Progressive Multi-image Registration based on Feature Tracking

Albert F. Cervantes*, Eun-Young Kang
albert.cervantes@gmail.com*, eykang@calstatela.edu
California State University, Los Angeles
5151 State University Drive
Los Angeles, CA 90032 USA

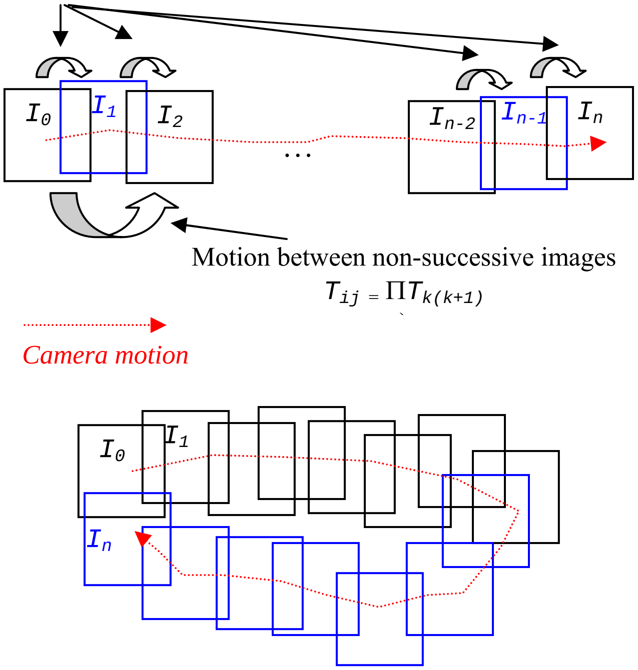
Abstract

Many previous multi-image registration methods perform their global registration process when all images of the input sequence are available. However, these methods are not suitable for real-time applications that aim to progressively register the current incoming image with respect to the previous images. In this paper, we present an innovative progressive multi-image registration method that uses an efficient feature-based pairwise local registration and feature tracking. First, transformations for temporally successive frames are estimated and concatenated to relate non-successive images. This pairwise registration recovers 2D transformation parameters using feature correspondences and an iterative least squares method. Then, overlapping images are detected among the previous and current image by mapping all images to the reference image coordinate space. When multiple overlaps exist between the current image and previous images, multi-image registration is performed. To do so, features in the overlapping regions are tracked, i.e. same features are identified from non-successive frames within the overlaps and linked together. These are the same features used earlier for pairwise local registration. Tracked features contribute to compute geometric errors caused by the concatenation of pairwise transformations and are used to correct mis-alignment of the current image. Our global registration method extends methods used for local registration (feature matching and parameter estimation) to feature tracking and parameter modification for the global registration in a unified way. The accuracy of our method is demonstrated by creating mosaic images from sets of images.

Keywords: Local and Global Image Registration, Feature Matching and Tracking, Parameter Estimation

1. Introduction

Many video processing applications such as video surveillance and video compression deal with video sequences captured by moving cameras and perform motion estimation in order to stabilize video frames, remove redundancies within a video sequence, and detect and track same objects appearing in different video frames. Among many motion estimation methods, the pairwise method, using a 2D parametric model, has been widely used for its simplicity. This approach estimates 6-affine motion parameters (or 8-homography parameters) for only temporally successive frames and recovers motion parameters for non-successive frames by concatenating pairwise motions. This approach works well for video sequences that exhibit a linear camera motion such as a camera panning in one direction as shown in Figure 1 (top). However, when a video sequence exhibits non-linear camera motions (Figure 1 bottom), the obtained result from pairwise motion estimation is insufficient to express accurate motions between frames that are overlapping spatially but not temporally. It is because the concatenation of these pairwise motion parameters produces a gross error accumulation. The process to find an accurate motion between non-successive frames while reducing the accumulated error is called “global registration” and is essential for many video applications.



Motion between non-successive images
$T_{ij} = \Pi T_{k(k+1)}$

Camera motion

Observe accumulated error between I_0 and I_n using $T_{0n} = \Pi T_{k(k+1)}$

Figure 1 Pairwise registration and accumulated error

Many approaches to global registration have been developed in recent years. [7] uses a patch based method that sub-divides overlapping regions in an image set and detects feature correspondences to measure displacement. This measurement is used to minimize the projected differences between features resulting in global alignment. The correspondence search of this method performs slowly, and the algorithm becomes biased when moving objects are introduced into the image sequence.

[2] proposed a method that iteratively aligns multiple images while simultaneously reducing lens distortion. In doing so, a large number of parameters must be calculated making this approach computationally expensive for image sequences longer than a few images.

[4] expressed global registration as the product of many local registration matrices. Known image quantities are used to minimize the error of the numerous local registration matrices that can be used to calculate a global registration matrix. Unfortunately, a method for obtaining the said image quantities is not presented. Our method of global registration addresses this issue and allows for a progressive approach to global registration that may be suitable for real-time applications.

[2][5][6] presented graph-based global registration methods that perform the analysis of the image topology in the mosaic space and re-adjusts the matrices between the reference frame and other frames. The re-adjustment of the matrix is based on finding and reallocating shared features within overlapping regions.

[1] presented a graph based approach that minimizes accumulated mis-registration error in pairwise alignments. Concatenation of pairwise mappings is reduced based on a calculated topology. The mapping $M_{i,m}$ is calculated using the shortest path linking image I_i to mosaic space I_m. Using a pre-defined set of grid-points and the concatenated mappings, bundle adjustment is used to minimize projection errors between these points in mosaic space, globally registering the images.

These previous methods have demonstrated good results for many challenging sequences. However, all of these methods perform the global registration as an off-line process. They require all images during calculation making them unsuitable for a real-time process that can provide a globally registered image progressively. To achieve this goal, our global registration method assumes the first image of the sequence to be the reference image, executes the local and global registration as any new image enters the system, and generates the globally registered images. The local and global registration processes are feature-based. Features are identified and tracked as each image of the sequence is processed. The correspondences acquired are used to calculate pairwise transformations between temporally successive images. Once calculated, these transformations are used to map the tracked features to mosaic space. The locations of the mapped features are compared against their actual locations based on the tracking. The discrepancies between mapped and tracked feature locations are used to minimize mis-registration error by modifying the concatenated matrices, thus globally registering the image. The contribution of our method of global registration is in our progressive approach. As the next image in the sequence is processed, overlapping features are found in mosaic space, and the image is registered.

2. Methodology

Figure 2 outlines our approach. The algorithm takes a set of images or frames of a video sequence as input. The order of input images is assumed to be sequential. In other words, the $(k+1)^{th}$ image of the input is acquired immediately after the k^{th} image with more than 50% overlap. Then, each pair of consecutive two images goes through the following three steps to produce a globally registered image. These three steps are indicated as shaded boxes in Figure 2.

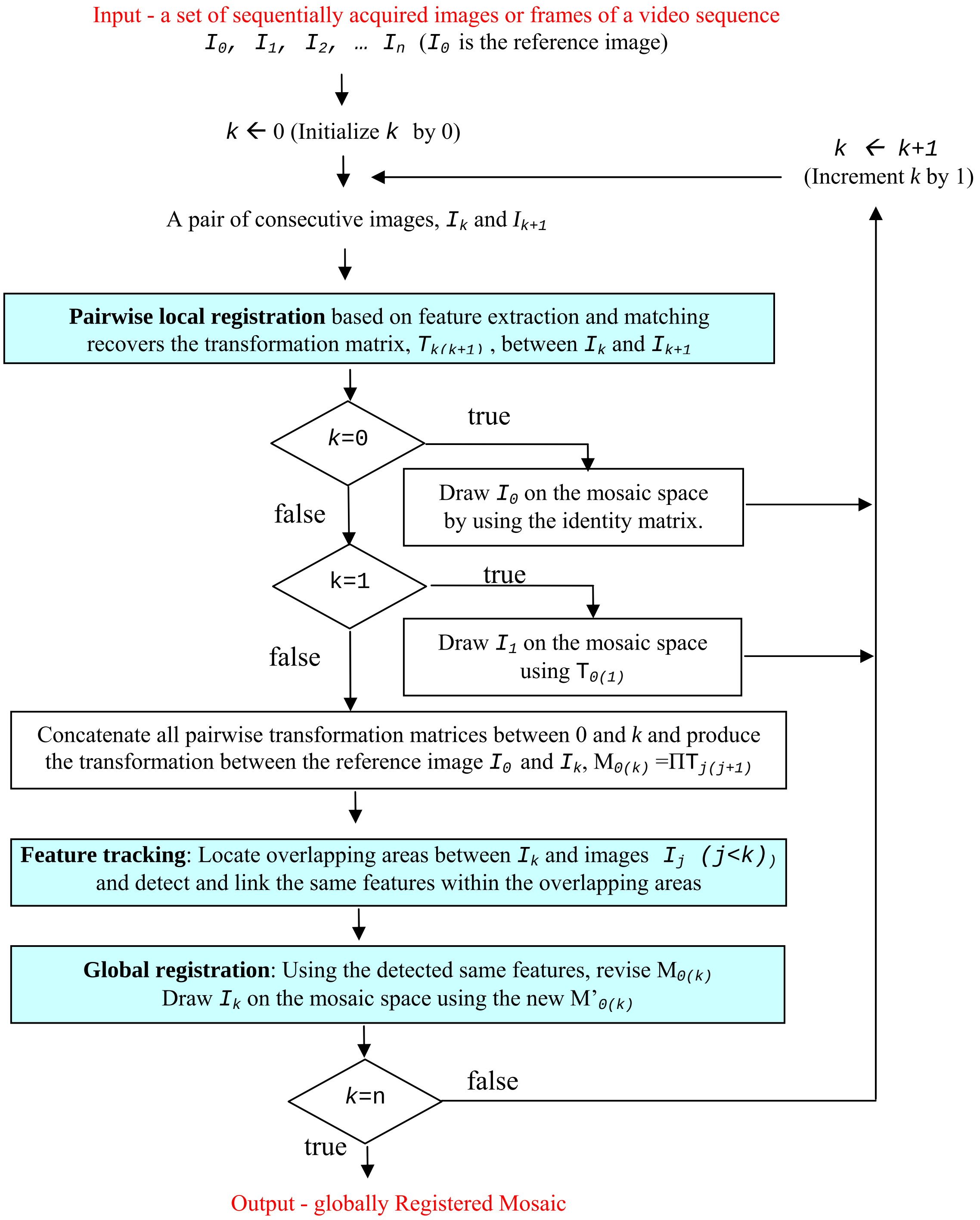


Figure 2 The Proposed Approach for Global Registration

Step 1: Pairwise Local Registration - Pairwise motion, $T_{k(k+1)}$, is estimated for pairs of temporally successive images, I_k and I_{k+1}, using a 2D parametric model. Particularly, we estimate 6-affine parameters in this paper. The motion parameters are calculated based on the feature detection and matching (a.k.a. feature correspondences). Cross-correlation criterion is used to determine correct correspondence. The recovered pairwise transformations are concatenated to create mappings between images that are not temporally successive. That is computed by $M_{0(k)} = \Pi T_{j(j+1)}$ where 0 and k are not successive imges.

Step 2: Feature tracking - Features detected from each image are tracked across temporally and non-temporally successive images. Some features disappear and reappear across images due to occlusions or out-of-frame movement and these features need to be tracked for global registration. To do so, our approach uses pairwise mappings to identify a rough topology of processed images so far, detect spatially overlapping images, and link features from non-successive images using the cross-correlation criterion.

Step 3: Global Registration - Tracked features of each image are mapped to mosaic space, and the mapped location is compared to the location of the same tracked features. The discrepancies between mapped and tracked feature locations are used to recalculate the locally aligned matrices and reduce the error that was accumulated during concatenation, thus globally registering the images.

2.1 Pairwise Local Registration

Feature Detection and Matching

Image regions exhibiting abrupt gradient changes are identified as features. These regions are detected using a simple edge detector. Our approach uses a 2D first derivative operator comprised of 2x2 convolution kernels, as shown in Figure 3. These kernels can be conveniently calculated using the operator in part (c) of Figure 3.

+1	0
0	-1

Gx

0	+1
-1	0

G_y

A	B
C	D

$|G_{x,y}| = |A-D| + |B-C|$

(a) Gradient in x-direction (b) Gradient in y-direction (c) Gradient in both x- and y-directions

Figure 3.

$|G_{x,y}|$ is computed for every image pixel. Then each image is sub-divided into smaller blocks and the location of greatest magnitude in each of these blocks is identified as a feature of that image. Through this sub-division, we guarantee to have a sufficient number of features and the least amount of distance between features. The detected features are used to find corresponding features from the other image of that pair. For each feature extracted from one image, we search for a corresponding feature from the other image. The criterion used to identify the best corresponding feature within a search area is the *cross correlation* and it is defined as:

$$c = \frac{\sum_i \left(I_k(i) - m_{I_k}\right) * \left(I_{k+1}(i) - m_{I_{k+1}}\right)}{\sqrt{\sum_i \left(I_k(i) - m_{I_k}\right)^2} \sqrt{\sum_i \left(I_{k+1}(i) - m_{I_{k+1}}\right)^2}}$$

where $I_k(i)$ is the value of image *k* at *i*, and *m* is the average value of the cross correlation window.

The computation involved in this process greatly depends on the search area per feature. For this purpose, our system provides an efficient GUI that allows the user to click few initial correspondences and uses them to limit the search area for finding other feature correspondences. This method is very useful especially when dealing with a sequence that contains less than 60% overlap. The detected features in temporally successive images are linked for use in both local and global registration. These links allow for the calculation of pairwise mappings during local registration, and the modification of concatenated mappings during global registration.

Affine parameter estimation

Correspondences between temporally successive images are used to estimate pairwise affine transformation parameters. Let us suppose *(x,y)* and *(x',y')* to be corresponding features in an image pair. The affine transformation between these features is defined by the following equations:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

$$x' = ax + by + t_x$$

$$y' = cx + dy + t_y$$

where $a,b,c,d,t_x,$ *and* t_y are the affine parameters. The equations can be rewritten as

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ t_x \\ c \\ d \\ t_y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

We use a set of corresponding features (more than 10 in general) to solve these linear equations using the linear least squares method.

2.2 Feature Tracking

Our approach to global registration is based on feature tracking. As addressed before, pairwise transformations are recovered using feature correspondences, and indirect mappings between non-successive images are calculated by concatenating these pairwise transformations. Unfortunately, the concatenated transformation accumulates a significant amount of error, especially for image sequences exhibiting a non-linear camera path. To correct this accumulated error, we propose to track spatially shared features over time and use the correspondences to modify the concatenated matrix.

While recovering pairwise transformations and concatenating these transformations, our system maintains the list of features shared by several images. In other words, the same features appearing across multiple images will be tracked and linked to each other. It is a challenging task to detect these features and link them together because some features disappear and reappear as the camera moves. In our approach, we use the initially concatenated transformation matrix to estimate the location of the current image with respect to the reference image in the mosaic space. This allows us to identify a subset of images and image regions overlapping with the current image and find missing links of the same features as shown in Figure 4. The same criterion, cross correlation that we used for the local registration, is used to make a decision for linking two features.

2.3 Global Registration

With these tracked feature correspondences, the concatenated matrices can now be modified. To modify these matrices, we first calculate the location of each feature of the current image (I_k) in the reference image (I_0)'s image space (mosaic space). A feature *(x,y)* maps to location *(x',y')* in the reference frame using M_{0k} as in Figure 2. However, the same feature from the previous images might map to a different location *(x'', y'')* instead of *(x',y')* due to the accumulated errors (see Figure 5). Therefore, we need to adjust the current matrix M_{0k} to map *(x,y)* to *(x'', y'')* so that we can globally register the multiple overlapping images. The modification of M_{0k} is achieved by solving the affine equations constructed by the corresponding features *(x',y')* and *(x'', y'')* using the linear least squares method. The result gives the modification necessary to correct the concatenated matrix, thus globally registering the image.

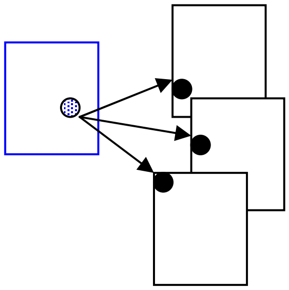


Figure 4. Detection of the same features in three overlapping images in the mosaic space

Moreover, the error for images that do not share common features with the reference image can still be reduced by first mapping the image to an already globally aligned image, and then mapping to the reference image space.

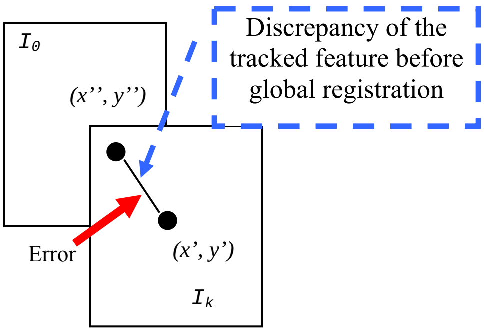


Figure 5. Feature in Image *k* mapped to the reference image I_0.

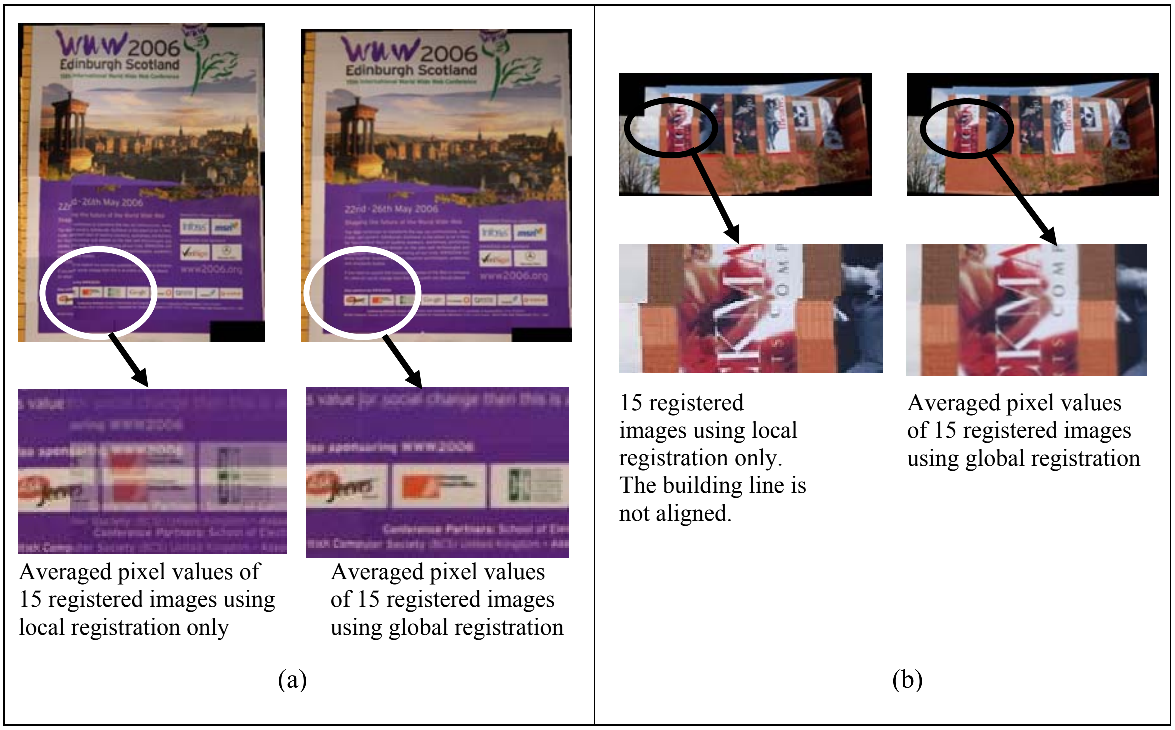
Our global registration process is unique for two reasons. The adjustment will be made for mapping of the current image M_{0k} only. In other words, previously mapped images will not be re-adjusted with respect to the current image. This enables us to perform the progressive global registration that provides the globally registered sequence as a new input image enters. Therefore, the global registration process does not wait until all images of the sequence are available. Knowing the exact location (x'', y'') for (x,y) via feature tracking is a huge contribution of our approach. Many previous methods attempt to minimize the algebraic discrepancy among several transformations M_{0k} computed via different paths to the reference image and they neglect to perform actual feature matching that gives accurate geometric errors. As our approach performs an efficient feature matching and tracking, we use the accurate geometric errors to update the original mapping M_{0k}.

3. Experimental Results

We have tested our global registration method for several challenging image sets. The original input images contain some out-of-focus images, as the images were acquired under a non-controlled camera set-up and free-motion. To reduce execution time, the input images were scaled down from their original size. The image sequence of Figure 6 (a) captured a printed advertisement. The camera path (vertically flipped U path) is non-linear. The registered image using the concatenation of pairwise registration introduces huge accumulated errors, especially in the overlapping area of the first and the last image of the sequence. This mis-registered part was dramatically improved by using our approach. Likewise, the image sequence of Figure 6 (b) is of the Luckman Fine Arts Building at CSULA. The camera motion (horizontally flipped U path) is also non-linear. Again, the local registration based on concatenation of the pairwise transformations produces a final mosaic exhibiting a gross misalignment, especially between the images near the beginning and end of the sequence. After our method was used to globally register all frames, the error was greatly reduced as is seen in Figure 6 (a) and (b). In order to demonstrate the accuracy of the global registration, each pixel value in these result mosaic images is computed as the average of all overlapping pixels. Due to blurry image pixels, some areas of the result also appear blurry. By using median filter, a much clearer mosaic image can be produced.

4. Conclusion

We present a novel approach to the global registration problem. Our approach uses feature tracking to correct the accumulated error caused by the concatenation of pairwise registration. In our approach, the scheme for detection and matching features and parameter estimation for local registration is naturally extended for the global registration in a unified way. The detection of overlapping images and image areas is efficiently computed by analyzing the topology of the images in the mosaic space. In addition, the actual feature comparison within overlapping regions is executed to compute accurate geometric errors between corresponding features and used for the global registration computation. For future work, we plan to enhance our approach to process a video sequence in real-time. The work includes the development of better and faster automatic feature detection and matching and automatic outlier removal using RANSAC in the parameter estimation.



Averaged pixel values of 15 registered images using local registration only

Averaged pixel values of 15 registered images using global registration

(a)

15 registered images using local registration only. The building line is not aligned.

Averaged pixel values of 15 registered images using global registration

(b)

Figure 6. (a) Comparison between locally aligned images and globally aligned images for an image sequence (printed advertisement) captured by a non-linear camera motion (b) Another comparison between locally aligned images and globally aligned images for an image sequence (CSULA Luckman Fine Arts Building) captured by a non-linear camera motion

References

[1] R. Marzotto, A. Fusiello, and V. Murino, “High Resolution Video Mosaicing with Global Alignment”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Vol. 1, pp. 692-698, 2004.

[2] H. S. Sawhney, S. Hsu, and R. Kumar, “Robust Video Mosaicing through Topology Inference and Local to Global Alignment”, in Proceedings of the 5th European Conference on Computer Vision, Volume II - Volume II, June 02 - 06, 1998.

[3] H. S. Sawhney and R. Kumar, “True Multi--Image Alignment and its Application to Mosaicing and Lens Distortion Correction”. CVPR, 1997.

[4] J. Davis, “Mosaics of scenes with moving objects”, in Proceedings of Computer Vision and Pattern Recognition Conference, 1998.

[5] E. Kang, I. Cohen, and G. Medioni, “A Graph-Based Global Registration for 2D Mosaics”, in Proc. of the 5th IAPR International Conference on Pattern Recognition (ICPR), Sept. 3-8, Barcelona, Spain, 2000.

[6] E. Kang, I. Cohen, and G. Medioni, "Yet Another Fast and Robust Feature-based Motion Estimation and Background Layer Extraction", in in Proc. of World Multi-Conf. on Systemics, Cybernetics and Informatics (SCI), July 2004.

[7] H. Shum and R. Szeliski, “Construction and Refinement of Panoramic Mosaics with Global and Local Alignment”, in Proceedings of the Sixth international Conference on Computer Vision (ICCV), 2003.