registration stage of interpolation reconstruction method, has been used. The 'cemetry' image, which contains a text area, has been chosen and four LR images were produced from it with the way explained in Section 3.2. Horizontal and vertical shifts of the three LR images with respect to the first LR image were {1,4,−1} and {−7,−5,10} pixels, respectively, and the rotation angles were {−5°,1°,2°}. The LR images are shown in Fig. 9.

For a better visual comparison, a small region, containing the text, has been enlarged and is shown in the Fig. 10.

The first LR image is shown in Fig. 10c; note that the lower text (NO THROUGHFARE PLEASE) is almost unreadable in the LR image. The result of Vandewalle method (Fig. 10d) suffers from bad registration. Although the results of Keren, RANSAC and OXYT-SIFT seems equal, but referring to Fig. 8, it indicates that values of the MSE of the two later methods are a bit smaller than that of the Keren method. This is also illustrated as absolute error map of the results and the original HR image in Fig. 11.

4 CONCLUSION

In this paper, a new registration method with the assumption of normal distribution of displacements of SIFT key-points' locations, after some modifications was proposed. The main idea is averaging of differences of key-points' orientations for rotation estimation. The key-points' rows and columns of the second image are rotated based on the estimated rotation angle, and vertical and horizontal displacements are approximated by averaging of differences of key-points' locations. Some modification and compensation have been carried out for accurate estimation of registration parameters.

In contrast to Keren *et al.*,[4] registration method, which is a repetitive method and cannot handle large rotation angles between images, the proposed approach is a one-step approach and can handle large rotation angles. In contrast to the frequency method of Vandewalle *et al.*,[6] the proposed approach does not need strong directionality in images. The

GC-BPM[7] is a fast registration algorithm, but produced poor results with respect to the proposed method. Finally, the proposed method is faster than famous RANSAC algorithm for registration parameters estimation. Moreover, the proposed method has some variations which can be used, when there are only vertical or horizontal shifts, rotation or combinations of them.

The main strength of the proposed method is in the situations where SIFT key-points of the images are known as a priority, and at the same time, the registration parameters are requested, for example, in an object recognition and tracking application based on SIFT key-points. In this case, the proposed method is faster than RANSAC about five times for parameter estimation, while its accuracy is competitive to RANSAC.

The only limitation of the proposed method is that it can be used only for Euclidean transformation model (translation + rotation), although it is a usual assumption such as many super-resolution applications. In the future work, we plan to extend the proposed method for other transformation model such as similarity model (Euclidean + isotropic scaling) with the aid of Yi *et al.*[12]

The various comparisons showed that the proposed registration method outperforms some other popular methods. The experimental results showed the high performance of the proposed method in a super-resolution problem.

In summury, the innovations of this paper are as follows:

- using SIFT key-points' orientations directly for a fast and powerful image registration approach
- using SIFT key-points with this manner for image registration in the super-resolution context
- justification that the displacements between corresponding SIFT key-points' locations, under global translational motion model, are approximately normally distributed, after some modifications, and under the null hypothesis, the estimated parameters have not meaningful difference from their true values.

ACKNOWLEDGEMENTS

The authors are indebted to anonymous referees for valuable comments. We would also like to thank Dr Patrick Vandewalle (Laboratoire de communications audiovisuelles, École Polytechnique Fédérale de Lausanne, Switzerland), for providing us with his super-resolution package, Dr Alan Brooks (Department of EECS, Northwestern University, USA, http://dailyburrito.com/m-files/) for his stabilisation Matlab codes and Professor David Lowe for his SIFT program. They also thank to Dr Peter Kovesi (http://www.csse.uwa.edu.au/_pk/research/matlabfns/) and Dr Hamid Sheikh for providing LIVE dataset[18], "Dr Gh. Mohajeri Minaei for video sequence 2" and A. Ghodsi (PhD Student in Spatial Processes, University Putra Malaysia) for his contribution.

REFERENCES

1 Zitová, B. and Flusser, J. Image registration methods: a survey. *Image Vision Comput*., 2003, **21**, 977–1000.

2 Pham, T. Q. 'Spatiotonal adaptivity in super-resolution of under-sampled image sequences', PhD thesis, Technische Universiteit Delft, Delft, The Netherlands, 2006.

3 Lucas, B. and Kanade, T. An iterative image registration technique with an application to stereo vision, Proc. 7th Int. Joint Conf. on Artificial intelligence: *IJCAI '81*, Vancouver, BC, Canada, August 1981, ACM, pp. 674–679.

4 Keren, D., Peleg, S. and Brada, R. Image sequence enhancement using sub-pixel displacement, Proc. IEEE Int. Conf. on Computer vision and pattern recognition: *CVPR '88*, Ann Arbor, Ml, USA, June1988, IEEE Computer Society, pp. 742–746.

5 Baker, S., Gross, R. and Matthews, I. Lucas–Kanade 20 years on: a unifying framework. *Int. J. Comput. Vision*, 2004, **56**, 221–255.

6 Vandewalle, P. 'Super-resolution from unregistered aliased images', PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2006.

7 Ko, S., Lee, S., Jeon, S. and Kang, E. Fast digital image stabilizer based on gray-coded bit-plane matching. *IEEE Trans. Consum. Electron*., 1999, **45**, 598–603.

8 Tsai, R. and Huang, T. In Advances in Computer Vision and Image Processing (Ed. R. Y. Tsai and T. S. Huang), 1984, Vol. 1, pp. 317–339 (JAI Press Inc., Greenwich, CT).

9 Lowe, D. G. Object recognition from local scale-invariant features, Proc. IEEE Int. Conf. on Computer Vision: *ICCV '99*, IEEE Computer Society, Vol. 2, p. 1150.

10 Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 2004, **60**, 91–110.

11 Mikolajczyk, K. and Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Patt. Anal. Mach. Intell*., 2005, **27**, 1615–1630.

12 Yi, Z., Zhiguo, C. and Yang, X. Multi-spectral remote image registration based on SIFT. *Electron. Lett*., **44**, 107–108.

13 Fischler, M. A. and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 1981, **24**, 381–395.

14 Brown, M. and Lowe, D. G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision*, 2007, **74**, 59–73.

15 Amintoosi, M., Fathy, M. and Mozayani, N. Reconstruction+synthesis: a hybrid method for multi-frame super-resolution, Proc. 5th Iranian Conf. on Machine vision and image processing: *MVIP '08*, Tabriz, Iran, November 2008, Tabriz University, pp. 179–184.

16 Yuan, Z., Yan, P. and Li, S. Super resolution based on scale invariant feature transform, Proc. Int. Conf. on Audio, language and image processing: *ICALIP 2008*, Shanghai, China, July2008, IEEE, pp. 1550–1554.

17 Amintoosi, M., Fathy, M. and Mozayani, N. Regional varying image super-resolution, Proc. IEEE Int. Joint Conf. on *Computational sciences and optimization*: *CSO 2009*, Sanya, China, April 2009, IEEE, Vol. 1, pp. 913–917.

18 Sheikh, H., Wang, Z., Cormack, L. and Bovik, A. LIVE image quality assessment database release 2 [online]. Available at: <http://live.ece.utexas.edu/research/quality>, June 1, 2011.

19 Vandewalle, P., Süsstrunk, S. and Vetterli, M. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP J. Appl. Signal Process.*, 2006, **2006**, 1–14.

20 Borman, S. 'Topics in multiframe superresolution restoration', PhD thesis, University of Notre Dame, Notre Dame, IN, USA, 2004.