Matching of Aerospace Photographs with the use of Local Features

R Malashin[1]
[1]Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, 49 Kronverkskiy pr., Saint Petersburg 197101, Russia

Email: malashinroman@mail.ru

Abstract. Matching of digital images is very challenging computer vision problem. The aim of investigation was developing of algorithms for matching real aerial and cosmic photographs. In the proposed methods, images are described locally by scale and rotation invariant descriptors. Reliability and high accuracy of the algorithms has been achieved by combining dense keypoint detector and robust descriptor with complex procedure of outlier elimination. The algorithms are capable of making correct decisions when number of local mismatches is more than 99%.

1. Introduction

The main difficulty in matching of aerospace images is that appearance of the Earth surface changes dramatically as the result of season, day-time, or viewpoint changes. In general case, object specific information also cannot be used due to wide diversity of possible content of these images. The target independent structural analysis is the only way for solving the task of matching of aerospace images [1], and there are two possible approaches for accomplishing it.

The first approach is to construct a very robust description of image that will capture only or almost only the invariant image features. The examples of successful implementation of this approach in the task of registration of aerospace images can be found in [2,3]. These methods are based on contour structural elements such as segments of straight lines, blobs and corners. These elements are not unique enough. That's why, plane structure of the Earth surface is assumed, and complex mutual relations of elements are necessarily taken into account. These methods cannot deal with non-planar surfaces.

The approach applied in this paper is based on matching of local features of images. Descriptors invariant to scale and rotation (like those applied in SIFT [4] and SURF [5]) are constructed in the regions of interest of images and then are matched to each other independently. Intuitively, our approach seems to be worthwhile of usage because some regions of aerospace images are almost insensitive to season and day-time changes. The local features extracted from such stable regions can be matched correctly, but another hard problem should be overcome. The great number of outliers arises from unstable regions and non-overlapped regions of images. Therefore the overall quality of matching depends significantly on the applied algorithms of eliminating outliers.

2. Proposed methods

Below there is a brief description of the proposed algorithms for feature detection, description and matching.

2.1. Detection and description of local features

Recently, there have been developed many different algorithms of keypoint detection and description. The experiments carried out with cosmic and aerial pictures have shown that the best results can be achieved using the SURF detector and the SIFT descriptor together.

The SIFT detector (as compared to the SURF detector) produces too many small keypoints that are less discriminative. The MSER [6] algorithm detects too few keypoints and can't provide needed "density" of description. The ORB [7] detector (and some other detectors) can't provide uniform distribution of keypoints in scale space.

The SIFT descriptor appeared to be more robust than other algorithms of local description. It is able to tolerate significant appearance changes of objects. The SURF descriptor captures too much object specific information and may be used when the alterability of territory appearance is not excessive. Binary descriptors (BRISK [8] and others) appeared to be inapplicable to aerospace images because they operate directly with pixel brightness (instead of intensity gradients) that is very sensitive to day-time and season changes.

2.2. Keypoint matching

Despite the appropriate selection of keypoint detector and descriptor (as it was discussed above), most of keypoint descriptors still are matched incorrectly. The outliers can be eliminated by means of a structural analysis that is described below.

2.2.1. Nearest Neighbor ratio versus 1-to-n matches exclusion. One of the most popular technique to eliminate outliers is use of nearest neighbor ratio as signal of mismatch [4]. Several experiments with manually labeled data were made. They showed that significant decrease of the number of inliers would result from thresholding ratio of distances to two closest neighbors in feature space when the images being matched are rather "difficult" for comparison. Figure 1 shows that most of inliers of "hard" image pairs refer to high ratios of distances to closest neighbors. It appeared that better results can be achieved by eliminating the ambiguity of matches that link several keypoints from the first image with one point of the second image. A single unambiguous match corresponding to the shortest distance in multidimensional feature space remains after such elimination. The n-to-1 matches have higher probability to be incorrect, and moreover, they have higher probability to form incorrect (false) clusters of matches with relevant parameters (Figure 2). The algorithms of cluster analysis applied for elimination of wrong matches are described below.

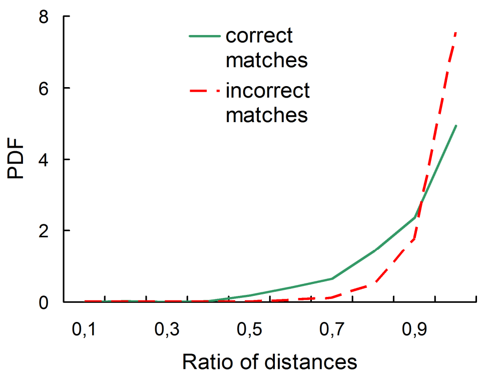

Figure 1. Probability density function of correct and incorrect matches. Generated from manually labelled data.

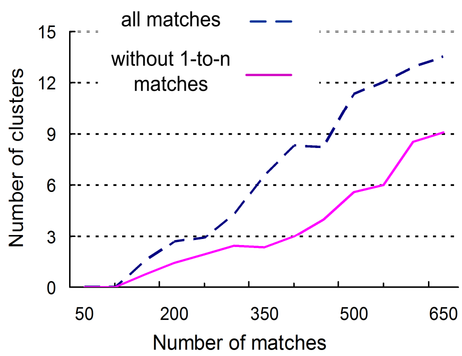

Figure 2. Average number of false clusters of matches (with 3 and more votes) before and after exclusion of 1-to-n matches. The graph was generated by matching hundreds of images having non-overlapping content.

2.2.2. Geometric constraints. Many mismatches can be eliminated by fitting a planar model to data generated by matching of descriptors. As soon as the Earth surface is not planar in general case, significant deviations from planar model should be accepted. Similarity transform model appeared to be more robust (with respect to outliers and "non-planar data") than more general affine transform model. Preliminarily, rough parameters of similarity transform are estimated by Hough Clustering of parameters of local co-ordinate transforms that are described by individual matches [4]. The bin size of Hough accumulator smaller than in [4] is used because the nearly planar scenes are considered. By the same reason, only the biggest cluster in Hough accumulator is further analyzed. The RANSAC algorithm is then used to estimate the parameters of similarity transform precisely and eliminate some more outliers from cluster. The mutual position, scale and orientation of matched keypoints are taken into account when fitting the model. Finally, the parameters of mutual affine transform of co-ordinate systems of two matched images are estimated by least square solution [4]. If the number of matches that passed all listed above tests falls below predefined threshold (with respect to initial number of matches) two photographs are considered to differ too strong from each other.

3. Experimental results

The proposed algorithms where successfully tested based on a series of aerial and cosmic photographs with strong appearance changes caused by season, day-time and viewpoint variation. The anthropogenous (manmade) changes were also considered by matching old photographs of city landscapes with modern ones. And moreover, SAR images were successfully matched with ones formed by optical sensors. The reliability and accuracy of the algorithms are achieved by eliminating of huge number of outliers by the use of geometrical constraints. The correct solutions can be made even based on small correct subsamples of the whole initial multitude of point-to-point matches (in some cases based on less than 0.1% of the whole multitude of matches).

Acknowledgements

This work was supported by the Ministry of Education and Science of the Russian Federation, and by the Government of the Russian Federation, Grant 074-U01.

References

[1] Lutsiv V, Malyshev I, Pepelka V and Potapov A 2002 *Proc. of SPIE* **4741** 351–62
[2] Lutsiv V., Malyshev I and Potapov A 2003 *Proc. of SPIE* **5238** 164–75
[3] Lutsiv V, Malyshev I and Pepelka V 2001. *Proc. of SPIE* **4380** 174–83
[4] Lowe D 2004 *Int. J. of Computer Vision* **60** 91-110
[5] Bay H, Tuytelaars T and Van Gool L 2006 SURF: Speeded Up Robust Features *Proc. 9th European Conf. on Computer Vision (Graz, Austria, May 7-13)* pp 404–417
[6] Matas J, Chum O, Urban M and Pajdla T 2002 Robust wide baseline stereo from maximally stable extremal regions *Proc. of British Machine Vision Conf (Cardiff, UK, September 2-5)* pp 384–396
[7] Rublee E, Rabaud V, Konolige K and Bradski G 2011 ORB: An efficient alternative to SIFT or SURF *Proc. of Int. Conf. on Computer Vision (Barcelona, Spain, November 8-11)* pp 2564–2571
[8] Leutenegger S, Chli M and Siegwart R 2011 BRISK: Binary Robust Invariant Scalable Keypoints *Proc. of Int. Conf. on Computer Vision (Barcelona, Spain, November 8-11)* pp 2548–2555