

## A Modified Automatic Image Stitching Model for Textile and Garment Research Area

Qi.Zhang, YiHong.Zhang and Jing.Xu

**Abstract.** This paper presents some techniques for constructing panoramic image stitching from sequences of images captured by cameras from different angle in garment and textile research area. The image stitching representation associates a transformation matrix with each input image. In this paper, an improved SIFT(Scale Invariant Feature Transform) algorithm was used to locate the feature points in the scanned images targeting at the problem of automatically stitching of textile images which were scanned in divided pieces. RANSAC (random sample consensus) method is proposed to estimate image transformation parameters and to find a solution that has the best consensus with the data. Techniques for estimating and rening camera focal lengths are also presented. In order to reduce accumulated registration errors, global alignment (block adjustment) method is proposed for the whole sequence of images, which results in an optimally registered image stitching. A local alignment technique is also developed which warps each image based on the results of pairwise local image registrations to compensate for small amounts of motion parallax introduced by translations of the camera and other unmodeled distortions.

**Keywords:** panoramic image stitching, invariant features, SIFT(Scale Invariant Feature Transform), RANSAC (random sample consensus), global alignment, local alignment, block adjustment.

1

---

Qi.Zhang (✉)

Development and Liaison Office, DongHua University, RenMin Road 2999,  
ShangHai, China

Email: Zhangqi0901@dhu.edu.cn

Yihong.Zhang

College of Information Science & Technology, DongHua University RenMin  
Road 2999, ShangHai, China

Email: Zhangyh@dhu.edu.cn

Jing.Xu

Development and Liaison Office, DongHua University, RenMin Road 2999,  
ShangHai, China

## 1 Introduction

The automatic panoramic image stitching has an extensive research literature and several commercial applications in the fields of photogrammetry, computer vision, image processing, and computer graphics. For image mosaic applications, global registration is performed based on a variety of predefined camera motion models[1]. For applications aiming to create special effect by engrafting image objects[2], input images are initially registered by manual dragging or assuming static camera configuration. In all these situations, even a small misalignment may cause local intensity or structure inconsistency and produce visual artifacts.

In the previous research literature methods for automatic image alignment and stitching fall broadly into two categories: direct[3],[4],[5],[6] and feature based[7][8][9]. Direct methods have the advantage that they use all of the available image data and hence can provide very accurate registration, but they require a close initialisation. Feature based registration does not require initialisation, but traditional feature matching methods (e.g., correlation of image patches around Harris corners [10][11]) lack the invariance properties needed to enable reliable matching of arbitrary panoramic image sequences.

Some other stitching methods[12][13][14] optimize a blending function that minimizes the intensity difference in the vicinity of the overlapped area in image stitching. There is, however, or structures will also be aligned. Structure misalignment causes image ghosting or blurring artifact, where a salient edge fades out as it enters the overlapped area and fades in just a few pixels away but in a shifted position. To align image features, nonparametric and patch-based techniques have been recently proposed in texture synthesis based on texture deformation[15][16]. To synthesize a natural texture image, the detected features are matched and deformed inside the overlapped texture samples. Therefore, local structure across patch boundaries can be maintained after synthesis. However, these techniques fail to handle input images with significant color or intensity inconsistency. Moreover, complex global structures and detailed patterns typical of natural images will significantly increase the ambiguity in their 2D patch matching process.

In this an optimal invariant feature based on SIFT and RANSAC method was presented for fully automatic panoramic image stitching special for fabric image stitching area. An efficient bundle adjustment implementation including global adjustment and local adjustment was also conducted in this paper. Figure1 shows the proposed automatic panoramic image stitching procedures.

## 2 Featuring Matching

The first step in the panoramic recognition algorithm is to extract and match SIFT features between all of the images. SIFT features are located at scale-space

maxima/minima of a difference of Gaussian function. At each feature location, a characteristic scale and orientation is established. This gives a similarity-invariant frame in which to make measurements. Although simply sampling intensity values in this frame would be similarity invariant, the invariant descriptor is actually computed by accumulating local gradients in orientation histograms. This allows edges to shift slightly without altering the descriptor vector, giving some robustness to affine change. This spatial accumulation is also important for shift invariance, since the interest point locations are typically only accurate in the 0-3 pixel range[19]. Illumination invariance is achieved by using gradients (which eliminates bias) and normalising the descriptor vector (which eliminates gain).

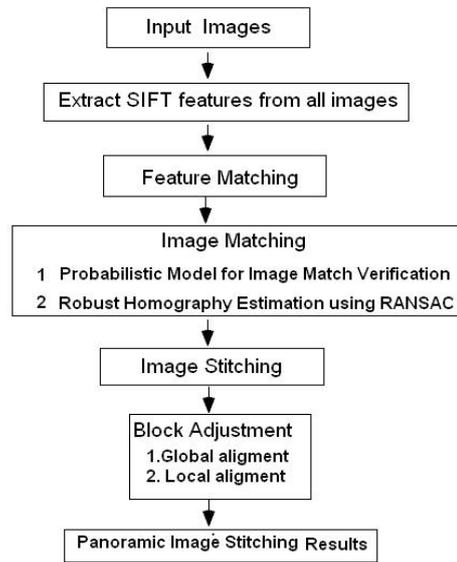


Fig.1 Automatic panoramic image stitching procedures

It is clear that image stitching experiment by using traditional feature matching techniques such as correlation of image patches around Harris corners. Ordinary is not so well and (translational) correlation is not invariant under rotation, and Harris corners are not invariant to changes in scale.

Assuming that the camera rotates about its optical centre, the group of transformations the images may undergo is a special group of homographies. We parameterise each camera by a rotation vector  $\theta = [\theta_1, \theta_2, \theta_3]$  and focal length  $f$ . This gives pairwise homographies  $\tilde{\mathbf{u}}_i = \mathbf{H}_{ij} \tilde{\mathbf{u}}_j$

where

$$\mathbf{H}_{ij} = \mathbf{K}_i \mathbf{R}_i \mathbf{R}_j^T \mathbf{K}_j^{-1} \quad (1)$$

and  $\mathbf{u}'_i, \mathbf{u}'_j$  are the homogeneous image positions ( $\mathbf{u}'_i = s_i [\mathbf{u}_i, 1]$ , where  $\mathbf{u}_i$  is the 2-dimensional image position). The 4 parameter camera model is defined by

$$K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & f_i \end{bmatrix} \quad (2)$$

and (using the exponential representation for rotations)

$$R_i = e^{[\theta]_x}, [\theta]_x = \begin{bmatrix} 0 & -\theta_{i3} & \theta_{i2} \\ \theta_{i3} & 0 & -\theta_{i1} \\ -\theta_{i2} & \theta_{i1} & 0 \end{bmatrix} \quad (3)$$

### 3 Image Mathching

At this stage the objective is to find all matching (i.e. overlapping) images. Connected sets of image matches will later become panoramas. Since each image could potentially match every other one, this problem appears at first to be quadratic in the number of images. However, it is only necessary to match each image to a small number of overlapping images in order to get a good solution for the image geometry.

From the feature matching step, we have identified images that have a large number of matches between them. We consider a constant number  $m$  images, that have the greatest number of feature matches to the current image, as potential image matches (we use  $m = 6$ ). First, we use RANSAC to select a set of inliers that are compatible with a homography between the images. Next a probabilistic model to verify the match will be applied.

#### 3.1 Robust Homography Estimation using RANSAC

RANSAC (random sample consensus)[21] is a robust estimation procedure that uses a minimal set of randomly sampled correspondences to estimate image transformation parameters, and finds a solution that has the best consensus with the data. In the case of panoramas we select sets of  $r = 4$  feature correspondences and compute the homography  $\mathbf{H}$  between them using the direct linear transformation (DLT) method [20]. We repeat this with  $n = 400$  trials and select the solution that has the maximum number of inliers (whose projections are consistent with  $\mathbf{H}$  within a tolerance pixels).

Given the probability that a feature match is correct between a pair of matching images (the inlier probability) is  $pi$ , the probability of finding the correct transformation after  $n$  trials is

$$p(\mathbf{H} \text{ is correct}) = 1 - (1 - (pi)^r)^n \quad (4)$$

#### 3.2 Probabilistic Model for Image Match Verification

For each pair of potentially matching images we have a set of feature matches that are geometrically consistent (RANSAC inliers) and a set of features that are inside the area of overlap but not consistent (RANSAC outliers). The idea of our verification model is to compare the probabilities that this set of inliers/outliers was generated by

a correct image match or by a false image match.

For a given image, it denote that the total number of features in the area of overlap  $n_f$  and the number of inliers  $n_i$ . The event that this image matches correctly/incorrectly is represented by the binary variable  $m \in \{0, 1\}$ . The event that the  $i$ th feature match  $f^{(i)} \in \{0, 1\}$  is an inlier/outlier is assumed to be independent Bernoulli, so that the total number of inliers is binomial

$$p(f^{(1:n_f)} | m = 1) = B(n_i; n_f, p_1) \quad (5)$$

$$p(f^{(1:n_f)} | m = 0) = B(n_i; n_f, p_0) \quad (6)$$

where  $p_1$  is the probability a feature is an inlier given a correct image match, and  $p_0$  is the probability a feature is an inlier given a false image match. The set of feature match variables  $\{f^{(i)}, i = 1, 2, \dots, n_f\}$  is denoted  $f^{(1:n_f)}$ . The number of inliers  $n_i = \sum_{i=1}^{n_f} f^{(i)}$  and  $B(\cdot)$  is the binomial distribution

$$B(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (7)$$

values  $p_1 = 0.6$  and  $p_0 = 0.1$ . The posterior probability could be evaluated that an image match is correct using Bayes' Rule

$$p(m = 1 | f^{(1:n_f)}) = \frac{p(f^{(1:n_f)} | m = 1) p(m = 1)}{p(f^{(1:n_f)})} \quad (8)$$



**Fig.2.** Experiment results by feature matching and image matching by SIFT and RANSAC Method

In **Fig.2**, it is obvious that a visible gap or overlap would exist between the first and last images in a sequence, even if these two images were the same as had been appeared in the previous section.

#### 4 Block Adjustment (Global and local alignment)

The global alignment algorithm is a feature-based technique, i.e., it relies on first establishing point correspondences between overlapping images, rather than doing direct intensity difference minimization (as in the sequential algorithm). To find features, each image was divided into a number of patches (e.g.,  $16 \times 16$  pixels), and

use the patch centers as prospective feature points.

To compensate for these effects, the amount of mis-registration and to then locally warp each image will be qualified so that the overall mosaic does not contain visible ghosting (double images) or blurred details. If our mosaic contains just a few images, we could choose one image as the base, and then compute the optical flow between it and all other images, which could then be deformed to match the base. Another possibility would be to explicitly estimate the camera motion and residual parallax[23], but this would not compensate for other distortions.

The block adjustment algorithm (13) provides an estimate  $p_j$  of the true direction in space corresponding to the  $j$ th patch center in the  $k$ th image,  $x_{jk}$ . The projection of this direction onto the  $k$ th image is:

$$x_{jk} \square V_k R_k \frac{1}{n_{jk} + 1} \sum_{l \in N_{jk} U_k} R_l^{-1} V_l^{-1} x_{jl} = \frac{1}{n_{jk} + 1} \left( x_{jk} + \sum_{l \in N_{jk}} x_{jl} \right) \quad (9)$$

This could be converted into a motion estimate:

$$\bar{u}_{jk} = \bar{x}_{jk} - x_{jk} = \frac{1}{n_{jk} + 1} \sum_{l \in N_{jk}} (\tilde{x}_{jl} - x_{jk}) = \frac{1}{n_{jk} + 1} \sum_{l \in N_{jk}} u_{jl} \quad (10)$$

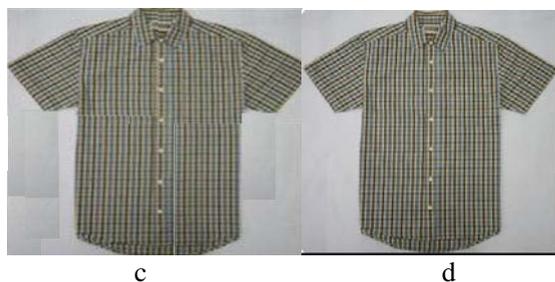
Even though the local alignment was formulated by using rotational mosaic representation, the deghosting equation (14) is valid for other motion models (e.g., 8-parameter perspective) as well. We need only to modify (14) as equation (15):

$$\bar{x}_{jk} \square m_k \frac{1}{n_{jk} + 1} \sum_{l \in N_{jk} U_k} M_l^{-1} x_{jl} = \frac{1}{n_{jk} + 1} \left( x_{jk} + \sum_{l \in N_{jk}} \tilde{x}_{jl} \right) \quad (11)$$

## 5 Experiment Result and Conclusions

Two results of applying our global and local alignment techniques to image stitching experiments were presented . It is widely used in image stitching in modern garment or textile defect detecting system. Figure3 illustrated the results of fabric images stitching without block adjustment and with block adjustment proposed in the present study.





**Fig.3** (a) is the garment images for stitching. (b) is the stitching results without global and local alignment. (c) is the stitching results with global alignment but Without local alignment. (d) is the stitching results with both global alignment and local alignment

**Fig.3** showed a image stitching example which was normally appeared in the textile product industry. This sequence was shot using the camera's automatic mode, which allowed the aperture and exposure time to vary, and the flash to fire on some images. Despite these changes in illumination, the SIFT features match robustly and the multi-band blending strategy yielded a seamless panorama. The output was  $360^\circ \times 100^\circ$  degrees and had been rendered in spherical coordinates  $(\theta, \varphi)$ . All 57 images were matched fully automatically with no user input, and a  $4 \times 57 = 228$  parameter optimization problem was solved for the final registration. The  $2272 \times 1704$  pixel input images were matched and registered in 9 seconds. Tests were conducted on a 2.6GHz Pentium M.

It could be concluded that the proposed method for fabric image stitching is feasible and applicable. It had several advantages over the previous approaches. Firstly, the use of invariant features enabled a reliable matching of panoramic image sequences despite the rotation, zoom and illumination change in the input images. Secondly, by viewing image stitching as a multi-image matching problem, it could automatically discover the matching relationships between the images, and recognise panoramas in unordered datasets.

## References

1. Szeliski R,(2004) "Image Aliment and Stitching :A Tutorial," Technical Report MSR-TR-2004-92, Microsoft Research .
2. Agrawala A, M.Dontcheva,M.Agra(1996) Image Aliment Method.
3. Szeliski, R. and Kang, S. (1995) Direct methods for visual scene reconstruction. In IEEE Workshop on Representations of Visual Scenes. Cambridge, MA, 26–33.
4. Irani, M. & Anandan, P. (1999). About direct methods. In Triggs, B.,Zisserman, A., and Szeliski, R. (eds.), Vision Algorithms: Theory and Practice, number 1883 in LNCS. Springer-Verlag, Corfu, Greece, 267–277.
5. Sawhney, H. & Kumar, R. (1999) True multi-image alignment and its application to mosaicing

- and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243.
6. Shum, H. & Szeliski, R. (2000) Construction of panoramic mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130.
  7. Zoghalmi, I., Faugeras, O., & Deriche, R. (1997) Using geometric corners to build a 2D mosaic from a set of images. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Puerto Rico. IEEE.
  8. Capel, D. & Zisserman, A. (1998) Automated mosaicing with superresolution zoom. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR98)*, pp. 885–891.
  9. McLauchlan, P. & Jaenicke, A. (2002) Image mosaicing using sequential bundle adjustment. *Image and Vision Computing*, 20(9–10):751–759.
  10. Harris, C. (1992) Geometry from visual motion. In Blake, A. and Yuille, A., (eds.), *Active Vision*. MIT Press, pp. 263–284.
  11. Shi, J. & Tomasi, C. (1994) Good features to track. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR94)*. Seattle.
  12. Burt, P.J. & Adelson, E.H. (1983) “A Multiresolution Spline with Applications to Image Mosaics,” *ACM Trans. Graphics*, vol. 2, no. 4, 217–236.
  13. Levin, A., Zomet, A., Peleg, S. & Weiss, Y. (2004) “Seamless Image Stitching in the Gradient Domain,” *Proc. European Conf. Computer Vision*.
  14. Uyttendaele, M., Eden, A. & R. Szeliski, (2001) “Eliminating Ghosting and Exposure Artifacts in Image Mosaics,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.
  15. Fang, H. & Hart, J.C. (2004) “TextureShop: Texture Synthesis as a Photograph Editing Tool,” *Proc. ACM SIGGRAPH '04*.
  16. Jia, J. & Tang, C.K. (2003) “Image Registration with Global and Local Luminance Alignment,” *Proc. Int'l Conf. Computer Vision*, vol. I, 156–163.
  17. Brown, M. & Lowe, D. (2003) Recognising panoramas. In *Proceedings of the 9th International Conference on Computer Vision (ICCV03)*. Nice, vol. 2, 1218–1225
  18. Lowe, D. (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.
  19. Brown, M., Szeliski, R., & Winder, S. (2005) Multi-image matching using multi-scale oriented patches. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR05)*. San Diego.
  20. Hartley, R. & Zisserman, A. (2004) *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, ISBN: 0521540518
  21. Fischler, M. & Bolles, R. (1981) Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM*, 24:381–395.
  22. Kang, S.B. & Weiss, R. (1996) Characterization of errors in compositing panoramic images. Technical Report 96/2, Digital Equipment Corporation, Cambridge Research Lab.
  23. Uyttendaele, M., Eden, A., & Szeliski, R. (2001) Eliminating ghosting and exposure artifacts in image mosaics. In *Proceedings of the International Conference on Computer Vision and Pattern*

Recognition (CVPR01). Kauai, Hawaii, vol. 2, 509–516.