OVERVIEW AND EXPERIENCES IN AUTOMATED MARKERLESS IMAGE ORIENTATION

Fabio Remondino[1], Camillo Ressl[2]

[1] Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland, fabio@geod.baug.ethz.ch
[2] Institute of Photogrammetry and Remote Sensing, TU Vienna, Austria, car@ipf.tuwien.ac.at

KEY WORDS: Features, Orientation, Adjustment, Matching, Automation

ABSTRACT
Automated image orientation is still a key problem in close-range photogrammetry, in particular if wide baseline images are employed. Nowadays, within the image-based modeling pipeline, the orientation step is the one which could be fully and reliably automated, exploiting the potentiality of computer and image processing algorithms. In this paper, we summarize recent developments in this field and apply them in three different workflows to automatically extract markerless tie points from close-range images of different types (video sequence, large and wide baseline images). Furthermore we compare the results obtained from bundle block adjustment using the automatic tie points with the results obtained by manual measurements and show how the accuracies of the automatic tie point extraction can further be improved by including least squares matching techniques.

1. INTRODUCTION

Image orientation is the first and thus very important step within the 3D modeling pipeline. To achieve the best results together with accuracy estimates the image orientation is usually performed by means of a bundle block adjustment. In order to speed up the entire modeling pipeline an automation of the orientation step is necessary. This has to operate on two issues: (i) the automatic measurement of tie points (without requiring to stick markers (i.e. signalized targets) on the object) and (ii) the automatic provision of initial orientation parameters for the bundle block adjustment. Whereas (ii) is nowadays easily solved – once the image correspondences are given – using perspective [Cronk et al., 2006] or projective [Hartley and Zisserman, 2001] geometry based formulations of the relative orientation of calibrated and uncalibrated images, the automatic measurement of *markerless* tie points is still a challenging topic especially in close-range images.

Commercial photogrammetric digital stations have some tools for the automated and markerless relative orientation of stereo pairs (HATS from Helava/Leica, ISDM from Z-I, MATCH-AT from Inpho). These systems, however, are mainly designed for (aerial) images acquired in the photogrammetric normal case and thus they generally fail with tilted close-range images. On the other hand, systems able to automatically calibrate and orient a set of close-range images using signalized *coded target* are already available (e.g. iWitness^{TM}). Commercial systems for automatic measurement of *markerless* tie points in close-range images, however, are still missing.

In the literature a lot of work on automated markerless tie point extraction from images can be found [Beardsley et al., 1996; Fitzgibbon and Zisserman, 1998; Pollefeys et al., 1999; Roth and Whitehead, 2000; Nister, 2001; Mayer, 2005]. Most of these point-based systems rely on very short baseline between consecutive frames and work only based on cross-correlation matching procedures. On the other hand, wide baseline images are also receiving great attention [Matas et al., 2002; Lowe, 2004; Georgescu and Meer, 2004; Tuytelaars and Van Gool, 2004]. Although the reported methods seem to be successfully applied on images with very large baselines and with wide intersection angles, still further research in this area is needed. Therefore automated markerless sensor orientation is one of the most attractive and difficult research themes in close-range photogrammetry and computer vision, in particular if wide baseline images are used.

In this paper we show how the current methods can be applied on three different scenarios and how the accuracy can be improved using Least Squares Matching (LSM) [Gruen, 1985]. In this way we check the feasibility and accuracy of automated extraction of markerless correspondences from different data sets. The found correspondences are then imported in a bundle adjustment software to retrieve the orientation parameters. We also compare the automated results with those coming from manual tie point measurements.

We consider the following three scenarios: 'short-range motion', 'long-range motion' and 'wide baseline' images. 'Short-range motion' sequences have a very short baseline between the images and are typically acquired with a video-camera. 'Long-range motion' sequences contain a significant baseline compared to the distance between camera and scene. Wide baseline images present a very large baseline and the intersection angle of homologues rays can be wider than 20-25 degrees. Different tests have been performed, using self-acquired images with a still-video or a video-camera as well as digitized movies without imposing any hard restriction on the camera motion.

In these images natural tie points are automatically extracted by different strategies with increasing complexity (depending on the base-to-distance ratio). This is to show that, depending on the application, simpler or more involved extraction and matching routines should be applied.

The workflows in these three different scenarios require no or only little manual intervention. The main sources for manual measurements are control point measurements, if required, to define the global scale and orientation of the image block and an initial guess about the disparity between the analyzed frames (only for the long-range motion images).

In all these scenarios we work with calibrated cameras or with fixed interior parameters. Although the interior orientation can be determined by self-calibration using the extracted tie points, the geometry of the images acquired for 3D modeling often

does not allow for an adequate (i.e. accurate and reliable) determination of the interior parameters. In fact typical sequences include images acquired in only one direction or the imaged object shows no depth variation. A weak determination of the interior orientation also deteriorates the accuracy of the object reconstruction. Therefore, in practical cases, rather than carrying out the calibration and reconstruction simultaneously, it may often be better first to determine the camera calibration (including the non-linear distortion parameters) using the most appropriate network (with or without control points) and afterwards recover the object geometry using the calibration parameters.

2. ORIENTATION OF SHORT-RANGE MOTION IMAGE SEQUENCES

Sequences with a 'short range motion' between consecutive frames present a very small parallax (often in one unique direction) which can be exploited during the search of the correspondences. Usually these images are acquired with a video-camera and all the frames are analyzed. Due to the small camera displacement, given the location of a feature in the reference image, the position of the same feature in the consecutive frame is found with a tracking process, as long as the feature is visible and matchable. When the frame-to-frame displacement is larger than a few pixels, the tracking process must be replaced with a more robust stereo matching. Optical flow techniques and feature tracker methods are widely used in the vision community if sufficiently high time frequency sequences are used. One of the most known feature tracker is the one developed by [Shi and Tomasi, 1994], based on the results of [Lucas and Kanade, 1981] and [Tomasi and Kanade, 1991]. More recent works were presented in [Nister, 2001].

We developed a feature tracker based on interest points and Least Squares Matching (LSM). The procedure tracks interest points through the images according to the following steps:

1. *Extraction of interest points from the first image*: different operators like [Förstner and Gülch, 1987], [Harris and Stephens, 1988], [Heitger et al., 1992] or [Smith and Brady, 1997] can be employed.
2. *Prediction of the position in the next frame*: due to the very short baseline, the images are strongly related to each other and the image position of two corresponding features is very similar. Therefore, for the frame at time t+1, the predicted position of a point is the same as at time t.
3. *Research of the position with cross-correlation*: around the predicted position a search box is defined and scanned to identify the position which has the highest cross-correlation value. This position is considered only as an approximation for the corresponding point in frame at time t+1.
4. *Establishment of the precise correspondence's position*: the approximation found with cross-correlation is refined using LSM, which provides a more precise sub-pixel location of the feature.
5. *Replacement of the lost features with new interest points*: new interest points are extracted in the areas where the matching process has failed or if a feature is no more visible in the image.

At the end of the tracking process, the correspondences which are visible in at least 2 frames are used for the successive bundle adjustment to recover the camera parameters.

Some commercial software is available to automatically solve the feature tracking problem [e.g. 3D Equilizer™, MatchMover™, Boujou™]. They work properly with sequences acquired with a video-camera (high frame-rate and very short baseline) and they can reliably extract the image correspondences if there are no rapid changes of the camera position. They are mainly used in the film industry (movies, advertisements) and industrial design. Once the features are extracted, the camera poses are recovered and a virtual object can be seamlessly inserted into the sequence and rendered for special effects.

2.1 Experiments

A video digitized from the television was used to test the tracking system and recover the camera poses. The images are acquired from a helicopter which is flying above a hotel (Figure 1). The hotel is approx. 30 m wide and 10 m high while the mountain in background is ca. 60 m far away from the hotel.

Figure 1: Six images of a sequence digitized from the Swiss TV (SF1) consisting of 240 frames (720x576 pixel).

The tracking process extracted approximately 1900 correspondences which were then used to recover the camera exterior parameters within a bundle adjustment (Figure 2). The project was scaled using the width of an hotel's window.

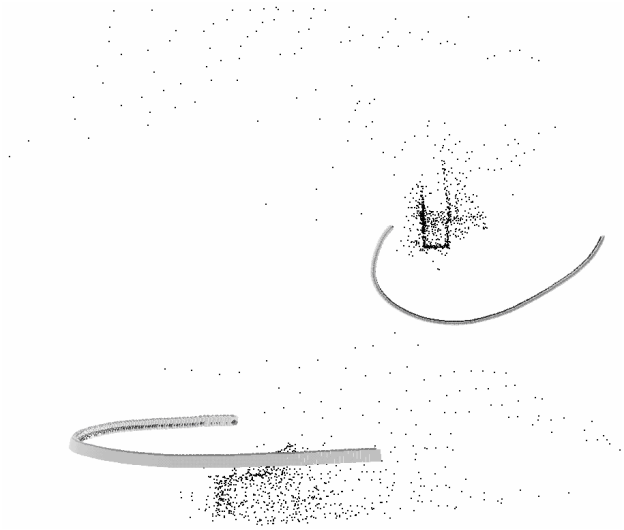

Figure 2: Top view (above) and front view (below) of the recovered camera poses of the sequence acquired from the helicopter.

The final average theoretical precision of the computed object coordinates is σ_x = 0.27 m, σ_y = 0.15 m, σ_z = 0.21 m. As depicted in Figure 2, the smooth trajectory of the helicopter could be successfully recovered. The recovered average distance between two consecutive projection centers is ca 0.5 m. The average distance between the camera and the hotel resulted as ca 45 m while the average distance with the background mountain is ca 120 m.

3. ORIENTATION OF LONG-RANGE MOTION IMAGE SEQUENCES

Long-range motion image sequences contain a significant baseline compared to the distance between camera and scene. They can be acquired with a video-camera (but not all the frames are used) or a still-video camera. The approaches for automatically orienting such image sequences (typically called 'shape-from-video' or 'video-to-3D') [Van Gool and Zisserman, 1996; Fitzgibbon and Zisserman, 1998; Pollefeys et al., 1999; Läbe and Förstner, 2006] require large overlap and good features. In practical situations, such conditions are not always satisfied, due to occlusions, illumination changes and lack of texture.

Our approach, after an initial guess of the average disparity between the images, extracts automatically corresponding points based on the following 5 steps:

1. *Interest points identification*. A set of interest points or corners in each image of the sequence is extracted using detectors like [Förstner and Gülch, 1987], [Harris and Stephens, 1988] or [Heitger et al., 1992]. According to the image size, a threshold on the number of corners extracted is defined and a good point distribution is assured by sub-dividing the images in small patches (e.g. 9x9 pixel on an image of 1600x1200) and keeping only the points with the highest interest value in those patches.

2. *Features matching*. All extracted feature points between adjacent images are matched at first with cross-correlation and then the results are refined using least squares matching (LSM). The point with biggest correlation coefficient is used as approximation for the LSM matching process. Cross-correlation alone cannot always guarantee the correct match while LSM, with patch rotation and reshaping, provides more accurate results. The process returns the best match in the second image for each interest point in the first image.

3. *Epipolar geometry between image pairs*. A pairwise relative orientation and an outlier rejection are performed afterwards. Based on the coplanarity condition, the fundamental matrix is computed with the Least Median of Squares (LMedS) method. Because LMedS estimators solve non-linear minimization problems by minimizing the median of the squared residuals, they are very robust in case of false matches or outliers due to false localisation. The computed epipolar geometry is then used to refine the matching process of step 2, which is now performed as guided matching along the epipolar line.

4. *Epipolar geometry between image triplets*. Not all the correspondences that support the pairwise relative orientation are necessarily correct. In fact a pair of correspondences can support the epipolar geometry by chance (e.g. a repeated pattern aligned with the epipolar line). These kinds of ambiguities and blunders are reduced considering the epipolar geometry between three consecutive images. A linear representation for the relative orientation of three frames is represented by the trifocal tensor T [Shashua, 1997]. T is represented by a set of three 3x3 matrices and is computed using image correspondences without knowledge of the motion or calibration of the cameras. In our process, the tensor is computed with a RANSAC algorithm [Fischler and Bolles, 1981] using 7 correspondences that support two adjacent pairs of images and their epipolar geometry. RANSAC is a robust estimator, which fits a model (tensor T) to a data set (triplet of correspondences) starting from a minimal subset of the data. The found tensor T is used (1) to verify whether the image points are correct corresponding features between three views and (2) to compute the image coordinates of a point in a view, given the corresponding image positions in the other two images. This transfer is very useful in case not many correspondences were found in one view. As result of this step, for each triplet of images, a set of corresponding points, supporting the related epipolar geometry, is recovered.

5. *Tracking image correspondences through the sequence*. After the computation of the trifocal tensor for each consecutive triplet of images, we consider all the overlapping tensors (e.g. T_{123}, T_{234}, T_{345}, ...) and we look for those correspondences which support consecutive tensors. That is, given two adjacent tensors T_{abc} and T_{bcd} with supporting points (p_a, p_b, p_c) and (q_b, q_c, q_d), if (p_b, p_c) in the first tensor T_{abc} is equal to (q_b, q_c) in the successive tensor T_{bcd}, this means that the point in the images a, b, c and d is the same and therefore this point must have the same identifier. Each point is tracked as long as possible in the sequence and the obtained correspondences are used as tie points for a subsequent bundle adjustment.

3.1 Experiments

For the 3D modeling of the empty niche of the Great Buddha of Bamiyan, Afghanistan, five images were acquired with a Sony Cybershot F707. The camera was pre-calibrated in the laboratory. For the image orientation, the tie points were firstly measured semi-automatically by means of LSM and then imported in a bundle adjustment to recover the orientation parameters. Then, the results were used as reference and compared with the results achieved by extracting the tie points automatically.

Figure 3: Three (out of five) images (1920x2560 pixel) of the empty niche of the Great Buddha of Bamiyan, Afghanistan, approximately 60 m high and 20 m wide.

The automated procedure, run with the [Förstner and Gülch, 1987] operator, could extract a high number of correspondences (388 points) which were then used for the image orientation.

Figure 4: The camera poses recovered using the automatically extracted tie points.

After the adjustment, the estimated theoretical precision of the computed 3D object coordinates turned out to be the same for the manual and for the automatic measurements (Table 1). This suggests that for normal network configurations and good image contents, automated markerless orientation procedures can be as good and reliable as manual measurements.

	Manual	Automated
Numb. of images	5	5
Numb. of tie points	24	388
Points in 2 images	-	253
Points in 3 images	24	135
STD X [m]	0.014	0.012
STD Y [m]	0.017	0.019
STD Z [m]	0.021	0.021

Table 1: Comparison between manual and automated tie point measurements. Number of extracted points and estimated theoretical precisions (STD) are reported.

4. ORIENTATION OF WIDE BASELINE IMAGE SEQUENCES

In some applications, due to acquisition constraints or occlusions, images are acquired from substantially different viewpoints. In this cases, the baseline between the images is very large (e.g. Figure 6) and the intersection angle between homologues rays may be larger than 25 degrees. A standard automated tie point extraction procedure, based on corner detectors, would fail because of the big perspective effects generated by the large camera displacement. Due to these effects interest points (e.g. points or corners simply described with their image location) cannot be correctly matched across the images, as:

– The patches in the search image ought to have large enough size in order to contain enough signal information. Due to the big perspective effects, however, the transformation of large patches between template and search image can no longer be described by a simple affine transformation. A small patch would probably allow the matching process, but it might not contain enough signal information to perform correctly the matching.
– The initialization of an LSM refinement will not work, as LSM requires rather precise approximate values for the parameters.

For these reasons, different researchers tried to solve the challenging problem of automatically orienting widely separated views and interest point detectors have been replaced with region detectors and descriptors [Pritchett and Zisserman, 1998; Baumberg, 2000; Matas et al., 2002; Schaffalisky and Zisserman, 2002; Xiao and Shah, 2003; Georgescu and Meer, 2004; Lowe, 2004; Mikolajczyk and Schmid, 2004; Tuytelaars and Van Gool, 2004; Mikolajczyk et al., 2005]. Indeed while corners might be occluded, regions could still be visible and matchable. Generally local features are extracted independently from the images, then characterized with invariant descriptors and finally matched (by means of the Euclidean or Mahalanobis distance between the descriptor elements). These descriptors (usually a vector of attributes) are invariant under affine transformation and illumination changes and can help in matching homologues points in widely separated views.

[Mikolajczyk and Schmid, 2003] have shown experimentally that the Lowe operator [Lowe, 2004] is the most robust algorithm for wide baseline matching and different applications [Roth, 2004; Roncella et al., 2005; Läbe and Förstner, 2006] have also shown its great potentiality.

For the automated orientation of images acquired with a very wide baseline, a strategy has been developed according to the following steps:

1. *Interest regions identification* by means of Lowe detector and SIFT descriptor [Lowe, 2004];

2. Using the vector of attributes extracted by the descriptor, *matching of corresponding points* (centroid of the extracted regions) is done by searching for points with minimal Euclidean distance between their attribute vectors;

3. *Wrong matches removal* by robust computation of the epipolar geometry (described by the fundamental matrix) between image pairs;

4. *Guided matching* by exploiting the epipolar geometry constraint and increasing the localisation accuracy by applying LSM on previously matched regions.

5. Retrieve the *epipolar geometry between image triplets* by means of the trifocal tensor and further refined checking of extracted correspondences.

As shown in [Remondino, 2006] the orientation derived from correspondences based on region detectors/descriptors has worse accuracy than from correspondences based on point detectors. The reason is that, although the correct regions are found to be corresponding, the centroids of the regions (i.e. the points used for computing the image orientation) might be shifted due to perspective effects. However, refining the centroid using LSM, the location accuracy can be improved. LSM requires approximations for the affine transformation parameters. These can be derived from the regions descriptors, which usually include an ellipse representation, whose parameters (major and minor axis and inclination) are derived from the eigenvalues of the second moment matrix of the intensity gradient [Lindeberg, T., 1998; Mikolajczyk, K. and Schmid, C., 2002]. LSM can cope with different image scales (up to 30%) and significant camera rotations (up to 20 degrees), if good and weighted approximations are used to constraint the estimation in the least squares adjustment. An example is shown in Figure 5: given a detected affine region in the template image and its ellipse parameters, LSM is computed in the search image without and with initial approximations, showing the improved matching results depending on the approximations.

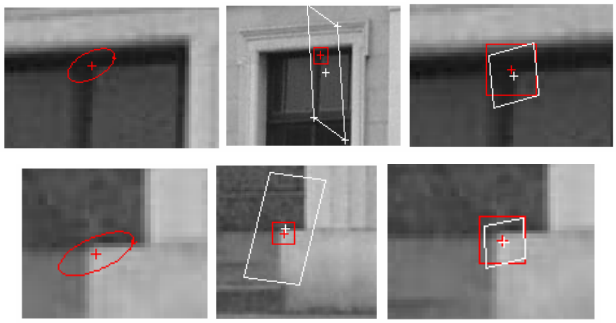

Figure 5: Detected affine regions in the template image (left). Wrong LSM results in the search image (center) with strongly deformed image patches (white patch), initialized by the centroid of the region detector (red patch) and without approximations for the affine transformation parameters. Correct LSM results (right) obtained using the approximations derived from the region descriptors.

4.1 Experiments

A building (Figure 6) has been imaged with two widely separated views. The base-to-distance ratio is approximately 1:0.77 and the camera interior parameters are known. The tie points are automatically extracted as previously described by means of region detectors and descriptors and the extracted correspondences (Figure 7) are used for the relative orientation of the image pair (Figure 8). The final mean RMSE of the image residuals of the computed object coordinates is 0.17 pixels.

A second example consists of three images (Figure 9), representing a Bayon Buddha statue [Gruen et al., 2001], acquired with an analogue Minolta camera and digitized afterwards. Due to the large baseline, there is a very small overlap between the first (A) and the third image (C). The images were firstly processed with the Wallis filter [Wallis, 1976] for radiometric equalization and especially contrast enhancement. The filter enables a strong enhancement of the local contrast by retaining edge details and removing low-frequency information in an image.

For the automated image orientation, the extracted regions are matched between the two adjacent pairs and then the epipolar geometry between the triplet is computed (Table 2).

All the extracted tie points are afterwards imported in a bundle adjustment to retrieve the exterior parameters (Figure 10). A total of 1047 object points are computed and the final RMSE of the image residuals is 0.71 pixels.

Figure 6: Two widely separated images (courtesy of S. El-Hakim, NRC Canada). The base-to-distance ratio is approximately 1:0.77.

Figure 7: The 55 correspondences, which where automatically extracted using the Lowe region detector and used for the orientation procedure.

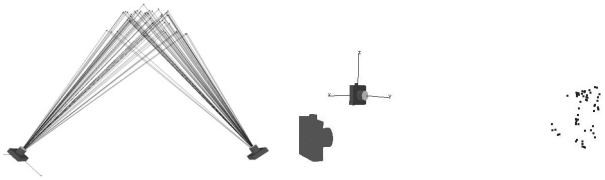

Figure 8: Top and side view of the recovered camera poses of the two widely separated views.

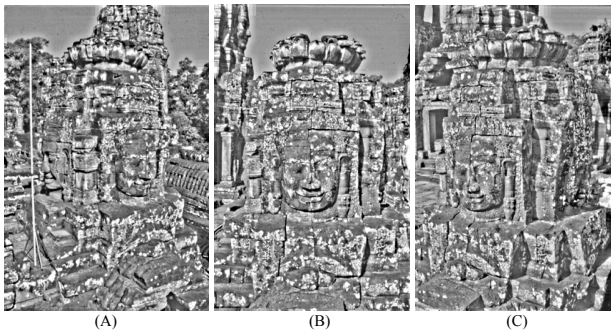

Figure 9: The three analyzed images of the smiling Buddha in Angkor Wat, Cambodia, after the Wallis filter enhancement.

	A	*B*	*C*
Extracted regions	18122	16778	17715
Matched A-B	197		
Matched B-C		902	
New matched A-B after guided matching	16		
New matched B-C after guided matching		7	
Points in 3 images	10		
Total number of 3D points	1047		

Table 2: Results of the tie point extraction between the three widely separated images of Figure 9.

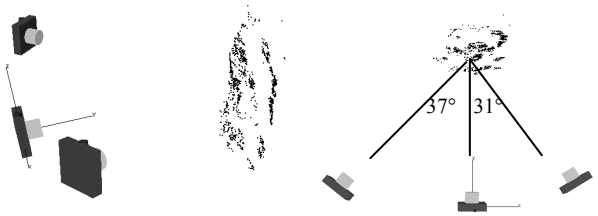

Figure 10: Recovered camera poses of the three widely separated views. The two relative rotation angles between the images are approximately 37 and 31 degrees.

5. CONCLUSIONS AND OUTLOOK

In this article we gave an overview on recent developments in automated orientation of close-range images and demonstrated their applicability on different data sets of real images.

The methods for automated image orientation have to deal with two issues: (i) the automatic measurement of markerless tie points and (ii) the provision of initial orientation parameters for the bundle block adjustment. The second issue is nowadays simply solved (once the correspondences are known) e.g. using orientation methods based on projective geometry (fundamental matrix and trifocal tensor) or with robust relative orientation procedures based on Monte Carlo type strategy [Cronk et al., 2006] – even for uncalibrated images. Although image orientation and calibration are often combined in the literature, we argue that both should be separated if possible. Image sequences acquired for object reconstruction usually do not allow for a proper calibration and consequently, the accuracies of the results will deteriorate. Therefore we prefer to work with calibrated images.

The automatic measurement of markerless tie points, however, is still a difficult and active research topic in close-range photogrammetry and computer vision. A clear fact is that no commercial solutions are still available. Depending on the baseline length between consecutive images, the approaches for automatic tie point extraction can be divided in point and region-based procedures. Whereas the point-based procedures apply simple and well-known extraction and matching techniques, the region-based procedures require more processing time and involved techniques, however, with the benefit that they are able to deal also with wide baseline images. Although the region-based techniques are the most general and are thus also applicable for short baselines, the original results do not exploit the full accuracy potential. However, by using LSM to refine the extracted features, it was shown that the accuracies can be significantly improved [Remondino, 2006].

We can safely conclude that the success of automatically orienting close-range images depends on the following main issues: (i) image arrangement (baselines and viewing directions) and (ii) the imaged scene properties (geometry and texture, even if it was shown that the image content can be enhanced for tie point extraction by image preprocessing).

Even if the image orientation step can be fully automated, within the 3D image-based modeling pipeline some user interaction is still required – especially in the subsequent modeling-phase – as the extracted features, even if well distributed for the orientation, are not sufficient for the object reconstruction, as not located in the salient object areas.

Although the presented examples show the applicability of automated image orientation to a variety of image configurations, in the future we plan to conduct more tests to further investigate the feasibility and limitations of present methods – especially with respect to wide baseline images.

REFERENCES

Baumberg, A., 2000: Reliable feature matching across widely separated views. Proc. of CVPR, pp. 774-781

Beardsley, P, Torr, P. and Zisserman, A., 1996: 3D model acquisition from extended image sequences. Proceedings of ECCV'96, Lecture Notes in Computer Sciences, Vol, 1065, pp. 683-695

Cronk, S., Fraser, C.S. and Hanley, H., 2006: Automatic calibration of colour digital cameras. The Photogrammetric Record. In press

Fischler, M. A. and Bolles, R. C., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, Vol. 24, pp 381-395

Fitzgibbon, A and Zisserman, A., 1998: Automatic 3D model acquisition and generation of new images from video sequence. Proceedings of European Signal Processing Conference, pp. 1261-1269

Förstner, W. and Guelch, E., 1987: A fast operator for detection and precise location of distinct points, corners and center of circular features. ISPRS Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281-305

Georgescu, B. and Meer, P., 2004: Point Matching under Large Image Deformations and Illumination Changes. PAMI, Vol. 26(6), pp. 674-688

Gruen, A., 1985: Adaptive least square correlation: a powerful image matching technique. South African Journal of PRS and Cartography, Vol. 14(3), pp. 175-187

Gruen, A., Zhang, L., Visnovcova, J., 2001: Automatic Reconstruction and Visualization of a Complex Buddha Tower of Bayon, Angkor, Cambodia. Proceedings of 5th International Conference on Optical 3-D Measurement Techniques, Vienna, Austria

Hartley, R. and Zisserman, A., 2001: Multiple View Geometry in Computer Vision, Reprinted Edition. Cambridge University Press, UK

Harris, C. and Stephens, M., 1988: A combined edge and corner detector. Proc. of Alvey Vision Conference, pp. 147-151

Heitger, F., Rosenthalter, L., von der Heydt, R., Peterhans, E. and Kuebler, O., 1992: Simulation of neural contour

mechanism: from simple to end-stopped cells. Vision Research, Vol. 32(5), pp. 963-981

Läbe, T. and Förstner, W., 2006: Automated relative orientation of images. Proceedings of the 5th Turkish-German Joint Geodetic Days, 29-31 March, Berlin

Lindeberg, T., 1998: Feature detection with automatic scale selection. International Journal of Computer Vision, Vol. 30(2), pp. 79-116

Lowe, D., 2004: Distinctive image features from scale-invariant keypoints. IJCV, Vol. 60(2), pp. 91-110

Lucas, B.D. and Kanade, T., 1981: An iterative image registration technique with an applicatiuon to stereo vision. Proc. 7th Intern. Joint Conference on Artificial Intelligence

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2002: Robust wide baseline stereo from maximally stable extremal regions. Proc. of British Machine Vision Conference, pp. 384-393

Mayer, H., 2005: Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. IAPRS, Vol. 36(3/W36), Beijing, China

Mikolajczyk, K. and Schmid, C., 2003: A performance evaluation of local descriptors. Proc. of CVPR

Mikolajczyk, K. and Schmid, C., 2004: Scale and Affine Invariant Interest Point Detectors. Int. Journal Computer Vision, Vol. 60(1), pp. 63-86

Mikolajczyk, K. Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L. 2005: A comparison of affine region detectors. Int. Journal of Computer Vision. In press

Nister, D., 2001: Automatic dense reconstruction from uncalibrated video sequences. PhD Thesis, Computational Vision and Active Perception Lab, NADA-KHT, Stockholm, 226 pages

Pollefeys, M., Koch, R. and Van Gool, L., 1999: Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. IJCV, 32(1), pp. 7-25

Pritchett, P. and Zisserman, A. 1998: Matching and reconstruction from widely separated views. 3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506

Remondino, 2006: Detectors and descriptors for photogrammetric applications. IAPRS, Commission III Symposium, Bonn, Germany. In press

Roncella, R., Remondino, F. and Forlani, G., 2005: Photogrammetric bridging of GPS outages in mobile mapping. Videometrics VIII, Beraldin/El-Hakim/Grün/Walton (Eds), SPIE Electronic Imaging, Vol.5665, pp. 308-319

Roth, G. and Whitehead, A., 2000: Using projective vision to find camera positions in an image sequence. Proceedings of 13th Vision Interface Conference

Roth, G., 2004: Automatic correspondences for photogrammetric model building. IAPRS, Vol. 35(B5), pp. 713-718

Schaffalitzky, F. and Zisserman, A., 2002: Multi-view matching for unordered image sets. Proc. of ECCV

Shashua, A., 1997: Trilinear Tensor: The Fundamental Construct of Multiple-view Geometry and its Applications. Int. Workshop on Algebraic Frames for the Perception Action Cycle (AFPAC), Kiel, Germany

Shi, J. and Tomasi, C., 1994: Good features to track. IEEE Proceedings of CVPR, pp. 593-600

Smith, S.M. and Brady, J.M., 1997: SUSAN – a new approach to low level image processing. IJCV, Vol. 23(1), pp. 45-78

Tomasi, C. and Kanade, T., 1991: Shape and motion from image streams: a factorizaton method - part 3 on 'Detection and Tracking of Point Features'. Technical Report CMU-CS-91-132, Carnegie Mellon University, Pittsburgh, PA, USA

Tuytelaars, T. and Van Gool, L., 2004: Matching widely separated views based on affine invariant regions. IJCV, Vol. 59(1), pp. 61-85

Van Gool, L. and Zisserman, A., 1996: Automatic 3D model building from video sequences. Proceedings of European Conference on Multimedia Applications, Services and Techniques, pp. 563-582

Xiao, J. and Shah, M. 2003. Two-frame wide baseline matching. IEEE Proceeding of 9th ICCV, Vol. 1, pp. 603-610

Wallis, R., 1976: An approach to the space variant restoration and enhancement of images. Proc. of Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, CA