A Graph-based Global Registration for 2D Mosaics

Eun-Young Kang, Isaac Cohen and Gerard Medioni
University of Southern California
Integrated Media Systems Center
{elkang, icohen, medioni}@iris.usc.edu

Abstract

We describe a graph-based global registration method for creating 2D mosaic images. When multi-frames overlap in space, global registration is necessary to minimize the accumulated registration errors. We use a graph to represent the temporal and spatial connectivity and show that global registration can be obtained through the search for an optimal path in the constructed graph. The definition of an adequate objective function characterizing the global registration provides a direct manipulation of the graph. The framework presented here allows the automatic construction of the graph, and the construction of a consistent mosaic from a collection of frames using projective transformations.

1 Introduction

The creation of a mosaic from a sequence or a collection of images has been popular in recent years due to its large number of applications. Most of the work have focused on the estimation of the parameters modeling the camera motion from a pair of images [4,9]. The obtained results are insufficient to generate a seamless mosaic of a complete scene. Although the pairwise registration is accurate, the concatenation of these pairwise transformations leads to global alignment errors. Therefore, the use of a global registration technique is required to build a consistent large view from a collection of images.

Several methods have been proposed, such as the solution of a linear system derived from the collection of pairwise registration matrices [2] or the frame-to-mosaic scheme [3], which registers each newly acquired frame to the mosaic as it is being built. The methods dedicated to panoramic mosaic construction attempt to minimize the difference between the ray directions of corresponding points in the least square sense [7] or the difference between quadratic surfaces using simulated annealing [6]. Recently, an approach based on a graph representation of the topology of the swaths was presented in order to consider explicitly the 2D topology of the input frames [5].

However, none of these methods did provide a unified framework of a global registration method, which supports: explicit 2D topology representation, error measurement by direct image evaluation and the inexpensive search of correspondences simultaneously.

In this paper, we propose a graph-based approach merging local and global registrations into a single framework. The graph representation of the topology of the swaths allows us to search for the optimal path connecting every frame in the sequence to the considered reference frame. This search is based on a cost function associated to each arc of the graph. This cost reflects the pairwise registration accuracy based on image correlation and a grid deformation representing the set of feature points selected in the mosaic projection space.

2 Pairwise Local Registration

We perform a pairwise registration recovering a projective transformation for every consecutive pair of frames. The projective transformation, $P_{i,i+1}$, between frame *i* and frame *i+1* is computed based on a set of feature points extracted from the two frames. For two arbitrary frames *i* and *j*, the projective transform $P_{i,j}$ is defined by the concatenation of the consecutive projective transformations between *i* and *j*.

The estimation of the projective parameter starts with the initial estimation of a of rigid transformation in the *frequency domain* [1]. This approach overcomes the local minimum problem faced by an arbitrary initialization [9]. The expensive search for the parameters in the frequency domain is reduced by using a *multiresolution approach*. This scheme also allows us to refine the estimated projective parameters gradually by matching two images at different resolutions. A *feature-based* approach is used to match the images leading to the solution of a set of the linear equations derived from the feature-point correspondences. This linear system is solved by *an iterative least square* method as in [8].

3 Global Registration

The obtained pairwise registration parameters provides an accurate transformation for registering pairs of frames. However, the concatenation of these pairwise transformations leads to global alignment errors as depicted in Figure 2(a) and 3(a). This is partly due to the accumulation of errors in the pairwise registration but mostly due to overlooking the topology of the swaths, or relative position of the frames. In order to build consistent large view from a collection of images, temporally and/or spatially adjacent frames have to be registered.

In order to achieve a globally registered mosaic image, given a reference frame R, each projected pixel q in the mosaic has to minimize the following error measure:

$$\varepsilon = \sum_i (I_M(q) - I_i(q'))^2 \qquad (1)$$

where $I_M(q)$ denotes the mosaic intensity level at the considered point q and $I_i(q')$ the intensity of a pixel from the i^{th} frame projected to the same mosaic location, q. Therefore q' is defined by: $q' = P_{i,R}(q)$.

Implementing such an approach is computationally infeasible, unless an efficient representation of the topology of the frames is used. Also, minimizing the error measure given by equation (1) for every single pixel is computationally expensive.

We adopt a graph structure to represent 2D topology of the swath along with the temporal and spatial adjacencies. Also, we use a tessellation of the mosaic into a regular grid instead of considering all the pixels. The points defining this grid are considered as feature points, for which the error measure (1) will be minimized. In the following, we show a global registration of the frames by searching for an optimal path connecting each image to the selected reference frame on top of this representation.

3.1 Graph-based Representation

Frame-Graph: The frame-graph is composed of frames and their correlation. A *node* represents a single frame from the collection of frames to be processed. An *edge* or *arc* is created between two nodes if the corresponding frames are adjacent in time or space. The temporal adjacency produces a linear graph. The spatial adjacency is determined by the overlap between two frames. The approximate overlapping area is computed by projecting the center of each input frames into the mosaic coordinate with respect to the reference frame by using the concatenation of initial pairwise transformation. The arc between nodes is created only if overlapping area is significant and the local registration between these two frames produces a high correlation. Figure 2(c) and 3(c-d) show the constructed graph for the processed sequences given as illustrations of the method. The edge has the correlation between two frames as property.

Properties of the frame-graph: As stated above, we consider the feature points of a uniform tessellation of the mosaic image. For each node of the frame-graph, we locate the grid points of the mosaic. These grid points act as anchors on each frame and represent the core of the proposed global registration. In other words, each node of the frame-graph has a list of the grid points, and each grid point of the node has a list of correspondences to other grid points located in other nodes.

Figure 1 illustrates the selection and representation of these grid points for each frame of the image sequence.

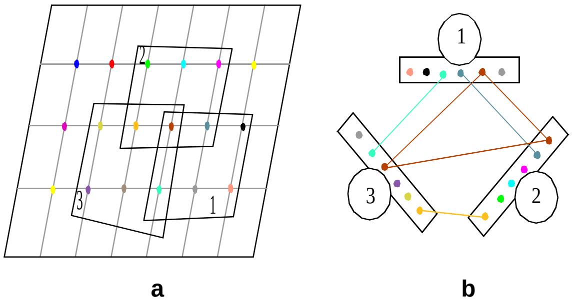


Figure 1. Illustration of the use of the grid points in the graph representation. **a**. Displays the grid points and the rectangle depicting the projected frames. **b**. The corresponding graph representation.

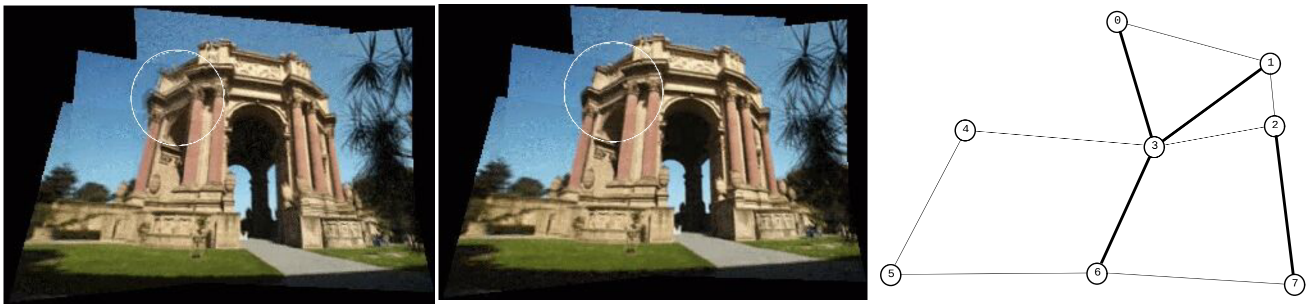
Multiple overlaps, i.e. the presence of a grid point in several frames of the image sequence is used for performing the global registration by minimizing the image error given by equation (1). We compute a local normalized correlation between a feature point f_i of image I_i and f_j of image I_j. This correlation is denoted by $c_{ij}(f_i, f_j)$ where f_i and f_j are the two corresponding pixels in the frames I_i and I_j. In order to simplify the tracking of similar grid points we store also for each pair of pixels f_i and f_j corresponding back projection into the reference frame. This is stored as a relative displacement with regard to the original grid point. This displacement is defined as follows:

$$d_{ij}(f_i, f_j) = P_{i,R}(f_i) - P_{j,R}(f_j) \qquad (2)$$

where $P_{i,R}$ and $P_{j,R}$ are the projection transforms of the frames i and j into the reference frame R and f_i and f_j are the corresponding pixels in the frames I_i and I_j.

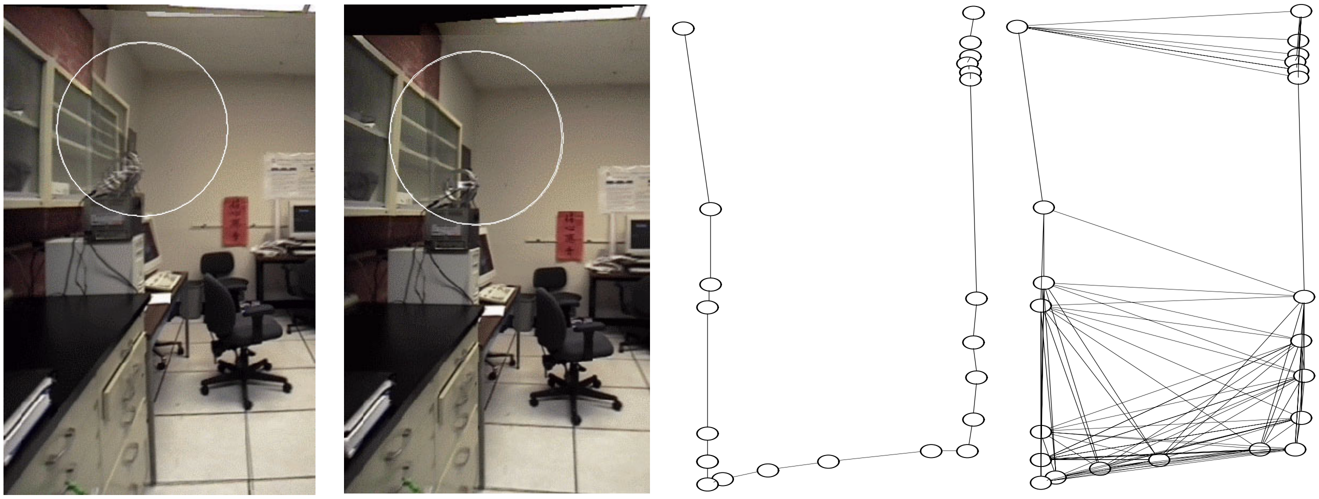
3.2 Selection of the Optimal Paths

The frame-graph representation presented in the previous section relates the temporal and spatial contiguity and provides an efficient tracking of the grid points along the whole collection of frames. For the global registration, the error measure given by equation (1) is performed by finding the set of correspondences for each grid point to adjust the pairwise registration to the selected reference frame.



a. Before Global Registration **b**. After Global Registration **c**. The constructed Graph Structure

Figure 2. The Mosaic Image and its graph structure for the “palace” scene. In the graph, the thin line is initial graph and the thick line is augmented spatial adjacency.



a. Before Global Registration **b**. After Global Registration **c**. Initial Graph **d**. Modified Graph

Figure 3. Another mosaic example for “room” sequence using the proposed global registration method

In order to find a globally registered location for each grid point, we exploit the correspondence information constructed on top of the frame-graph data structure.

The method presented here is based on the selection of the shortest path connecting the two points along camera path. The estimation of the shortest path is based on the geometric distance and the correlation between the grid point f_i and a matching point f_j, the geometric distance between these two point is defined by the norm of the associated offset $d_{ij}(f_i, f_j)$. Similarly, the length of a path connecting two feature points in the graph is the norm of the accumulated offsets. This definition of the path length has the main advantage of privileging the selection of paths that have the smaller geometric distortion. Indeed, the geometry offset are directional, therefore the accumulated geometry offset from a source to a destination via several nodes gives the point with less geometry distortion and similar image intensity level. The shortest path is extracted by a search algorithm which can is carried very efficiently since the costs are cumulative.

Figure 4 illustrates how to find correspondences for a feature point, f_6, through optimal paths.

3.3 Refining Frame Projections

The optimal path approach described above characterizes all the correspondences f_i of a feature point f_m and its real geometric offsets derived by accumulating the estimated geometric offsets along the optimal path. The location of the selected grid point is re-adjusted by considering the centroid of all the reprojected frame locations. The new location of the grid point is defined by the following equation:

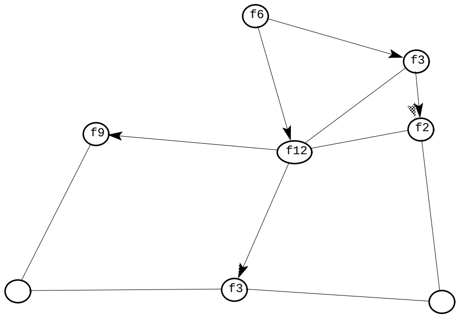


Figure 4. Searching the optimal path connecting a grid point to an image feature point

$$f_m' = \frac{1}{\sum_k w_k(f_m, f_k)} \cdot \left(\sum_k w_k(f_m, f_k) f_k\right) \qquad (3)$$

where the weights w_k are the correlations $c_{ij}(f_m, f_k)$. This allows us to give less weight to erroneous correspondences of the grid point in the overlapping frames. Once we have the new locations of each feature point f_m for a specific node, we can set up a system of linear equations to estimate the adjustment transformation. This adjustment corresponds to the refinement of the concatenation of the pairwise projective transforms. This corresponds to the global registration of the node i with the reference frame and the adjustment transformation is denoted by A_i. Therefore, the global registration of a frame i with a selected reference frame R is defined as follows:

$$P'_{i,R} = A_i P_{i,R} \qquad (4)$$

where P_{iR} is the concatenation of projective transforms. Figure 2(b) and 3(b) show the results using the proposed global registration method.

The system presented processes an AVI video stream of 320x240 color/gray frames at the rate of 6 images per second. For the global registration, it requires $O(mn)$ registrations, where m denotes the maximum number of arcs connected to a node and n denotes the number of input frames processed, usually $m << n$. In the case of a complete graph, we can reduce the number of connections by characterizing the redundancies in the input frame based on the estimation of the rigid transform parameters. The computational complexity associated to finding correspondences and computing the optimal path for each feature points is very low since a fixed mosaic grid is considered and every optimal path contains also all the subpaths corresponding to the overlapping frames. Therefore for each grid points, the optimal path is computed only once, given a reference frame.

4 Conclusion

Global multi-frame registration is a necessary step to produce seamless mosaic. In this paper, we presented a unified framework to represent connectivity and correspondences for the global registration. Also, we proposed an efficient way to find correspondences through an optimal path search in the graph structure that leads to a fast implementation of global registration. An important issue to be addressed in the future is the detection and correction of 3D parallax.

Acknowledgment

This research has been funded by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152.

References

[1] B. S. Reddy and B. N. Chatterji, "An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration", *IEEE Trans. On Image Processing, Vol. 5, No. 8,* 1996

[2] James Davis, "Mosaics of Scenes with Moving Objects", *IEEE Comp, Soc. Conf. On Computer Vision and Pattern-Recognition (CVPR 1998)*, 1998.

[3] H.S. Sawhney, R. Kumar et al., "VideoBrush: Experiences with Consumer Video Mosaicing", *Fourth IEEE Workshop on Applications of Computer Vision*, 1998.

[4] H. Sawhney and R. Kumar, "True Mult-Image Alignment and Its Application to Mosaicing and Lens Distortion Correction", *IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 21, No. 3*, 1999

[5] H. Sawhney and S. Hsu and R. Kumar, "Robust Video Mosaicing through Topology Inference and Local to Global Alignment", *Fifth European Conference on Computer Vision*,1998

[6] Y. Xiong and K. Turkowski, "Registration, Calibration and Blending in Creating High Quality Panoramas", *Fourth IEEE Workshop on Applications of Computer Vision*, Princeton, New Jersey, October 19-21, 1998,

[7] H. Shum and R. Szeliski, "Construction and Refinement of Panoramic Mosaics with Global and Local Alignment", *In Proc. Intl. Conf. on Computer Vision*, 1998

[8] R. Hartley and P. Sturm, "Triangulation", *Computer Vision and Image Understanding,* Vol. 68, No. 2, 1997

[9] R. Szeliski and H. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps", *SIGGRAPH* 1997.