ROBUST MATCHING OF AERIAL IMAGES WITH LOW OVERLAP

M. Mizotin [a], G. Krivovyaz [a], A. Velizhev [a], A. Chernyavskiy [b], A. Sechin [c]

[a] Moscow State University, Dept. of Computational Mathematics and Cybernetics, Russia – {mmizotin, gkrivovyaz, avelizhev}@graphics.cs.msu.ru
[b] State Research Institute of Aviation Systems (FGUP GosNIIAS), Moscow, Russia - achern@gosniias.ru
[c] Racurs Co., Moscow, Russia - sechin@racurs.ru

Commission III, Working Group III/1

KEY WORDS: Image Matching, Image Registration, Aerial Triangulation, Low Overlap

ABSTRACT:

This paper addresses the problem of aerial image matching. We analyze existing approaches to this problem and show that, though the modern algorithms cope with the task quite well, their results are deteriorated in case of low overlap and significant rotation angle between the images. A two-stage feature-based image matching scheme is presented. It is shown that preliminary stage of simplified (shift-rotation) model estimation is crucial in case of low overlap and influences significantly the whole matching scheme. We introduce a novel method of shift-rotation model estimation, based on the voting procedure in parameter space, which allows finding the correct model in case of extremely difficult input data (overlap area is less than 10%) without using any additional information. Finally, the experimental results of our method on both synthetic and real data are presented. We compare our results with one of the state of the art SAC-based model estimators and show that our algorithm outperforms existing methods in case of very small overlap.

1. INTRODUCTION

Automatic aerial image matching, or aerial triangulation, is the field of active research at the moment. Matching is necessary for further creation of orthophotomaps and 3D modelling, both finding an application in such areas as geodesy, cartography, Earth monitoring and others. With the rapid development of geoinformation systems and the growth of input data amount, a fully automated technology of aerial image processing is becoming the main goal in this area.

A lot of different techniques have been introduced recently and commercial software is already available on the market. But the existing state of the art algorithms still are not able to ensure the stable work of automated systems. Because of the complexity of aerial images obtaining process, some difficulties inevitably appear in the input data. Among them are the cases of very small overlap area and significant rotation angle that usually occur when matching a whole block of images divided into sequences. The aim of our work was to develop a technique that outperforms existing matching methods in these cases. Note that image scale is assumed to be constant as it is a typical situation for aerial images. See Figure 1 for an example of input data.

The structure of the paper is the following. In Section 2 we discuss existing approaches to aerial images matching. The two-stage feature-based matching scheme is given in Section 3 and the novel voting-based method of shift-rotation model estimation is introduced in Section 4. We present our experimental results in Section 5. Section 6 concludes the paper.

2. RELATED WORK

Aerial triangulation operates with hundreds or even thousands of images but actually this complex procedure is based on image pair matching. Aerial images are usually very large (hundreds of mega pixels) making it necessary to use hierarchical scale pyramid.
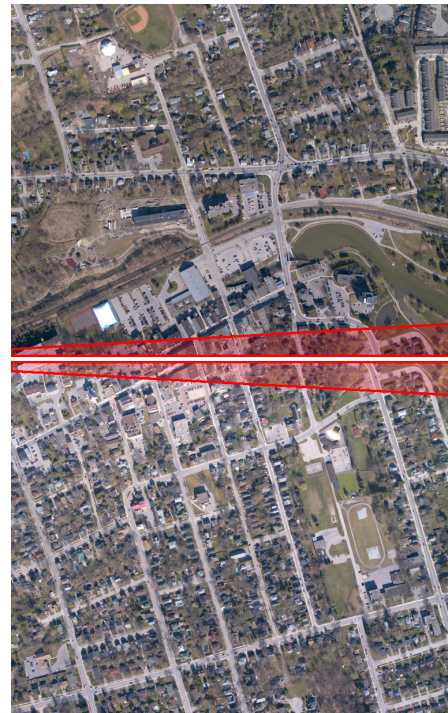

Figure 1. An example of input aerial images. The overlap area (5.7%) detected by the proposed method is outlined

On the highest level of the pyramid two images are small enough to be compared within the whole area. If this step is successful then we only need to refine the matching results on the lower pyramid levels. Thus, we concentrate on the high level image matching in our work.

Image matching methods can be classified into two separate groups: area-based and feature-based. The area-based methods (Kaneko et al., 2003; Zheng et al., 1993; Grün, 1985) normally apply a floating window to compare source and target images using the correlation technique. The most popular similarity measures are the sum of squared differences and the normalized cross-correlation. According to (Zitova et al., 2003), the straightforward correlation-based approaches are practically applicable only in case of shift and small rotation between images due to a huge computational complexity needed for determination of an arbitrarily rotation angle. The more advanced group of methods employing Fourier transform, following the idea of (Reddy et al., 1996), can handle arbitrarily rotation, but these methods are dealing mainly with the case of different image scales when one of the images may be considered as a template, so that the overlap area is quite big.

Feature-based methods (Lowe, 2004; Matas et al., 2002; Flusser et al., 1994) exploit different kinds of features presented in the image such as segments, lines, curves and interest points. Specially constructed descriptors are computed for each feature in order to be compared and to produce a set of feature pairs (putative matches). We refer to (Tuytelaars et al., 2008; Li et al., 2008; Remondino et al., 2006) as exhaustive surveys on feature detectors and descriptors. Finally, an affine or an epipolar model is fitted by the use of robust model estimators. Feature-based methods do not depend on the rotation transform, but their crucial weak point is robust model fitting. One of the most popular robust estimators is the RANSAC algorithm (Fischler et al., 1981). It can handle a significant percent of false putative matches (outliers) efficiently, but the probability of finding the correct model decreases rapidly when outlier level exceeds 50%. But this basic algorithm is being constantly improved with a new modification of RANSAC being published practically each year. In our work we refer to PROSAC (Chum et al., 2005) as one of the best modifications for matching purposes (Choi, 2009) that have been proposed up to date. The PROSAC algorithm can employ information about descriptor's similarity which increases the probability of finding a correct model even if the number of outliers is much higher than 50%. However, as we show in Section 5, even PROSAC fails to provide enough robustness in case of very low image overlap.

Meanwhile, some special methods of matching slightly overlapping images have been already introduced. In the work (Pires et al., 2004) authors use a special adaptive sliding window. But this method was tested only on a few samples and there was no example with a considerable rotation. The paper (Begg et al., 2004) introduces a brute force method. In order to estimate image overlap authors compute image similarity for all possible image position on the highest pyramid level. This method works only with shift transformation. In our opinion, the advanced feature-based matching scheme (see Section 5 for details) outperforms all other existing methods, thus we compare our algorithm only with this scheme.

In the context of this paper it is important to refer to the method described in (Xiong et al., 2006) which is based on a voting procedure. Voting schemes are widely used in computer vision algorithms. One of the most popular methods is Hough transform (Hough, 1962) which is very robust to noise and can be used in very challenging cases of parameters estimation. Unlike SAC-based methods, adaptability of Hough transform depends on model complexity. But if it is possible to carry out the voting procedure efficiently, this approach turns out to be quite powerful as it is fully deterministic, its runtime does not depend on the inliers percentage and it localizes the peak in parameters space more precisely because all points are used, not only a subsample. The method (Xiong et al., 2006) does not use descriptors at all. Image matching is divided into two stages: rotation and scale estimation. On the first step lots of corners are detected in both images and small image patches surrounding each corner are extracted. For each patch the main angle direction is computed using principal component analysis of pixel intensities. Then the voting procedure is used to get the rotation angle. For this purpose the differences between each pair of corners are summarized into angular bins. The maximum of histogram defines the angle between images. This method works only with heavily overlapped images but the proposed voting idea is very promising. We also point to the papers (Seo et al., 2003; Seedahmed et al., 2003) in which very similar matching approaches are proposed.

3. IMAGE MATCHING OUTLINE

In this paper we consider the feature-based image matching scheme that consists of the following consecutive steps:
1) Putative matching
2) Shift-rotation model estimation
3) Overlap detection
4) Rematching in overlap area
5) Final model estimation

Although this scheme seems to be quite obvious, we have not seen it mentioned in the literature. In this paper we focus on the second step, which is discussed in detail in the next section.

Putative matching and model estimation are the two basic steps of feature-based image matching process. But in case of low overlap, when the percentage of true point matches is low, very few points are usually left after model estimation. Even if those points are inliers the estimated model may be too rough to use it on the lower levels of image pyramid, especially if the desired model is a complex one (e.g. a fundamental matrix). In this case one can first estimate the overlapping region by computing a simple shift-rotation model and then repeat the matching procedure only within that region. This produces more matching points uniformly distributed along the overlap area and, as a result, a complex model can be estimated robustly and precisely. Thus, in our work we concentrated on creating a robust shift-rotation model estimation algorithm.

4. A VOTING SCHEME FOR SHIFT-ROTATION MODEL ESTIMATION

The main contribution of this paper is a new method for shift-rotation model estimation based on a voting scheme. Unlike the algorithm in (Xiong et al., 2006), the proposed method takes putative matches as an input. As we deal with the highest level of the image pyramid, it is possible to carry out an efficient putative matching procedure before running model estimation (about the way we obtained putative matches - see Section 5).

The workflow of our algorithm is divided into two stages. At the first stage the rotation angle between the images is

estimated. It is absolutely necessary to compensate this angle in order to further shift estimation work correctly. Recall that our algorithm deals with the general case of an arbitrary rotation angle. Once the rotation is compensated, the second stage begins and the shift is computed. Both rotation and shift estimation are performed in the way similar to Hough transform. These two stages are separated because the applied voting technique is computationally efficient only if the dimensionality of accumulator is not greater than 2. Thus, we independently employ voting for an angle in a 1D space and voting for a shift in a 2D space. The detailed description of these voting procedures is given below.

4.1 Rotation angle estimation

To compute the rotation angle between the images we run through all available pairs of matches and accumulate their votes. Let M be the set of putative matches and $m_1 \in M, m_2 \in M$ - two concrete matches from this set. Assume $m_1 = \{P_1, P_1'\}$ and $m_2 = \{P_2, P_2'\}$, where P_1, P_2 and P_1', P_2' are the points within the images I and I' correspondingly. Then the vector $\overline{P_1 P_2}$ from the image I corresponds to the vector $\overline{P_1' P_2'}$ from I' and the rotation angle can be extracted out of this correspondence. Each pair of matches votes for its own angle, and then the accumulator is smoothed with a Gaussian filter. The highest peak in accumulator corresponds to the desired rotation angle between the images (see Figure 2). The choice of the bin size and the Guassian width will be discussed in Section 5. The ratio of the second highest peak to the first one can be used as an indicator of solution reliability.
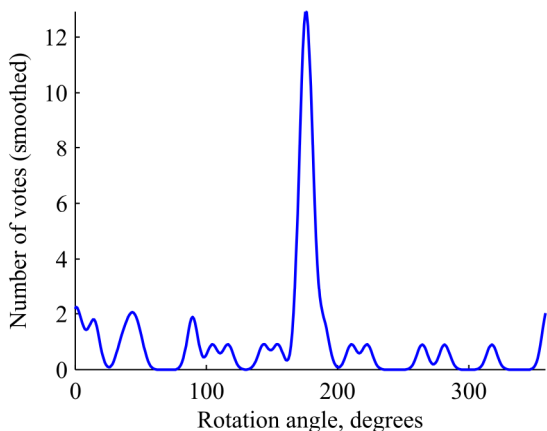

Figure 2. An example of a smoothed angle votes accumulator

In order to preserve the accumulator from too much noise that comes with outliers we applied a simple metric limitation to each pair of putative matches. A pair's vote is considered only if the following condition is met: $\rho(P_1, P_2) - \rho(P_1', P_2') \le d$, where $\rho(x, y)$ denotes the distance between points x and y within the image and d is the threshold. Note that it is not a kind of threshold that needs to be accurately tuned. As incorrect putative matches in most cases include points that are placed in arbitrary way within the image, the metric error is usually large enough for outliers. But at the same time, due to perspective projection, this difference for some inliers may be considerable. Thus, the threshold should not be too strict. The value of 4% of image dimensions for this threshold worked fine in all our experiments.

The similar technique can be employed for scale change detection if the used descriptor is scale-invariant. In this case, first of all, the corresponding pairs of points vote for the scale change $\rho(P_1, P_2)/\rho(P_1', P_2')$ and then the angle is estimated as described above.

It is clear that the complexity of the described procedure is $O(n^2)$, where n denotes the number of putative matches. In order to speed-up computations when the working time is crucial, one can limit the number of pairs to run through. But to avoid a significant loss of quality it is necessary to sort the matches beforehand. The Euclidian distance between descriptors can be used as a metric of reliability of a match. Once all the matches are rearranged in the order of decreasing reliability then it is reasonably to form pairs only with first $k < n$ matches from the list.

4.2 Shift estimation

Shift estimation starts only after rotation of the images has been compensated. At this stage each putative match votes for its own shift along both axes, thus filling the 2D accumulator. The accumulator is then smoothed with a Gaussian filter in order to achieve some robustness to distortion that is present in voting inliers due to perspective effects and non-ideal localization of feature points by detector. The element with the maximum value in the smoothed accumulator is found and the highest peak is localized by approximating the point and it is neighbourhood with a 2D Gaussian function. The least squares method is used for that approximation.

The second highest peak is also detected (see Figure 3) and the ratio of this peak's height to the height of the largest peak may be used as a measure of reliability of the found solution, as at the angle estimation stage. If this ratio is not greater than 0.5, then the shift has been estimated correctly with high probability.
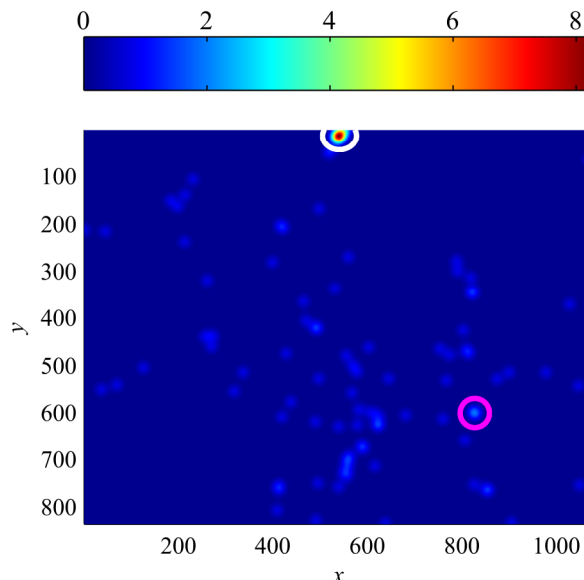

Figure 3. The highest peak (white ellipse) and second highest peak (violet ellipse) detected in the 2D accumulator. Logarithmic scale is used

5. EXPERIMENTS

To confirm the efficiency of our model estimation algorithm we have compared it with another possible overlap detection scheme which is based on sample consensus technique. Having two overlapping images and a set of putative matches, we apply sample consensus to estimate the shift-and-rotation model. We have chosen PROSAC among other sample consensus methods due to its high convergence speed, ability to cope with high outlier's level and suitability to our task: provided putative matches, it is possible to sort them using Euclidean distance between descriptors as a quality function.

We conducted our experiments on both synthetic and real data. Their detailed description is given below.

5.1 Synthetic tests

The synthetic test consists of a number of trials (we used 1000), with the full process of image pair matching being modeled at each trial. First, we randomly choose shift and rotation angle for the pre-selected percentage of image overlap, thus defining the correct transformation. A number of points are placed in the overlap area of the first image and transferred to the second image. This forms the correct matches (inliers), which are then perturbed by noise. After that, false matches (outliers) are added, being placed randomly within the images. The number of inliers and outliers is chosen in such a way that the percentage of inliers is equal to the percentage of overlap area and the total number of matches equals the desired value N.

The important aspect that must be considered in synthetic tests is the need of match ordering for PROSAC method. We take this into account by using the following procedure. The correct matches are assigned random descriptor distances according to the distribution provided in (Läbe et al., 2006). We modeled it with the Rayleigh distribution with $\sigma = 1$ and scaled by a factor of 10000. The incorrect matches are assigned uniformly distributed distances from the $[10^5, 3 \cdot 10^6]$ interval. The used values for descriptors distances are characteristic for the SIFT (Lowe, 2004) descriptor that we use for the real data (see Section 5.2). As it was already mentioned in the previous section, match ordering can also be employed by the proposed method to save computational resources. At the stage of angle estimation the number of combinations can become huge if

there are too many matches, thus it is reasonable to select only the first k matches, which are the most reliable ones, for this process.

Applying the described procedure, two tests have been made, both using $N = 1000$ matches with $k = 200$ most reliable of them employed for angle detection. The two experiments differ in the level of noise added to inliers: the first test was held with the maximum deviation of 4 pixels, while 8 pixels were used in the second one. The threshold for PROSAC was set according to this noise level parameter. For both methods, after inlier detection, the model was refined using the least-squares fitting. If the relative error between the found model $(\tilde{X}, \tilde{Y}, \tilde{\theta})$ and the true one (X, Y, θ) was less than 4%, i.e.

$$\sqrt{\left(\frac{\tilde{X}-X}{W}\right)^2 + \left(\frac{\tilde{Y}-Y}{H}\right)^2 + \left(\frac{\left|\theta - \tilde{\theta} - \pi\right| - \pi}{2\pi}\right)^2} \le 0.04,$$

then the match was considered successful. Here W and H are image width and height respectively. The results for varying overlap percentage are shown in Figure 4.

Due to the least-squares refinement of the final result, the choice of bin size and smoothing parameter for the proposed method does not affect the result in a wide range of values. Thus we used 5% of parameter range ($[0, 2\pi]$ for angle and $[-W, W] \times [-H, H]$ for shift) for the Gaussian filter width, while the bin size was equal to 1 degree for angle and to 1% of image dimensions in pixels for shift.

As graphs indicate, the PROSAC method performs quite well in both cases if overlap area is not less than 15% of an image size. But with the decrease of overlap percentage the performance deteriorates quickly, especially in case of stronger noise. In fact, this does not necessarily mean that PROSAC is unable to find inliers. More often it means that the found model differs too much from the real one. PROSAC tends to select a small subset of inliers, while the proposed method either does not find them at all (quite rarely) or finds almost all of them. This allows our method to tolerate greater noise and to achieve much better precision. Also, as it was said before, it does not require the noise threshold to be set at all.
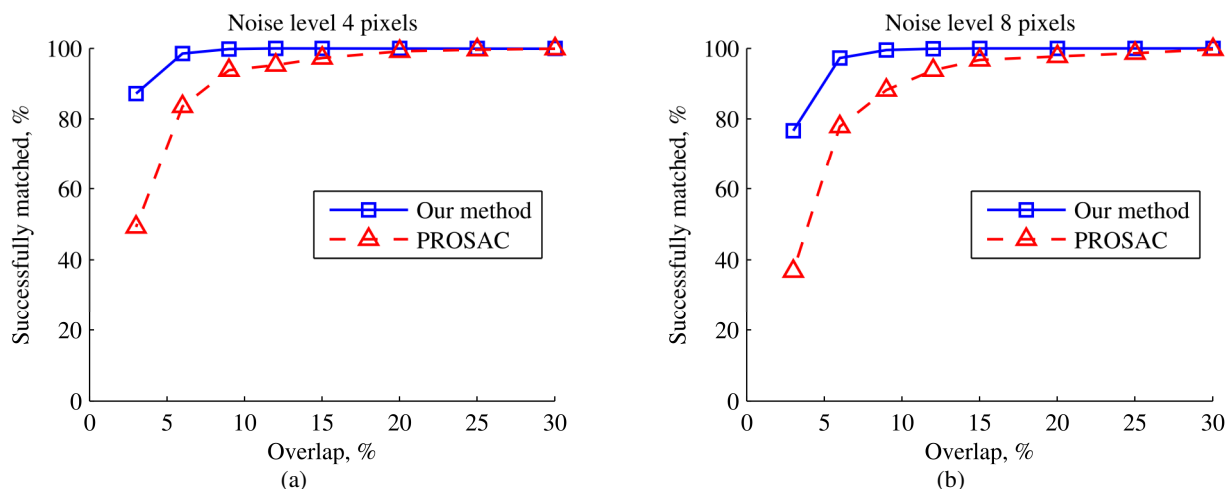

Figure 4. The comparative results of our algorithm and PROSAC on synthetic data. Note that the proposed method is more reliable in case of low overlap (less than 15%)

5.2 Experiments on real data

In order to compare our algorithm with PROSAC on real data we have collected a test set that consists of 28 pairs of overlapping aerial images. Those images have been taken with different cameras under varying conditions and include both digital and analog photos. The set also provides a reasonable variety of areas presented in the images as they include forests, fields, towns, mountains and rivers. All the images have been downsampled to the size of ~0.5 mega pixels. The overlap percentage is less than 10% for all the pairs, while the rotation angle varies from 0 to 180 degrees.

To provide both our algorithm and PROSAC with putative matches we first detected 1000 Harris corners (Harris et al., 1988) per image which is enough to cover the whole image area uniformly. SIFT descriptor with radius $r = 12$ pixels was then applied to those feature points and, finally, a simple nearest neighbor matching procedure was carried out. The descriptors distance threshold was set to $\sqrt{80000}$, as recommended in (Läbe et al., 2006). Matches for PROSAC were sorted by the nearest neighbor distance ratio.

(a)

(b)

Figure 5. Results of the proposed method on real data. Overlap area is 3.5% in (a) and 10% in (b)

Like in synthetic tests, our algorithm showed an advantage over the PROSAC-based scheme when applied to real data. Being absolutely deterministic, as opposed to SAC-methods, the proposed algorithm demonstrated better robustness and successfully matched 71% of the challenging test pairs, while PROSAC coped with 54%, often having to run through thousands of iterations to find a solution. The results are shown in Figure 5.

In our experiments we also tried an alternative approach to rotation angle detection. As we use SIFT descriptor, the dominant direction of each feature is known. Thus, it is possible to vote for the difference in directions by each match pair to determine the relative orientation of the images. But this approach turned out to be less robust than the one described in Section 4.1, due to instability of feature direction detection by SIFT and much lower total amount of votes. Moreover, this approach depends on a specific descriptor, while the proposed method is able to work with pure point correspondences.

5.3 Matching a set of images

The algorithm of shift-rotation model estimation has been used in a framework that matches a whole set of aerial images. The amount of images in a set is ~100-200. A set is usually divided into a number of routes. For each image the route to which it belongs is known beforehand and no other additional information is used. Note that, while the overlap among images belonging to the same route is substantial, typical overlap between neighbouring routes is on the order of 20-30% or even less in the most complex cases. Moreover, in general case the routes may be placed randomly with respect to each other. Thus, if no GPS / IMU data is available, pairs of overlapping images must be extracted automatically. We used the approach proposed in (Brown et al., 2007) for candidates' selection and then applied our method for image matching. The overlapping images were successfully detected in most cases which made an automatic assembly of image blocks possible.

6. CONCLUSIONS

We have presented a new method of shift-rotation model estimation that is based on a voting procedure in parameter space. The proposed method improves the stage of model estimation within the bounds of the discussed two-level scheme of matching aerial images. This allows a more robust and qualitative detection of overlap area which results in better matching of aerial images with low overlap. Our method has been shown to outperform the existing approaches, particularly the modern PROSAC algorithm. The results of the synthetic tests have been proved by experiments on real data that consisted of an exhaustive set of aerial images with small overlap. The advantages of the proposed method are determinacy and a straightforward set of parameters that do not require tuning.

REFERENCES

Begg, C., Mukundan, R., 2004. Hierarchical Matching Techniques for Automatic Image Mosaicing. In: *Proceedings of the Conference on Image and Vision Computing*, Akaroa, New Zealand, pp. 381-386

Brown, M., Lowe, D. G., 2007. Automatic Panoramic Image Stitching using Invariant Features. In: *International Journal of Computer Vision*, Vol. 74, Issue 1, pp. 59-73

Choi, S., Kim, T., Yu, W., 2009. Performance Evaluation of RANSAC Family. In: *The British Machine Vision Conference*, London, UK (on CD-ROM)

Chum, O., Matas, J., 2005. Matching with PROSAC - Progressive Sample Consensus. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, pp. 220-226

Fischler, M., Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: *Transactions of ACM - Graphics and Image Processing,* Vol. 24, pp. 381–395

Flusser, J., Suk, T., 1994. A moment-based approach to registration of images with affine geometric distortion. In: *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, pp. 382–387

Grün, A., 1985. Adaptive least squares correlation: a powerful image matching technique. In: *South African Journal of Photogrammetry, RS and Cartography*, Vol. 14 (3), pp. 175-187

Harris, C., Stephens M., 1988. A combined corner and edge detector. In: *Proceedings of The Fourth Alvey Vision Conference*, Manchester, UK, pp. 147–151

Hough, P., 1962. A method and means for recognizing complex patterns. U.S. Patent #3069654

Kaneko, S., Satoh, Y., Igarashi, S., 2003. Using selective correlation coefficient for robust image registration. In: *Pattern Recognition*, Vol. 36, pp. 1165–1173

Läbe, T., Förstner, W., 2006. Automatic Relative Orientation of Images. In: *Proceedings of the 5th Turkish-German Joint Geodetic Days*, Berlin, Germany (on CD-ROM)

Li, J., Allinson, N., 2008. A comprehensive review of current local features for computer vision. In: *Neurocomputing archive,* Vol. 71 , Issue 10-12, pp. 1771-1787

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints, In: *International Journal of Computer Vision*, Vol. 60 (2), pp. 91-110

Matas, J., Obdrzálek, S., Chum, O., 2002. Local affine frames for widebaseline stereo. In: *16th International Conference on Pattern Recognition*, Quebec, Canada, Vol. 4, pp. 363–366

Pires, B., Aguiar, P., 2004. Registration of images with small overlap. In: *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, pp. 255- 258

Reddy, S. B., Chatterji, B. N., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. In: *IEEE Transactions on Image Processing*, Vol. 5, Issue 8, pp. 1266-1271

Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Bonn, Germany, Vol. 36, Part 3, pp. 49-54

Seedahmed, G., Martucci, L., 2002. Automated Image Registration Using Geometrically Invariant Parameter Space Clustering. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Graz, Austria, Vol. 34, Part 3A, pp. 19–23

Seo, J., Jeong, S., Kim, K., 2003. Hierarchical corner matching for automatic relative orientation. In: *Proceedings of Geoscience and Remote Sensing Symposium*, Toulouse, France, Vol. 6, pp. 3958-3960

Tuytelaars, T., Mikolajczyk, K., 2008. Local Invariant Feature Detectors: A Survey. In: *Foundations and Trends in Computer Graphics and Vision*, Vol. 3, pp. 177-280

Xiong, Y., Quek, F., 2006. Automatic Aerial Image Registration Without Correspondence. In: *Proceedings of the 4th IEEE International Conference on Computer Vision Systems*, New York, USA, pp. 25-31

Zheng, Q., Chellappa, R., 1993. A computational vision approach to image registration. In: *IEEE Transactions on Image Processing*, Vol. 2 (3), pp. 311–326

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. In: *Image and Vision Computing*, Vol. 21, Issue 11, pp. 977-1000

7. ACKNOWLEDGEMENTS