

Improved Video Mosaic Construction by Selecting a Suitable Subset of Video Images

J. S. Jimmy Li and Sharmil Randhawa

School of Informatics and Engineering
Flinders University of South Australia
GPO Box 2100, Adelaide 5001, South Australia

jimmy.li@flinders.edu.au, sharmil.randhawa@flinders.edu.au

Abstract

By stitching together adjacent images from a video sequence surveying a scene, a video mosaic of the entire panorama can be formed. Since a video survey consists of a sequence of images having small relative displacements with respect to each other, there is redundant overlapping information in consecutive images so that not all consecutive video images are required to create a mosaic, and only a subset of suitable images needs to be chosen. Images that are misaligned due to subpixel translation, rotation or shear are difficult to perfectly re-align and stitching of such images can result in a mosaic where discontinuities are noticeable. We propose a new technique for the construction of a seamless mosaic to minimise discontinuities. Our technique partitions an image into four quadrants to register with those of successive images. These registration values are used to form a misalignment index for selecting the best images for stitching.

Keywords: video mosaic, registration, translation, rotation, shear, misalignment

1 Introduction

Mosaics form a larger field of view by stitching successive overlapping video images together. The construction of the mosaic involves two processes: image registration/alignment and image integration/stitching (Peleg and Herman 1997). Registration of two images entails ascertaining the transformation that aligns the images. These transformations include rigid and projective translation. Image integration involves the assembling of the aligned images to form the mosaic. Blending may be used to reduce seams in the stitching process.

There are different existing techniques that can be used to determine the transformation that aligns two images (Szeliski 1996, Tan *et al* 1995, Wolberg *et al* 2000a, Wolberg *et al* 2000b). Images that are aligned subpixelly or with a rotational transformation produce mosaics with

noticeable discontinuities. Hence inverse transformations will be required to re-align these images. Very often, the exact parameters for transformations, e.g. rotation angle, cannot be determined precisely. Incorrect estimation of these parameters may manifest themselves in the mosaic, resulting in visible discontinuities.

A hand-held camera surveying a scene will have a sequence of images with small relative displacements with respect to each other. These images will have rigid translations and/or projective translations. Since a sequence of video images has abundant overlapping areas, not all images are necessary to construct the mosaic. The aim of this paper is to propose a technique for selecting a subset of images for integration in order to minimize alignment errors and produce a mosaic with minimum discontinuities.

2 Detection of Misalignment

Registration involves determining the translation that aligns two images. In particular, we want to be able to detect images with linear translational, rotational and shear transformation. The most appropriate image for integration into a mosaic is the one that is aligned most accurately with the reference image. Accurate alignment can only take place when two images have pure integer linear translational shifts.

The goal of the proposed technique is to find the best image from a group of successive images to align with the reference image. The technique involves partitioning each image into four quadrants, as seen in Figure 1, and registering each quadrant pair of reference and successive images. The underlying principle behind this proposed method is that the registration values of the four quadrants are used as a criterion for choosing the best image for integration. If the registration values for all four quadrants are the same, this implies a linear translational shift. However, if they differ, a rotational, projective transformation or a combination of both has taken place.

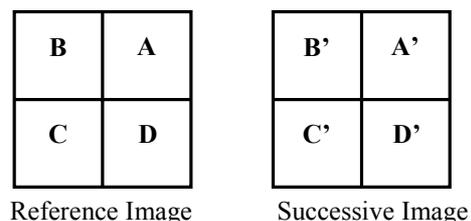


Figure 1: Quadrants of Images

By comparing the registration values of each quadrant pair, e.g. A-A', with other quadrant pairs, we can determine the type of translation involved. Images with linear translational shift between them will return the same registration values for all the quadrant pairs. This is because all the partitions would have shifted the same amount with respect to the whole image. However, those with rotational or shear transformations will give different registration values for the quadrant pairs. A standard deviation σ of these registration values can indicate the existence of misalignment between the two images.

Table 1 shows the registration values of two images and their four quadrants. (x, y) represents the linear translation between the two whole images, and (A_x, A_y) represents the translation of the A-A' quadrant pair along the x and y coordinates, and similar for the other quadrants.

Registration Values	
Whole Images	(x, y)
A-A'	(A_x, A_y)
B-B'	(B_x, B_y)
C-C'	(C_x, C_y)
D-D'	(D_x, D_y)

Table 1: Registration of two images

The standard deviation σ of the four quadrants is given by (1):

$$\sigma = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (1)$$

where σ_x and σ_y are the standard deviations of the x and y registration values of the four quadrants respectively, i.e.

$$\sigma_x = \sqrt{\frac{(A_x - \bar{x})^2 + (B_x - \bar{x})^2 + (C_x - \bar{x})^2 + (D_x - \bar{x})^2}{3}}$$

$$\text{where } \bar{x} = \frac{1}{4}(A_x + B_x + C_x + D_x) \quad (2)$$

and

$$\sigma_y = \sqrt{\frac{(A_y - \bar{y})^2 + (B_y - \bar{y})^2 + (C_y - \bar{y})^2 + (D_y - \bar{y})^2}{3}}$$

$$\text{where } \bar{y} = \frac{1}{4}(A_y + B_y + C_y + D_y) \quad (3)$$

2.1 Phase Correlation Method

The Phase Correlation Method (4-8) is a commonly used image registration technique. It is based on the Fourier shift property, where the translational motion of two images is transformed in the Fourier domain as linear phase differences.

The translational shift (x_0, y_0) between a pair of images, I_1 and I_2 , with translational shift of (x_0, y_0) can be determined by phase correlation (Kuglin and Hines 1975).

$$I_2(x, y) = I_1(x - x_0, y - y_0) \quad (4)$$

The Fourier transform of each image

$$F_1(u, v) = \mathfrak{F}(I_1) \text{ and } F_2(u, v) = \mathfrak{F}(I_2) \quad (5)$$

can be related using the phase shift property.

$$F_2(u, v) = F_1(u, v)e^{-j2\pi(ux_0+vy_0)} \quad (6)$$

Hence the phase shift can be determined by:

$$e^{-j2\pi(ux_0+vy_0)} = \frac{F_1^*(u, v)F_2(u, v)}{|F_1^*(u, v)F_2(u, v)|} \quad (7)$$

where F_1^* is the complex conjugate of F_1 .

Applying the inverse Fourier transform to the phase shift will result in a delta-like function, whose peak denotes the point of registration.

$$\delta(x - x_0, y - y_0) = \mathfrak{F}^{-1} \left[\frac{F_1^*(u, v)F_2(u, v)}{|F_1^*(u, v)F_2(u, v)|} \right] \quad (8)$$

The most remarkable property of the phase correlation method is the accuracy of detecting the peak of the correlation function (Shekarforoush 2002). The phase correlation method provides a distinct sharp peak at the point of registration if there is an integer translational shift between the images. Non-integer translations between two similar images will cause dispersion of the peak of the correlation matrix, spreading it across neighbouring pixels. To identify non-integer shifts in the spatial domain, one approach is to apply interpolation methods to determine the subpixel location of the peak (Tian and Huhns 1986).

The phase correlation method with interpolation has been applied to determine the registration values with subpixel accuracy for our method.

2.2 Detection of Integer Linear Translation

Figures 2 and 3 show two images of the Sydney Harbour Bridge, with linear translational shift. The registered values of the whole images as well as those of the quadrants are identical in value. This is depicted in Table 1. The standard deviation σ between all quadrant registration values is zero.



Figure 2: Reference Image



Figure 3: Linearly Translated Image

Registration Values	
Whole Images	(10, -10)
A-A'	(10, -10)
B-B'	(10, -10)
C-C'	(10, -10)
D-D'	(10, -10)
σ	0

Table 1:

Registered Values of Linearly Shifted Images

2.3 Detection of Non-integer Linear Translation

Figure 4 depicts a non-integer (subpixel) translated image of the reference image in Figure 2. The registration values between the four quadrant pairs are similar, and equal to the registration value of the whole images. The standard deviation σ of the four quadrant registration values is zero, as before. As a result, in order to distinguish between integer and non-integer linear translation, (9) is applied to detect subpixel shift.

$$S = \sqrt{(x - [x])^2 + (y - [y])^2} \quad (9)$$

where (x, y) are values from the whole image registration and $[.]$ is the round function.



Figure 4: Non-integer Translated Image

Registration Values	
Whole Images	(10.5, -9.5)
A-A'	(10.5, -9.5)
B-B'	(10.5, -9.5)
C-C'	(10.5, -9.5)
D-D'	(10.5, -9.5)
σ	0

Table 2: Registered Values of Non-integer Translated Images

2.4 Detection of Rotational Translation

In rotational translation, as in Figure 5, each of the four quadrants will be shifted a different amount in the x and y-directions with respect to each other, and the registration values of the four quadrant pairs will reflect this.

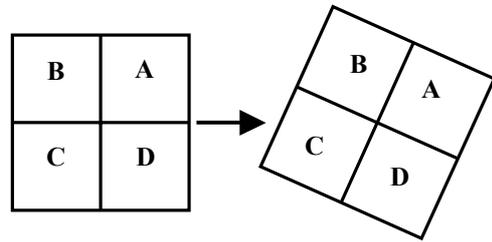


Figure 5: Rotational Translation

Figure 6 shows a rotated image by 1-degree counter clockwise with respect to the reference image. Table 3 shows the registration values of the four quadrant pairs of these images. The non-zero value of σ indicates that translation other than linear has occurred. The standard deviation will increase with the amount of rotation.



Figure 6: Image with Rotational Translation

Registration Values	
Whole Images	(0.9, -1.1)
A-A'	(0.2, 0)
B-B'	(0.9, -1.4)
C-C'	(-0.1, -1.8)
D-D'	(-0.6, 1.7)
σ	1.70

Table 3: Registered Values of Images with Rotational Translation

2.5 Detection of Shear Translation

In shear transformation, the whole image is shifted in the z-direction direction, as seen in Figure 7. This means that each of the four quadrants will shift varying amounts in the x and y-directions with respect to each other, and the registration values of the four quadrant pairs will reflect this. Figure 8 shows a sheared image with respect to the reference image. Table 4 depicts the registration values of the four quadrant pairs of these images. Similar to rotational translation, the non-zero value of σ shows that translation other than linear has occurred. The standard deviation will increase with the amount of shear.

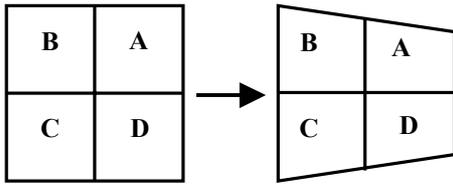


Figure 7: Shear Translation



Figure 8: Image with Shear Translation

	Registration Values
Whole Images	(7.1, -1.7)
A-A'	(7.7, -1.1)
B-B'	(0.9, -6.0)
C-C'	(0, 4.2)
D-D'	(7.2, 1.9)
σ	5.99

Table 4: Registered Values of Shear Aligned Images

3 Misalignment Index

A misalignment index M is introduced to detect and quantify the amount of misalignment with respect to a reference image for a segment of image sequence. M is defined by (10):

$$M = \sigma + K S \quad (10)$$

where σ is the standard deviation of the registration values from the four quadrant pairs, given by (1), S is the amount of subpixel shift, given by (9), and K is a

weighting constant that can be determined experimentally. For small rotations and shears, K can be set to 1. An image with the smallest misalignment index will be selected for mosaic integration.

For images with integer linear shift, the value of the misalignment index will be zero, since σ and S will both be zero.

For images with a non-integer shift, the misalignment index will be non-zero even though σ is zero. This is due to the value of S , which indicates the amount of subpixel shift.

For images with rotational shift or shear translation, the misalignment index is prominently based on σ with a value that increases with the amount of translation. This was clearly illustrated in Tables 3 and 4.

Table 5 shows the misalignment indices for the images in Figures 4, 6 and 8, with $K = 1$, registered with the reference image in Figure 2.

	Misalignment Index
Image in Figure 3 (Integer translation)	0
Image in Figure 4 (Non-integer translation)	0.71
Image in Figure 6 (Rotation translation)	1.85
Image in Figure 8 (Shear translation)	6.31

Table 5: Misalignment Indices for Images

For the above case, the image with linear translation will be selected for mosaic integration since it has the minimum misalignment index. The misalignment index criterion will also be able to handle images with a combination of translations.

4 Quality Index

The algorithm in this technique is depicted as a flowchart in Figure 9. It starts with a reference image, which can be chosen arbitrarily. Generally it can be the first image of the video sequence. The first group of consecutive images to be included for evaluation needs to be determined. The group size is dependent on a number of factors, including the speed of the camera. The image within the group with the minimum misalignment index will be chosen for integration, and this image will be set as the next reference image. The whole process will be repeated for subsequent images until the end of the video sequence. All the chosen reference images are then integrated to form the mosaic. The overall quality of the constructed mosaic can be measured by the reciprocal of the root mean square of the misalignment indices of all images in the selected subset, as shown in (11). The value of Q increases with the quality of the mosaic.

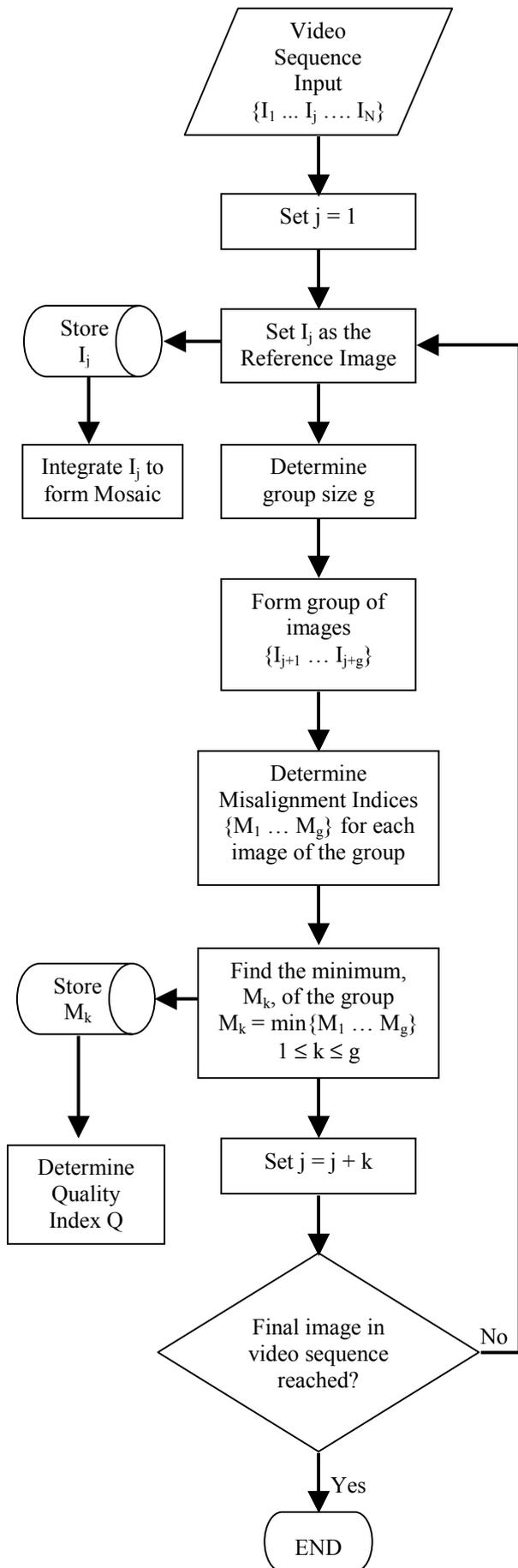


Figure 9: Flowchart of Selecting Suitable Images for Mosaic Construction

$$Q = \frac{1}{\sqrt{\frac{\sum_{i=1}^N (MI_i)^2}{N}}} \quad (11)$$

where MI_i is the misalignment index of the i^{th} image, and N is the total number of selected images in the subset.

If this quality index Q is below a set threshold, then the whole process can be repeated with a different starting reference image. Consequently, a different set of reference images will be selected to form another mosaic with a different Q .

5 Results

Figures 10, 11 and 12 show mosaic constructions of images from a real video sequence taken with a handheld video camera. No inverse transformations were applied in the construction of these mosaics. Figures 10 and 11 illustrate mosaics with higher quality indices than Figure 12. The first reference image for Figure 10 starts at the leftmost position, whereas that for Figure 11 starts at the rightmost. Both images have a similar quality index even though they have different starting reference images. For the mosaic in Figure 12, discontinuities are clearly visible in the sky area. The bridge is visibly disjointed and the umbrellas appear blurred, in particular the second umbrella from the left. Even without any inverse transformations applied, the mosaics in Figures 10 and 11 have almost invisible discontinuities.

6 Conclusion

A technique to select the most suitable video images for improved mosaic construction has been proposed. The technique is based on the idea that the most appropriate images for integration are those images with integer linear translation. A misalignment index determines the suitability of images for mosaic construction.

The images can be partitioned into more than four quadrants, 16 for example, if higher accuracy is needed for larger images.

Since exact translational parameters are often difficult to determine, the advantage of our method is that it very often does not require inverse transformations in order to produce a high quality mosaic, as shown in Figures 10 and 11. This proposed method is also computationally efficient when compared to methods that use inverse transformations.

7 References

- Peleg, S. and Herman, J. (1997): Panoramic mosaics by manifold projection. *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 338-343.
- Davis, J. (1998): Mosaics of Scenes with Moving Objects. *IEEE Proceedings of CVPR'98*, Santa Barbara, CA, USA.

- De Castro, E. and Morandi, C. (1987). Registration of Translated and Rotated Images Using Finite Fourier Transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 700-703.
- Kuglin, C.D. and Hines, D.C. (1975): The phase correlation image alignment method. *IEEE 1975 Conference on Cybernetics and Society*, pp. 163-165.
- Shekarforoush, H.F., Zerubia, J. and Berthod, M. (2002): Extension of Phase Correlation to Sub-Pixel Registration, *IEEE Trans. Image Processing*, vol. 11, Issue 3, pp. 188-200.
- Tan, Y.P., Kulkarni, S.R. and Ramadge, P.J. (1995): A New Method for Camera Motion Parameter Estimation. *In proc IEEE Int Conf Image Processing*, vol.1, pp. 406-409.
- Tian, Q. and Huhns, M.N. (1986): Algorithms for Subpixel Registration. *Computer Vision, Graphics and Image Processing*, 35, pp. 220-233.
- Szeliski, R. (1996): Video Mosaics for Virtual Environments. *IEEE Computer Graphics and Applications*, 16(2), pp. 22-30.
- Wolberg, G and Zokai, S. (2000a). Robust Image Registration using Log-polar Transform. *Proceedings of IEEE International Conference on Image Processing*, Canada, September 2000.
- Wolberg, G and Zokai, S. (2000b). Image Registration for Perspective Deformation Recovery. *SPIE Conf. on Automatic Target Recognition X*, Orlando, Florida, April 2000.



Figure 10: Mosaic with Quality Index $Q = 2.54$ (stitching from left to right)



Figure 11: Mosaic with Quality Index $Q = 2.48$ (stitching from right to left)



Figure 12: Mosaic with Quality Index $Q = 0.02$