Error Equalisation for Sparse Image Mosaic Construction

Bogdan J. Matuszewski, Lik-Kwan Shark, Martin R. Varley
Research Centre for Applied Digital Signal and Image Processing
University of Central Lancashire
Preston, PR1 2HE, UK
bmatuszewski1@uclan.ac.uk

John P. Smith
BAE SYSTEMS
Preston, PR4 1AX, UK

Abstract

This paper presents a new error equalisation method for construction of mosaics built from a large number of images that may not completely cover a scene ('sparse' coverage without all the overlapping images in the neighbourhood). The proposed method is shown to achieve consistent sparse image mosaic construction with the final alignment error no bigger, and in most cases smaller, than the error introduced by the transformation parameter estimation method. The performance of the proposed method is validated using simulation data as well as X-ray images acquired from non-destructive inspection of physically large aircraft components.

1 Introduction

To construct a mosaic, the correspondence between different images has to be established first, such that matched sub-areas of the images represent the same object point (object area). Various methods have been proposed to correctly model the displacement between images in a mosaic. These include modelling displacement for each image pixel or estimating the displacement of a coarse control grid [14] with the movement of each pixel interpolated, usually by two-dimensional splines [16]. Other displacement models often used include projective, affine, similarity or rigid transformations implemented globally or locally [1, 12]. With the transformation model selected next step is to estimate its unknown parameters. Many methods are available to perform this task, overview of which can be found in [1, 3, 10]. Most of the transformation's parameters estimation methods operate on a pair of overlapping images at a time and as a result small errors in alignment accumulate from one pair of images to the next. With large number of images in the mosaic this accumulated error can significantly reduce the overall mosaic quality, making the mosaic globally inconsistent. This is particularly critical for mosaics constructed from a sequence of images, which loops back on itself. There have been number of methods proposed to address this problem. One of the most popular solutions is to align images to the

BMVC 2003 doi:10.5244/C.17.52

actual mosaic as it is being composed [2, 11, 15]. Another method is based on subdividing the sequence of images into subsets, with which internally consistent sub-mosaics can be produced. The final mosaic is constructed from these sub-mosaics [15, 7]. A significantly different approach to the problem is proposed in [4, 13], where transformation's parameters are computed for all images in the mosaic in the same time (simultaneous registration of all images). This enables to impose global consistency constraints on the estimated parameters. The drawback of these methods is their computational cost, which practically prevents the use of pixel based similarity measures for the image registration. The complexity of the global alignment can be reduced by using an error equalisation methodology. In this case local registrations (between overlapping image pairs) are treated independently with the global consistency of the mosaic imposed by introduction of small changes to the estimated local transformations [17, 5, 6].

Using translation as the displacement model, this paper presents a method for mosaicing images that 'sparsely' cover a scene. The proposed method is based on error equalisation methodology and requires that each image in the mosaic is allocated to a single scanning path (given as a direction or a set of directions). Although true image positions can deviate from the scanning path, it is assumed that these deviations are bounded. This new method is a further development of the method previously proposed [8] for the construction of the 'dense' mosaic (complete coverage of the scene) with a rigid transformation used for image alignment.

2 Mosaic Consistency Problem

Mosaic inconsistency is one of the main problems in the construction of a mosaic from a large number of images. A mosaic is said to be consistent when the position of any image in the mosaic is unique (does not depend on the sequence of local transformations used to compute its position). The purpose of the error equalisation is to overcome mosaic inconsistency. Figure 1 shows a diagram representing an example of a mosaic constructed from 5 images.

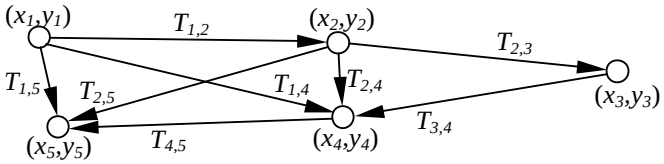

Figure 1: Diagram of a mosaic of five images

Using a common co-ordinate system, the centre positions of m images in a mosaic are denoted by:

$$\{(x_i,y_i) : 1 \leq i \leq m\} \tag{1}$$

Using the differences between the image positions, the local translations required between overlapping image pairs can be computed for the current positions of images in a mosaic and are denoted by:

$$\left\{T_{i,j} : 1 \leq i \leq m; 1 \leq j \leq m;\ j > i;\ \mathbf{M}_{i,j} = 1\right\} \tag{2}$$

where $\mathbf{M}_{i,j} = 1$ indicates ovelapping images i and j, and

$$T_{i,j} = [Tx_{i,j}, Ty_{i,j}]^T = [x_j - x_i, y_j - y_i]^T \tag{3}$$

Since these computed local translations are consistent (because they are obtained from the image positions in the mosaic), it implies, for the mosaic shown in Figure 1, that the following relations have to be true:

$$T_{1,5} = T_{1,2} + T_{2,4} + T_{4,5};\ \ T_{2,4} = T_{2,3} + T_{3,4} \tag{4}$$

In practical image mosaic construction the exact positions of the images are unknown and have to be computed from a sequence of local translations. These are also unknown and have to be estimated from the image contents. This is done for each overlapping image pair and the set of local translations estimated for m images in the mosaic are denoted by:

$$\left\{ \hat{T}_{i,j} : 1 \leq i \leq m; 1 \leq j \leq m;\ j > i;\ \mathbf{M}_{i,j} = 1 \right\} \tag{5}$$

It is unlikely for $\hat{T}_{i,j}$ to be consistent because they are contaminated by some inherent estimation errors and found based on the similarity existing in an image pair without relating to their positions in the mosaic. On the other hand if the centre positions of m images in the mosaic are known the local translations, required between all overlapping image pairs, can be computed as differences between the corresponding image positions, as in equation 3. The local translations computed in such a way are consistent although it does not necessarily mean that the images are properly aligned. Although the local translations, estimated based on image similarity, can be used to place the images into a mosaic as long as no closed loop is formed from them in the process, errors are accumulated as the local translations are cascaded to build the mosaic. Since the accumulated error of the local translations can cause a serious degradation in the mosaic quality, the estimated local translations based on image similarity should not be used directly for mosaic construction. The optimal image positions should not only satisfy the consistency conditions (position of the image should not depend on the sequence of local translations used), but also the local translations computed from them should be as close as possible to the estimated translations based on image similarity. This is achieved via the minimisation of the cost function $E(\mathbf{p})$ with $\mathbf{p} = [x_2, \ldots, x_m, y_2 \ldots, y_m]$ representing the positions of the image centres.

$$E(\mathbf{p}) = \sum_{\substack{i,j:1 \leq i \leq m; 1 \leq j \leq m \\ j > i;\ \mathbf{M}_{i,j} = 1}} \left(\left(Tx_{i,j} - \hat{T}x_{i,j} \right)^2 + \left(Ty_{i,j} - \hat{T}y_{i,j} \right)^2 \right) \tag{6}$$

3 Sparse Mosaic Construction

In sparse coverage of a scene only a small number of estimated local translations are available for the determination of the image positions in a mosaic. This can cause degradation in the mosaic quality, as the positions of the images computed based on minimisation of the cost function $E(\mathbf{p})$ are not sufficiently constrained by the adjacent images. To improve the mosaic quality, it is necessary to introduce extra constraints on the image positions. Figure 2 shows an example of the simulated sparse mosaic, where circles represent the image centres and lines represent the valid local translations computed between overlapping images. The constraints can be divided into two classes. The first class is an

inner-path constraint, illustrated graphically in Figure 3 with image centres represented by black circles, whereas the second class is an inter-path constraint.

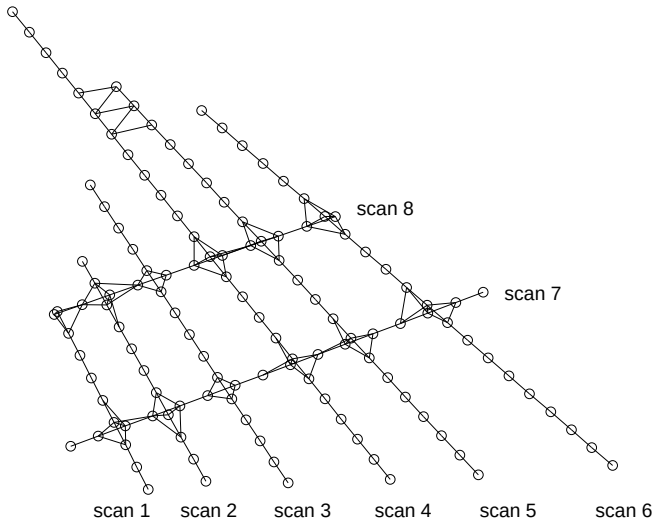

Figure 2: Simulated sparse mosaic

Two different constraints are proposed for the first class. The first type, applicable to linear scanning paths only, is to restrict the distance of an image from the corresponding scanning line.

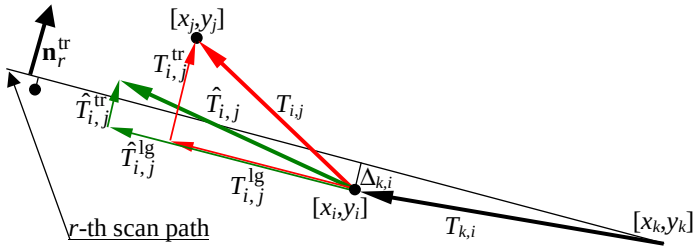

Figure 3: Illustration of constraints in the first class

Knowing the location of a scanning path defined by the centre position of the first image in the scanning path, $[x_k,y_k]$, and the vector normal to the r-th scanning path, $\mathbf{n}_r^{\mathrm{tr}}$, the condition restricting the distance of the i-th image from the scanning line to which this image belongs can be expressed as

$$\left|T_{k,i} \circ \mathbf{n}_r^{\mathrm{tr}}\right| = \Delta_{k,i} \leq d\Delta \tag{7}$$

where $i \in I_r$ and I_r represents a set of indices of all images which belong to the scanning path 'r', $T_{k,i} \circ \mathbf{n}_r^{\mathrm{tr}}$ is the inner product between vectors $T_{k,i}$ and $\mathbf{n}_r^{\mathrm{tr}}$, and $d\Delta$ is the limit allowed for an image location to deviate in the direction perpendicular to the scanning

direction. It is important to notice that the first image in each scanning path is free to move, but its movement affects all the images in the given scanning path; the only exception is the first image in the first scanning path as it is used to anchor the mosaic. The second type of the constraint in the first class is to limit the disparity between the local translations, estimated based on image similarity and computed based on image positions, for all overlapping images in the same scanning path. This constraint is particularly important for images, which overlap not only with images in the same scanning path, but also with images from a different scanning path. Without this constraint minimisation of the cost function given by equation 6 may cause excessive increase of the transversal component of the computed translation. To avoid this, the transversal component of the translation computed based on image positions, $T^{\text{tr}}_{i,j}$ (perpendicular to the scanning direction), should not deviate too much from the same component of the translation estimated based on image similarity, $\hat{T}^{\text{tr}}_{i,j}$. To achieve this, the following constraint is introduced for all overlapping images in the same scanning path:

$$\left|\left(T_{i,j} - \hat{T}_{i,j}\right) \circ \mathbf{n}^{\text{tr}}_r\right| = \left|T^{\text{tr}}_{i,j} - \hat{T}^{\text{tr}}_{i,j}\right| \leq \Delta T^{\text{tr}} \tag{8}$$

where i, j are the indices of the overlapping images in the same scanning path with $i, j \in I_r$; $j > i$; $\mathbf{M}_{i,j} = 1$, and ΔT^{tr} is the difference allowed between $T_{i,j}$ and $\hat{T}_{i,j}$ in the direction perpendicular to the scanning direction. ΔT^{tr} could be interpreted as the local limit for deviations from the scan direction, and $d\Delta$ as the global limit. Based on this interpretation inequality $\Delta T^{\text{tr}} < d\Delta$ can be justified.

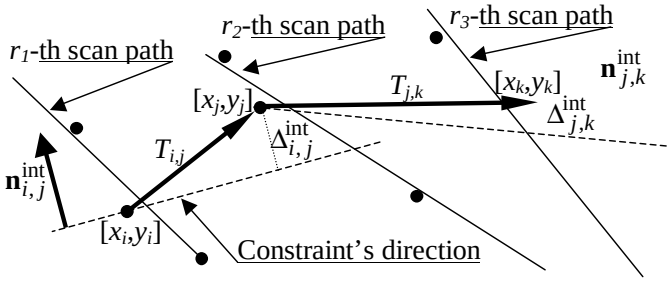

Figure 4: Illustration of the constraint in the second class

The constraint of the second class is illustrated in Figure 4. While the previous two constraints restrict the image positions along a scanning path, this constraint is implemented to control the relative position of the adjacent scanning paths. This is particularly useful when there is no or only a limited number of scans to establish the separation distance between different scanning paths. An example of such a situation would be to construct a mosaic for Figure 2 without scanning paths 7 and 8. The disparity between the direction required by the images in different scanning paths and the direction computed from image positions is constrained using:

$$\left|T_{i,j} \circ \mathbf{n}^{\text{int}}_r\right| = \Delta^{\text{int}}_{i,j} \leq d\Delta^{\text{int}} \tag{9}$$

where $\mathbf{n}^{\text{int}}_{i,j}$ is the vector normal to the direction derived from the estimated positions of images i and j, and $d\Delta^{\text{int}}$ is the maximum deviation allowed. As the number of images

for which this constraint is imposed is small, the constraining directions (vector $\mathbf{n}_{i,j}^{\text{int}}$) can be selected manually or, as in the example described in section five, they can be derived from data obtained from the acquisition system.

4 Minimisation of Cost Function

Writing the cost function given by equation 6 in a matrix form gives

$$E(\mathbf{p}) = \|\mathbf{A}\cdot\mathbf{p} - \mathbf{b}\|^2 \tag{10}$$

where $\mathbf{A}$ is a matrix describing the correspondence between image positions and computed local translations and $\mathbf{b}$ is a vector containing the estimated local translations. Similarly writing the constraints given by equations 7 - 9 in a matrix form gives:

$$\mathbf{C}\cdot\mathbf{p} \leq \mathbf{d} \tag{11}$$

where $\mathbf{C}$ is a matrix describing all the constraints imposed on image positions and $\mathbf{d}$ is a vector describing allowed deviations. To simplify the notation, the two subscript notation (defined by matrix $\mathbf{M}$) is replaced by a one subscript notation using columnwise indexing of matrix $\mathbf{M}$.

Assuming that the whole mosaic is anchored at the position of the first image (x_1, y_1) and this position is fixed at the co-ordinate origin, matrices $\mathbf{A}$ and $\mathbf{C}$, and vectors $\mathbf{b}$ and $\mathbf{d}$ have the following forms:

$$\mathbf{A} = \begin{bmatrix} & \overset{(i-1)}{\downarrow} & \vdots & \overset{(j-1)}{\downarrow} & \vdots & \overset{(i-2+m)}{\downarrow} & & & \overset{(j-2+m)}{\downarrow} & & \\ \cdots 0 & -1 & 0\cdots 0 & 1 & 0\cdots & & & & \cdots & 0\cdots & \leftarrow (k) \\ \cdots 0 & \cdots & & & \cdots 0 & -1 & 0\cdots 0 & & 1 & 0\cdots & \leftarrow (k+1) \\ & & & & \vdots & & \vdots & & & & \end{bmatrix} \tag{12}$$

$$\mathbf{C} = \begin{bmatrix} & \overset{(k-1)}{\downarrow} & \vdots & \overset{(i-1)}{\downarrow} & \vdots & \overset{(k-2+m)}{\downarrow} & \vdots & \overset{(i-2+m)}{\downarrow} & & \\ \cdots & -\mathbf{n}_r^{\text{tr}}(1) & \cdots & \mathbf{n}_r^{\text{tr}}(1) & \cdots & -\mathbf{n}_r^{\text{tr}}(2) & \cdots & \mathbf{n}_r^{\text{tr}}(2) & \cdots & \leftarrow (l) \\ \cdots & \mathbf{n}_r^{\text{tr}}(1) & \cdots & -\mathbf{n}_r^{\text{tr}}(1) & \cdots & \mathbf{n}_r^{\text{tr}}(2) & \cdots & -\mathbf{n}_r^{\text{tr}}(2) & \cdots & \leftarrow (l+1) \\ & & \vdots & & \vdots & & \vdots & & & \end{bmatrix} \tag{13}$$

$$\mathbf{b} = \begin{bmatrix} \vdots \\ \hat{T}x_{i,j} \\ \hat{T}y_{i,j} \\ \vdots \end{bmatrix} \begin{matrix} \\ \leftarrow (k) \\ \leftarrow (k+1) \\ \\ \end{matrix} \qquad \mathbf{d} = \begin{bmatrix} \vdots \\ d\Delta \\ d\Delta \\ \vdots \end{bmatrix} \begin{matrix} \\ \leftarrow (l) \\ \leftarrow (l+1) \\ \\ \end{matrix} \tag{14}$$

where equation 13 gives a part of matrix $\mathbf{C}$ related to the first constraint. From the form of equations 7 - 9 it is obvious that other two constraints can be expressed in a similar way. Equations 10 and 11 present a problem of the constraint linear least squares optimisation

which searches for the minimum of $E(\mathbf{p})$ defined by equation 10 with the constraints imposed on vector $\mathbf{p}$ by equation 11. To solve this problem, the quadratic programming active set method [9] was used.

5 Simulation Results

To investigate the performance of the proposed method, a series of tests were performed for different configurations of scanning paths. The simulation results shown in this section were obtained for the scanning path configuration shown in Figure 2. The image positions in each scanning path were generated randomly. From these image positions, the exact local translations are computed for all the overlapping images. The simulation of the local translations estimated based on image similarity, $\hat{T}_{i,j}$, is done by adding random errors to the exact local translations. The error distribution is uniform over the interval of ± 1 pixel for the transversal components (with respect to the scanning path) and ± 2 pixels for the longitudinal components. Based on the simulated local translations $\hat{T}_{i,j}$, the initial positions of the images in the mosaic were computed based on the sequence of the images in each scanning path and one image pair randomly selected from different scanning paths. Since there are two scanning paths (No. 7 and 8) to fix the relative position between other scanning paths, only the first two constraints were applied. To evaluate the effectiveness of the imposed constraints, histograms of $(\mathbf{C} \cdot \mathbf{p} - \mathbf{d})$ computed for image positions before and after error equalisation are shown in Figure 5. While the left histogram for the initial image positions shows that the conditions 7 and 8 are far from being fulfilled, the right histogram for the image positions after error equalization shows that all the conditions are satisfied.

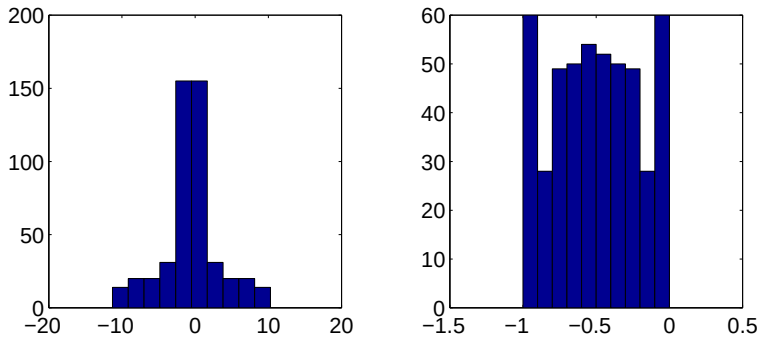

Figure 5: Histogram of $(\mathbf{C} \cdot \mathbf{p} - \mathbf{d})$ for initial image positions (left) and image positions after error equalisation (right)

Figure 6 shows the histograms of the errors computed between the ideal local translations and the local translations computed respectively from the initial image positions and the image positions after error equalisation. As shown in Figure 6, the errors for the initial image positions (7 pixels maximum due to the image inconsistency problem described in Section 2) are significantly reduced by the proposed error equalisation method to below the level of the modelled local translation estimation error (1.8 pixels maximum).

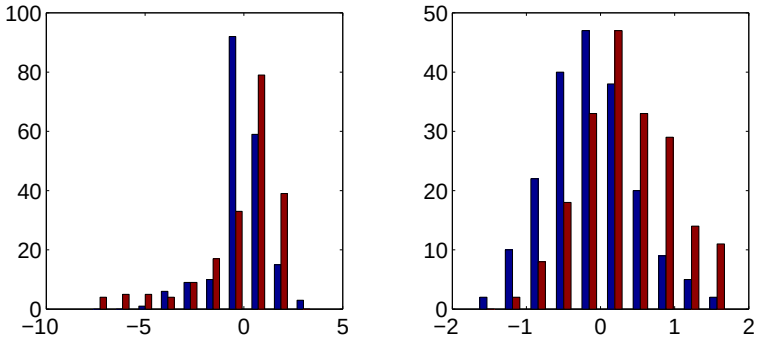

Figure 6: Histogram of local translation errors for initial image positions (left) and image positions after error equalisation (right)

6 Mosaic Construction for X-Ray Images

The described mosaic construction method has been applied to merge a large number of X-ray images obtained using a robotic inspection system into a single image representing the component under inspection. Figure 7 shows the mosaic constructed from 217 X-ray images using image positions obtained directly from the acquisition system, these positions exhibit random errors with maximum errors as big as 12 pixels. Although alignment of the global structures is relatively good, there is significant misalignment between overlapping images.

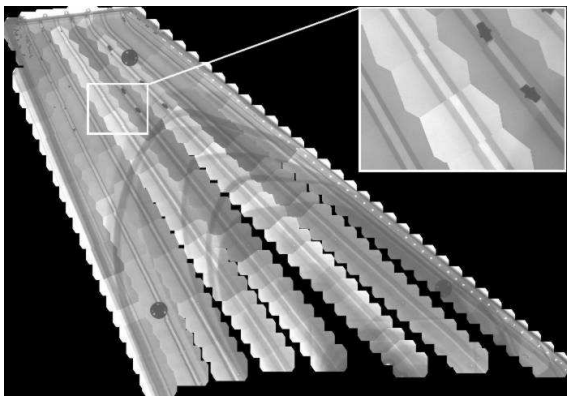

Figure 7: Mosaic constructed directly from the image positions obtained from the X-ray acquisition system

The mosaic built based on the local translations estimated from the image content is shown in Figure 8. All the local translations were estimated iteratively in a multi-resolution coarse-to-fine fashion with the cross-correlation coefficient used as the cost function. The image positions were found using a sequence of local translations along each scanning path from top to bottom. The relative positions between different scanning paths were established based on the local translations estimated using image similarity

between images at the top of each scanning path. In contrast to the mosaic in Figure 7, the images in Figure 8 are well aligned locally, but due to accumulation of the errors along each scanning path there are significant misalignments for global features (arc structures).

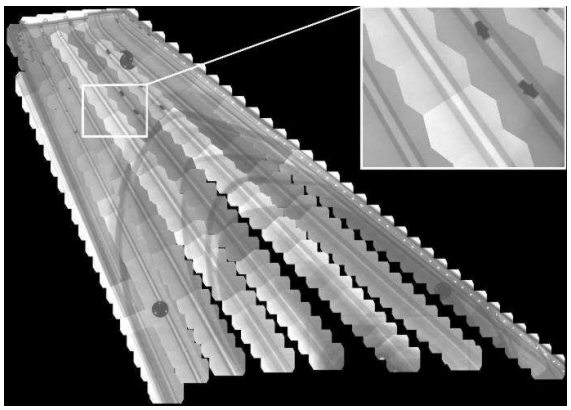

Figure 8: Mosaic constructed based on the estimated local translations without any error equalisation

Figure 9 shows the result of the proposed error equalisation method with all three constraints 7 - 9 implemented. The estimated local translations were exactly the same as those used in the mosaic of Figure 8. The scanning paths and constraints' directions were calculated from the positions obtained from the X-ray acquisition system. Those are exactly the same positions from which the mosaic in Figure 7 was constructed. Four constraints of the form given by equation 9 were used; one at the bottom, one at the top, and two equally spaced across the middle part of the mosaic. From the presented result it can be seen that not only are the images well aligned locally (with sub-pixel accuracy) but also, thanks to the proposed methods, the global structures represented in different scanning paths are fully aligned.

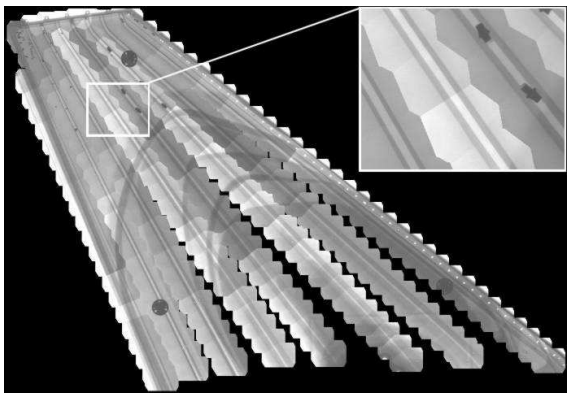

Figure 9: Mosaic constructed by error equalisations with all three constraints implemented

7 Conclusions

The paper presents a new method for automatic sparse mosaic construction with local translation error equalisation. Compared with the mosaic built directly using local translations estimated from image contents, the proposed method is shown to achieve a lower average misalignment error. Due to the high reliability and adequate precision ensured by this method, it is suitable for unsupervised mosaic construction of a large number of sparsely distributed images.

References

[1] Lisa G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[2] Peter Burt, Michael Hansen, and P. Anandan. Video mosaic displays. *Proceedings of SPIE International Society for Optical Engineering*, 2736:119–127, 1996.

[3] Peter Burt, Chihsung Yen, and Xinping Xu. Local correlation measures for motion analysis, a comparative study. In *Proceedings of Conference on Pattern Recognition and Image Processing*, pages 269–274, 1982.

[4] David Capel and Andrew Zisserman. Automated mosaicing with super-resolution zoom. In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pages 885–891, 1998.

[5] James Davis. Mosaic of scenes with moving objects. In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pages 354–360, 1998.

[6] Manuel Guilln Gonzlez, Phil Holifield, and Martin Varley. Improved video mosaic construction by accumulated alignment error distribution. In *Proceedings of British Machine Vision Conference (BMVC'98)*, pages 377–387, 1998.

[7] S. Mann and R.W. Picard. Video orbits of the projective group: A new perspective on image mosaicing. Technical Report 338, MIT Media Laboratory, 1995.

[8] Bogdan J. Matuszewski, Lik-Kwan Shark, John P. Smith, and Martin P. Varley. Automatic mosaicing with error equalisation of non-destructive testing images for aerospace industry inspection. In *Proceedings of NDT World Congress*, Rome, 2000.

[9] Walter Murray Philip E. Gill and Margaret H. Wright. *Practical Optimization*. Academic Press, 1981.

[10] Andr Redert, Emile Hendriks, and Jan Biemond. Correspondence estimation in image pairs. *IEEE Signal Processing Magazine*, pages 29–46, May 1999.

[11] Harpreet S. Sawhney and Rakesh Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pages 450–456, 1997.

[12] Lik-Kwan Shark, Bogdan J. Matuszewski, John P. Smith, and Martin R. Varley. Automatic mosaicing and construction of three-dimensional shearographic surface images. In *Proceedings of the IASTED Conference, (SIP'99)*, pages 183–187, 1999.

[13] Heung-Yeung Shum and Richard Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, 1997.

[14] Christoph Stiller and Janusz Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, pages 70–91, July 1999.

[15] Richard Szeliski. Video mosaic for virtual environments. *IEEE Computer Graphics and Applications*, pages 22–30, 1996.

[16] Richard Szeliski and James Coughlan. Hierarchical spline-based image registration. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 194–201, Seattle, June 1994.

[17] Richard Szeliski and Heung-Yeung Shum. Creating full view panoramic image mosaic and environment maps. In *SIGRAPH*, pages 251–258, 1997.