

# Real-Time Image Mosaicing Using Non-rigid Registration

Rafael Henrique Castanheira de Souza, Masatoshi Okutomi, and Akihiko Torii

Tokyo Institute of Technology

rafaelh.souza@ok.ctrl.titech.ac.jp , {mxo,torii}@ctrl.titech.ac.jp

<http://www.ok.ctrl.titech.ac.jp>

**Abstract.** Mosaicing is a classical application of image registration where images from the same scene are stitched together to generate a larger seamless image. This paper presents a real-time incremental mosaicing method that generates 2D mosaics by stitching video key-frames as soon as they are detected. The contributions are three-fold: (1) we propose a “fast” key-frame selection procedure based solely on the distribution of the distance of matched feature descriptors. This procedure automatically selects key-frames that are used to expand the mosaics while achieving real-time performance; (2) we register key-frame images by using a non-rigid deformation model in order to “smoothly” stitch images when scene transformations can not be expressed by homography; (3) we add a new constraint on the non-rigid deformation model that penalizes over-deformation in order to create “visually natural” mosaics. The performance of the proposed method was validated by experiments in non-controlled conditions and by comparison with the state-of-the-art method.

**Keywords:** mosaic, non-rigid, registration, feature based, real-time.

## 1 Introduction

Mosaicing is a classical application of image registration. Typically, a set of images is stitched together to simulate a camera with a larger field of view. Real-time mosaicing can be useful for medical imaging, augmented reality, digital camera panorama generation, etc. Online registration, i.e., stitching key-frames as soon as they are detected, is necessary for real-time processing.

In this work, we propose a method of online mosaicing that can generate 2D mosaics from video inputs acquired beyond homography assumptions. Classical mosaicing methods work under the assumption that the input images are related to each other by homography (*projective transformation*). This assumption holds true when the images are acquired under some limited conditions (camera rotation around its optical center or scene lying on a planar surface). Unless these conditions are satisfied, the images can not be perfectly aligned by registration and the results may be very poor. This problem may be alleviated by the application of non-rigid registration [5].

A naive approach to online mosaicing is to register and stitch the current key-frame into the previous key-frame. The process will accumulate registration error, which will grow with each new image added to the sequence.

This paper presents a method which uses a very efficient feature based non-rigid registration model in order to align images with high precision. At the same time, the over-deformation of the mosaic is avoided during the online mosaic creation. These two objectives are achieved by formulating the registration problem by enforcing smoothness while keeping the original proportions of the captured frame. Additionally, in order to achieve real-time processing, the key-frames are efficiently extracted from the video by a procedure which uses the distance distribution of matched feature descriptors.

In Sect. 2, the related methods are presented. Section 3 presents the proposed method. Section 4 shows the result of experimental validations. Finally, Sect. 5 shows the conclusions of this research and future research subjects.

## 2 Related Work

For the reader who is not familiar with mosaicing, Szeliski [13] presents a comprehensive tutorial about a variety of methods of registration and mosaic composition.

Since mosaic is a well studied area of computer vision, there are many approaches to 2D mosaicing. These works can be grouped in 3 classes: (1) offline methods that use homography or lower degree transformations, (2) offline methods that use higher degree transformations, and (3) online methods. The group (1) includes the works, [2,9,12,7], which are based on global transformations such as homography. The group (2) includes the works [4,3], which model the deformation as quadratic functions. The group (3), which is the most related to the proposed method, includes the works of [6,10]. The work in [6] uses 3D information for registering aerial images using a non real-time algorithm. The method in [10] is online and avoids the problem of over deformation by using fixed camera movements (*e.g.* translation, forward motion).

Although most of the presented works dealing with mosaicing make use of global transformations such as homography, there are more general registration methods that use non-rigid deformation. Some of them use feature based methods, *e.g.* [5,11,14]. Feature based methods are generally more computationally efficient than area based methods [13], specially in the case of non-rigid registration. The method in [14] can register correctly pairs of images even in the presence of a large ratio of outliers in real-time. However, this method is designed for pairs of images only.

On top of the state of the art, the contributions of the proposed method are: real-time performance, use of non-rigid registration, prevention of over-deformation of the mosaic, and less restrictions on camera movement.

### 3 Proposed Method

The mosaicing procedure consists of four steps: frame selection, feature matching, registration, and mosaic displaying. The frame selection module reads the input video and selects which key-frames will be used to create the mosaic. The feature matching module matches the feature points in the newly selected frame to the features in the previously selected frame. The pairwise registration module receives the set of matched features and registers the newly selected frame into the previously selected frame. The registered frame is then sent to the mosaic creation module, where it is added to the mosaic and displayed. The procedure is repeated again, until the end of the video. The modules are explained in more details in the following sections.

#### 3.1 Frame Selection

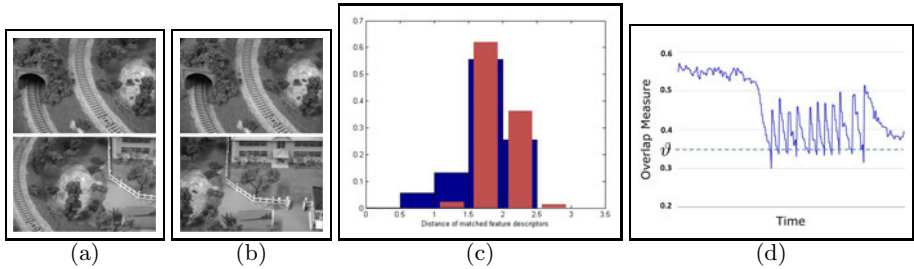
In order to create mosaics efficiently, only a small subset of the video frames must be selected. This key-frame set must be as sparse as possible, to reduce the number of registrations performed. At the same time, it must contain overlapping key-frames so that a mosaic can be composed out of them. To fulfill these requirements, it is necessary to estimate the overlap of pairs of frames. To do so, the following algorithm is proposed: (1) the features in both frames being compared are detected using SURF descriptors [1]); (2) the nearest-neighbor matching of the features is computed; (3) a histogram of the matched descriptors is computed; (4) the overlap measure (OM) is computed. The OM was defined as follows:

$$OM(H) = \sum_{j=1}^{n_{Bin}} G((j - 0.5)h_{size}, \varsigma) H_j \quad , \quad (1)$$

where  $n_{Bin}$  is the number of bins in the histogram,  $h_{size}$  is the size of each bin,  $(j - 0.5)h_{size}$  is the average range of the bin  $j$ ,  $G$  is a Gaussian weighting function with 0 mean and standard deviation  $\varsigma$ . This weighting function assigns larger weights to values near zero, and the weight decays quickly, so that the bins which probably contain correct matches receive a larger weight than the bins with wrong matches. So, using the OM, the key-frames are selected by the following algorithm: (1) the first video frame is selected and used as reference; (2) the next frame whose OM (comparing with the reference frame) is smaller than a given threshold is selected and becomes the new reference. Step (2) is repeated until the end of the video.

It was experimentally observed that the probability distribution of the descriptor distances changes according to the intersection size between the image pair. Fig. 1(a) shows two frames with a small overlap. The descriptor distance has a bell-shape like distribution (fig. 1(b)). Fig. 1(c) shows two frames with a larger overlap. The distribution becomes bimodal (fig. 1(d)). The smaller peak represents the inliers among the matched features. Fig. 1 shows the variation of

OM over time, in a video recorded by a translating camera. The value of OM decreases as the intersection becomes smaller and rises again when a new frame is selected.



**Fig. 1.** (a): pair of frames with a large overlap; (b): pair of frames with a small overlap; (c): histogram of the distance of matched descriptors; the blue bars represent pair (a) and the red bars the pair (b); (d): variation of the overlap measure over time

### 3.2 Registration

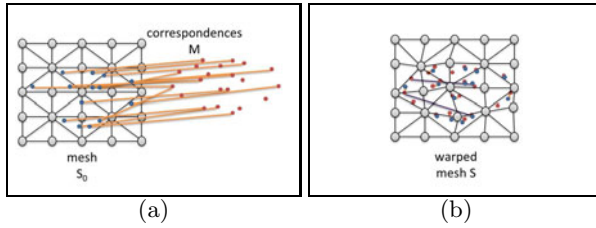
This section explains the registration model used in the proposed method. Two constraints must be met: the mosaic must be as seamless as possible and as similar as possible to the original captured frame (i. e., over-deformation must be avoided). For doing so, the proposed method applies a non-rigid deformation model that uses triangle meshes and a registration algorithm that uses feature points obtained by the frame selection procedure and pruned by RANSAC [8].

**Deformation Model for Image Registration.** A 2D mesh model is used to implement the non-rigid transformations. Each vertex (or control point)  $v_j$  is represented by its coordinates  $(x_j, y_j)$ . The entire mesh is written as  $S = (X, Y)$ , where  $X$  is a vector containing the  $x$  coordinates of the control points and  $Y$  the vector containing the  $y$  coordinates. The warp of any point  $p$  inside a mesh triangle defined by the vertices  $v_i, v_j$ , and  $v_k$  can be calculated using the barycentric coordinates of  $p$ :  $w(p, S) = \sum_{l \in \{i, j, k\}} B(p, v_l) [x_l, y_l]^T$ , where  $B(p, v_l)$  is the barycentric coordinate of  $p$  in relation to  $v_l \in \{v_i, v_j, v_k\}$  (computed in relation to the identity mesh  $S_0$ ). Fig. 2 illustrates the basic principle of this kind of transformation.

**Problem Formulation.** The initial model of pairwise non-rigid registration was drawn from Zhu *et al.*'s work [14], which was based on Pilet *et al.*'s work [11]. It is summarized by the equation below:

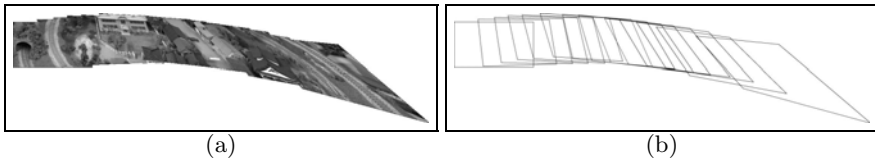
$$E(S) = E_C(S) + \lambda E_{S_m}(S) \quad , \quad (2)$$

where  $E_C$  is the correspondence energy function and  $E_{S_m}$  is the smoothness energy. The constant  $\lambda$  balances the compromise between precision and mesh smoothness. The registration is solved by finding the mesh  $S$  which minimizes



**Fig. 2.** Deformation using a mesh model: (a) shows the identity mesh, (b) shows the mesh  $S$  warped to reduce the projection error of the matched features

$E(S)$ . The correspondence energy is proportional to the projection error of warped features, while the smoothness energy measures the discontinuities on  $S$ ; this energy is important to remove outlier feature matchings. The initial formulation described by (2) is suitable for pairwise image registration, however. The registration of sequences of images poses some additional problems. If only pairwise registration is used to align a sequence of images, over-deformation due to error accumulation may occur (fig. 3).



**Fig. 3.** Error accumulation using homography: (a) rendered mosaic , (b) projected frame borders. The last frame is the most deformed.

To avoid error accumulation, a modified version of the previous energy function is presented. The new term,  $E_{\text{Ref}}(S - S_{\text{Ref}})$  is named *reference mesh energy*. The mesh  $S_{\text{Ref}}$  represents a model of how the mesh  $S$  should look like without over-deformation. Alternatively, it is how the user of the mosaic system would expect the image (warped by  $S$ ) to look like. The constant  $\mu$  regulates the reference mesh energy weight. The new formulation is presented below:

$$E'(S) = E_C(S) + \lambda E_{\text{Sm}}(S) + \mu E_{\text{Ref}}(S - S_{\text{Ref}}) \quad (3)$$

**Correspondence Energy.** The correspondence energy  $E_C(S)$  is a function of the projection error of the matched features. The matched feature set is represented by  $M$ . The matched feature pair  $c \in M$  is composed of two features  $(c_0, c_1)$ , where  $c_0$  is a feature found in the target image and  $c_1$  is its paired feature found in the image being warped. The warp function  $w(c_1, S)$ . The function  $v$  is the same robust estimator used by Zhu *et al.*[14]. It is defined below:

$$E_C(S) = \sum_{c \in M} v(c_0 - w(c_1, S), \sigma); v(\delta, \sigma) = \begin{cases} \frac{\|\delta\|^2}{\sigma^n} & \text{if } \|\delta\| \leq \sigma \\ \sigma^{2-n} & \text{otherwise} \end{cases} \quad (4)$$

The function  $v$  has two parameters: the projection error  $\delta$  and the radius of tolerance  $\sigma$ . The matches whose projection error are greater than the radius of tolerance are considered outliers and penalized. The radius of tolerance  $\sigma$  dictates which matched feature pairs will be considered outliers, conferring robustness to the registration procedure.

**Smoothness Energy.** The correspondence energy, if used alone, is sensitive to outliers among the matched features. A smoothness constraint is added to the model in order to avoid this problem. The proposed method uses the same smoothness constraint found in Zhu *et al.*[14] and Pilet *et al.*[11]. This energy is the sum of the approximate second derivative of the mesh  $S$ . Let  $E$  be the set of all collinear control points in  $S$  that define two adjacent edges. The smoothness energy is defined below:

$$E_{\text{Sm}}(S) = \sum_{i,j,k \in E} (-x_i + 2x_j - x_k)^2 + (-y_i + 2y_j - y_k)^2 = X^T K X + Y^T K Y, \quad (5)$$

where  $K = K'^T K'$ , and  $K'$  is a matrix containing one row per triplet in  $E$  and one column per mesh vertex. The row corresponding to the triplet  $(i, j, k)$  has all of its values zero except by values in columns  $i, j$ , and  $k$ , that have values  $-1, 2$ , and  $-1$ , respectively [11].

**Reference Mesh Energy.** The registration using the energy function in (2) is only suited for pairwise registration, because registration error may accumulate, as shown in fig. 3. The role of the reference mesh energy is to alleviate this problem. This energy is proportional to the  $L_2$  distance between the mesh  $S$  and the reference mesh  $S_{\text{Ref}}$ . The former is the registration solution and the latter is an approximation of how  $S$  should be if it has no over-deformation. The criteria selected to generate  $S_{\text{Ref}}$  was to make  $S_{\text{Ref}}$  look similar to the original captured image.  $S_{\text{Ref}}$  is defined as the similarity transformation (i.e., rotation, translation and scaling) that minimizes the correspondence energy. This mesh can be computed efficiently by reducing the projection error using the similarity transformations combined with RANSAC. The reference mesh energy is defined below:

$$E_{\text{Ref}}(S - S_{\text{Ref}}) = \frac{1}{2} \|S - S_{\text{Ref}}\|^2. \quad (6)$$

During the optimization process, the reference mesh energy is stronger in the regions of the mesh  $S$  where there are no features. While the region with features is deformed to minimize the projection error, the region without features is deformed by similarity transformations. These local differences in the deformation are not possible for global registration models.

**Optimization Routine.** As pointed in [14], the projection error  $\delta$  can be written as a linear system. Given that:  $c_0 = (c_{0x}, c_{0y})$ ,  $c_1 = (c_{1x}, c_{1y})$ :

$$\|\delta\|^2 = (c_{0x} - t^T x)^2 + (c_{0y} - t^T y)^2, \quad (7)$$

where  $x$  and  $y$  are the coordinates of the mesh and  $t_{c_1} \in R^N$  is a vector ( $N$  is the number of control points) representing the barycentric coordinates of the feature point  $c_1$ , which is inside the triangle defined by  $v_i, v_j, v_k \in S_0$ , (calculated in the identity mesh). The vector  $t_{c_1}$  has all its values 0, except in the coordinates  $i, j$ , and  $k$ , where the barycentric coordinates of  $c_1$  in relation to  $v_i, v_j$ , and  $v_k$  are set, respectively. Using (5) and (6), the energy  $E'(S)$  in (3) can be rewritten as:

$$E'(S) = \frac{1}{\sigma^n} \sum_{c \in M_{Inl}} \left( c_{0x}^2 + c_{0y}^2 - 2 \begin{bmatrix} c_{0x}t \\ c_{0y}t \end{bmatrix}^T S + S^T \begin{bmatrix} tt^T & 0 \\ 0 & tt^T \end{bmatrix} S \right) + |M_{Out}| \sigma^{2-n} + \lambda (X^T K X + Y^T K Y) + \frac{\mu}{2} \|S - S_{Ref}\|^2, \quad (8)$$

where  $M_{Inl}$  is the set of inlier matches,  $M_{Out}$  is the set of outlier matches. The following definitions are done for simplification:  $A = \frac{1}{\sigma^n} \sum_{c \in M_{Inl}} tt^T$ , and  $b = \begin{bmatrix} b_x \\ b_y \end{bmatrix} = \frac{1}{\sigma^n} \sum_{c \in M_{Inl}} \begin{bmatrix} c_{0x}t \\ c_{0y}t \end{bmatrix}$ . Computing the gradient of  $E'$  and setting it to zero, the mesh  $S$  can be found by solving a linear system:

$$S = \begin{bmatrix} \lambda K + A + \mu I & 0 \\ 0 & \lambda K + A + \mu I \end{bmatrix}^{-1} (b + \mu S_{Ref}) . \quad (9)$$

The optimization is repeated varying the value of  $\sigma$ , which decreases during the optimization procedure. In the beginning,  $\sigma$  is large, allowing many possible outliers to influence the result of the optimization process. However, since the module of the derivative of the  $E_C$  is small when  $\sigma$  is large,  $E_{Sm}$  and  $E_{Ref}$  have a larger weight and they initially guide the optimization. As the value of  $\sigma$  decreases, the weight of  $E_C$  increases, guiding the optimization to minimize the projection error of the remaining inliers. In this way, this registration method is robust to outliers. The process stops when  $\sigma$  is smaller than a given threshold.

In order to display the results, the mosaics are created by warping the registered frames one over the other. In order to avoid using regions without features, that may have large registration errors, only the convex hull of the correctly aligned feature points is warped.

## 4 Experiments

The objective of the experiments is to demonstrate four points: the proposed method has a smaller projection error comparing to the classical approaches, the mosaics created by the proposed method have less over-deformation, the proposed method can run in real-time, and that the results obtained by the proposed method are more robust than the results obtained by classical approaches in the kind of video considered.

### 4.1 Experimental Setup

The project was run in a computer with Intel(R) Core(TM) i7 CPU (2.93 GHz) and 4GB of RAM. The proposed method was implemented using the OpenCV library. The parameter setting is presented in Table 1.

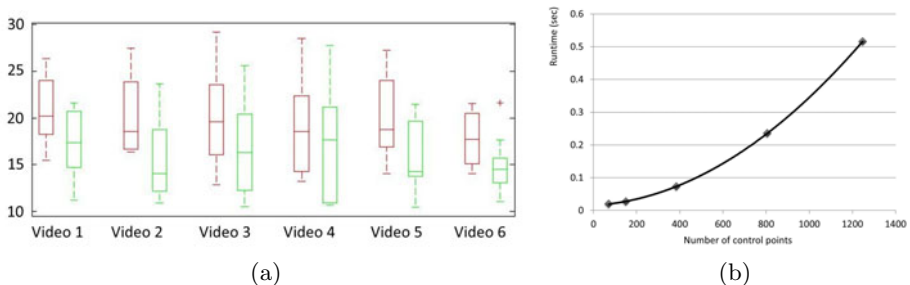
**Table 1.** Parameter settings for the proposed method

Param.	Value	Description
$\vartheta$	0.4	Frame selection threshold.
$\varsigma$	1.0	Frame selection weight function std. deviation.
$\lambda$	$10^{-6}$	Smoothness energy parameter.
$\mu$	$10^{-7}$	Reference mesh energy parameter.
$n$	4	Correspondence energy parameter.
$\sigma_0$	32	Registration parameter; initial radius of tolerance.
$\sigma_{min}$	3	Minimum radius of tolerance; i.e., projection error.
$\eta$	0.5	Radius of tolerance decay rate.

For the reference mesh computation, the precision of RANSAC is set to 99% in the presence of 70% of outliers. The size of the mesh was  $19 \times 28$  control points. The videos used on the experiments had a resolution of  $720 \times 480$ .

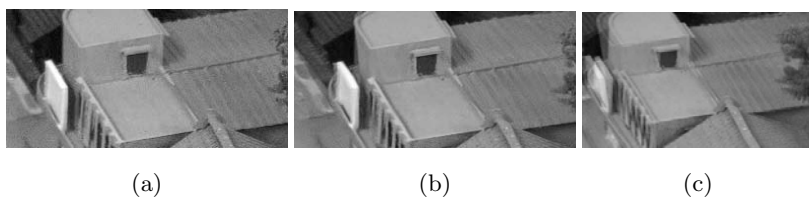
## 4.2 Registration Precision

This experiment presents the comparison between homography and non-rigid transformations concerning precision by means of mean appearance error, defined as the mean absolute difference between between all aligned pixels. The experiments were conducted by registering of pairs of images. Fig. 4(a) shows the results of the average error of pair-wise registration over different video sequences. Fig. 5 shows a detail of a pair of registered frames (the averaged image). As can be seen, the results achieved by the registration method used by the proposed method are always more precise than the results using homography. This happens because the deformation field between the pairs of images can not be precisely described by a global transformation like projection, since the displacement field depends on the geometry of the scene.



**Fig. 4.** (a): appearance error with homography and non-rigid transformations. The error is measured as the mean absolute difference between pixel gray-scale values of aligned pixels, in a set of videos. The red boxes show the results obtained by homography, and the green boxes represent the results of the proposed method; (b): execution time (seconds) in relation to number of control points.

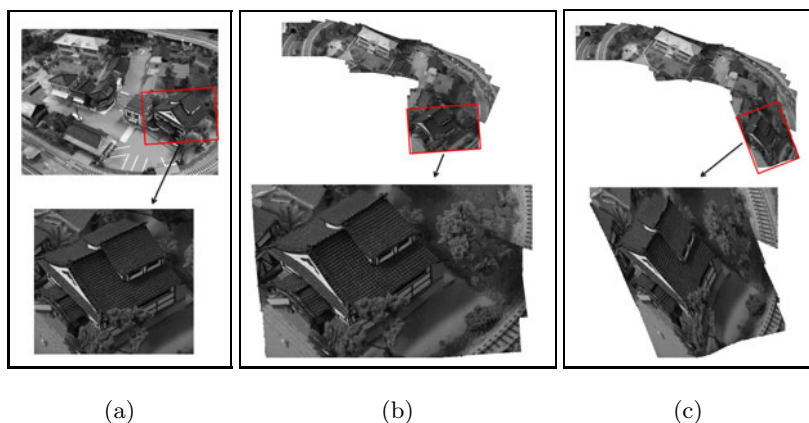




**Fig. 5.** Detail of a pair of registered frames, showing the average of the superposition of the frames, aligned by: (a) shows the original frame from the video, (b) proposed method and (c) homography

### 4.3 Over-Deformation Avoidance

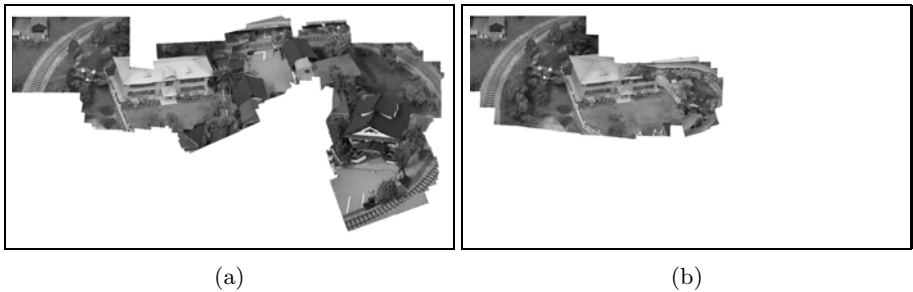
This set of experiments compares mosaics done by the proposed method and non-rigid registration as described by [14]. The comparisons are done regarding over-deformation. Figure 6 shows the results. Both methods use the same set of frames. As previously showed in fig. 3, using homography, the registration error tends to build up and cause the frames to over-deform. When using only non-rigid registration, without the reference mesh energy, error accumulation also happens, even though the alignment error is small when compared to homography. The proposed method, using the reference mesh energy, minimizes these amount of over-deformation. This result may be achieved by related methods using bundle adjustment, but the proposed method achieves the same by only doing pair-wise registration.



**Fig. 6.** Mosaicing results, regarding overdeformation; (a) detail of the city model used in the experiments, showing an expected undeformed frame; (b) shows the result obtained by the proposed method; and (c) shows the results obtained using only non-rigid registration without the reference mesh energy

#### 4.4 Comparison with a Standard Method

In this set of experiments, the proposed method was compared to a standard method, implemented by Microsoft Image Composite Editor (ICE), version 1.3.5. Using ICE, the user can choose different camera movements. The one which yielded the best result was selected. The proposed method used the parameters described in section 4.1. ICE and the proposed method used the same set of key-frames. Fig. 7 shows the mosaic created from a video taken by a camera moving over a city model. The results can be seen below.



**Fig. 7.** Comparison between the proposed method and a standard method; (a) shows the result of the proposed method; (b) shows the result of the standard method. The proposed method created a more complete mosaic since it can handle more complex camera movements.

Fig. 7(a) shows the results obtained by the proposed method. Fig. 7(b) shows the results obtained by ICE. As can be seen, the results obtained by the proposed method are more complete than the results given by ICE. This happens because of the complex camera movement and the non-planar surface, which violate the projection constraints used by ICE.

#### 4.5 Computational Complexity

The current implementation of the proposed method runs in about 32 frames per second with a tax of of 2 frames selected per second, what is reasonable for videos where the camera movement is not excessively fast.

Each iteration of frame selection takes approximately 0.031 seconds, so the frame rate is about 32 frames per second, enough for most videos. Fig. 4(b) shows runtime regarding only the registration procedure. It was executed 10 times for each quantity of control points (the computation of the reference mesh is included). As can be seen from the experiments, registration runtime grows slowly. This happens because the implementation that uses sparse matrices to represent the registration model. The runtime of the frame selection and mosaic creation procedure were also computed. Using approximately 1000 triangles, the registration can be done in about 3 frames per second. Regarding the mosaic

creation, each frame takes on average 0.4 seconds to be added into the mosaic, a tax of nearly 2 frames per second.

The conclusion is that the proposed method can run in real time, given the conditions above. Further optimization on the method may be performed in the future.

## 5 Conclusions and Future Work

This paper presented a new mosaicing technique based on feature based non-rigid registration. The proposed method can be used to create mosaics of non-planar surfaces in real-time. This model deals with the problem of over-deformation using only pairwise registration, and creates mosaics with smaller alignment error when comparing with standard approaches. For this purpose, the reference mesh energy was presented. An efficient method of key-frame selection, created to achieve real-time performance, was also presented. The proposed method has some restrictions. First, since there is no bundle adjustment, the generated mosaic is prone to error if a region of the scene is recorded twice (loop). This will require efficient loop closing method able to run in real-time. The proposed method also fails when sharp discontinuities in the optical flow are present, due to the smoothness constraint. These limitations will be tackled in our future research.

## References

1. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
2. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* 74, 59–73 (2007)
3. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: application to mosaicing the curved human retina. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 585–591 (2000)
4. Chaiyasarn, K., Kim, T.-K., Viola, F., Cipolla, R., Soga, K.: Image mosaicing via quadric surface estimation with priors for tunnel inspection. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 537–540 (2009)
5. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 89(2-3), 114–141 (2003)
6. Crispell, D., Mundy, J., Taubin, G.: Parallax-free registration of aerial video. In: *Proc. British Machine Vision Conf.* (2008)
7. Deng, Y., Zhang, T.: Generating panorama photos. In: *Proc. of SPIE Internet Multimedia Management Systems IV* (2003)
8. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 381–395 (1981)
9. Hsu, S., Sawhney, H.S., Kumar, R.: Automated mosaics via topology inference. *IEEE Computer Graphics and Applications* 22(2), 44–54 (2002)

10. Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A.: Mosaicing on adaptive manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(10), 1144–1154 (2000)
11. Pilet, J., Lepetit, V., Fua, P.: Real-time non-rigid surface detection. In: *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 822–828 (2005)
12. Sawhney, H.S., Hsu, S., Kumar, R.: Robust Video Mosaicing Through Topology Inference and Local to Global Alignment. In: Burkhardt, H., Neumann, B. (eds.) *ECCV 1998. LNCS*, vol. 1407, pp. 103–119. Springer, Heidelberg (1998)
13. Szeliski, R.: Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.* 2, 1–104 (2006)
14. Zhu, J., Lyu, M.R., Huang, T.S.: A fast 2d shape recovery approach by fusing features and appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1210–1224 (2009)