A UNIFIED FRAMEWORK FOR THE AUTOMATIC MATCHING OF POINTS AND LINES IN MULTIPLE ORIENTED IMAGES

Christian Beder

Institute of Photogrammetry, University of Bonn
Nussallee 15, 53115 Bonn, Germany
beder@ipb.uni-bonn.de

KEY WORDS: Matching, Reconstruction, Statistics, Geometry, Algorithms

ABSTRACT

The accurate reconstruction of the three-dimensional structure from multiple images is still a challenging problem, so that most current approaches are based on semi-automatic procedures. Therefore the introduction of accurate and reliable automation for this classical problem is one of the key goals of photogrammetric research.
This work deals with the problem of matching points and lines across multiple views, in order to gain a highly accurate reconstruction of the depicted object in three-dimensional space. In order to achieve this goal, a novel framework is introduced, that draws a sharp boundary between feature extraction, feature matching based on geometric constraints and feature matching based on radiometric constraints. The isolation of this three parts allows direct control and therefore better understanding of the different kinds of influences on the results.
Most image feature matching approaches heavily depend on the radiometric properties of the features and only incorporate geometry information to improve performance and stability. The extracted radiometric descriptors of the features often assume a local planar or smooth object, which is by definition neither present at object corners nor edges. Therefore it would be desirable to use only descriptors that are rigorously founded for the given object model. Unfortunately the task of feature matching based on radiometric properties becomes extremely difficult for this much weaker descriptors.
Hence a key feature of the presented framework is the consistent and rigorous use of statistical properties of the extracted geometric entities in the matching process, allowing a unified algorithm for matching points and lines in multiple views using solely the geometric properties of the extracted features. The results are stabilized by the use of many images to compensate for the lack of radiometric information. Radiometric descriptors may be consistently included into the framework for stabilization as well.
Results from the application of the presented framework to the task of fully automatic reconstruction of points and lines from multiple images are shown.

1 INTRODUCTION

This work deals with the matching of points and lines across multiple images. This problem has been addressed extensively for points (c.f. (Schmid and Mohr, 1997) or the classical textbooks (Horn, 1986) and (Faugeras, 1993)) and also for lines in (Schmid and Zisserman, 1997), (Baillard et al., 1999) and (Heuel and Förstner, 2001). All of those approaches generate matching pairs of features from the image intensity data and then use the known geometric information for forward intersection in order to obtain a 3D reconstruction of the depicted object. If you consider a situation like in figure 1, where a simple cube is depicted from all its six sides, it is obvious, that all those feature matching methods relying on radiometric information in the first place must fail, since no face of the object is visible in more than one view. On the other hand, all line features are visible in two views and all point features are visible in three views, thus a precise 3D reconstruction should be possible given the matches. The problem is of course the assumption, that the observed surface is local planar at the features, which is by definition neither the case at object corners nor at object edges, that are of primary interest for an accurate scene reconstruction. Using the known image geometry has been applied by (Jung and Paparoditis, 2003) for edgel matching across multiple views.

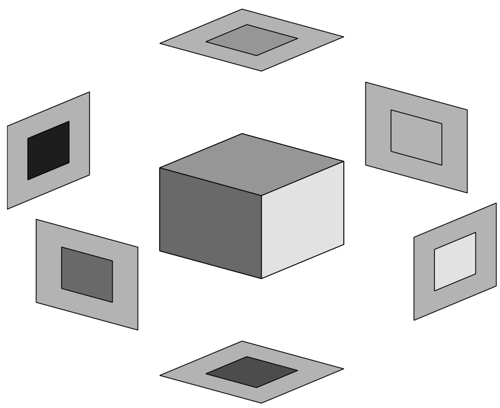

Figure 1: A cube depicted from all its six sides. No pair of pictures shows the same face, though a reconstruction is possible given the orientation of the cameras.

In the following a framework is presented, that allows to do matching of points and lines across multiple oriented images in a unified manner, by proposing the use of spatial filters that do not operate in the image domain, but use the known orientation to operate in the scene domain at the earliest possible stage. In doing so, not only unification is achieved, but also the effects of using radiometric image information are isolated, allowing more control over the use of intensity information in the matching process. Even the possibility of not using intensity information at all is given, allowing the fully automatic reconstruction based on geometric information alone as it is required in the scene depicted in figure 1.

In order to exploit the full geometric knowledge provided by the feature extraction, the statistical properties of the extracted features are used throughout the whole matching process, enabling the construction and operation on graphs, that represent the statistical relations between the objects.

2 FEATURE EXTRACTION

A prerequisite for feature matching is the extraction. The task of feature extraction from single images is well understood and many approaches are available (c.f. (Förstner, 1994), (Smith and Brady, 1997), (C.G. Harris, 1988), (Canny, 1986)). Even the statistical properties of the extracted features, i.e. points and lines, are obtainable as presented in (Förstner, 1994). If you know the exterior and interior orientation of the camera used, the uncertain projecting ray for every point and the uncertain projecting plane for every line segment can be computed according to (Heuel and Förstner, 2001). If you also know a lower and an upper bound on the distance of the depicted object from the camera, which is very simple accessible in many applications including aerial imagery, the locus of an image point x in space is a space line segment s together with its uncertainty Σ_{ss} and the locus of an image line segment l in space is a space quad q together with its uncertainty Σ_{qq}. Thus feature extraction in oriented images yields not only a set of image features together with a reference to the generating image

$$I_{FE} = \{(x_i, I_i)|i = 1..N\} \cup \{(l_j, I_j)|j = 1..M\}$$

but also a set of space objects together with their uncertainties

$$S_{FE} = \{(l_i, \Sigma_{l_i l_i}|i = 1..N\} \cup \{(q_j, \Sigma_{q_j q_j})|j = 1..M\}$$

Note, that there is a one-to-one mapping

$$s_{FE} : S_{FE} \rightarrow I_{FE}$$

between the two sets, associating each space object with its generating image object.

3 SPATIAL FILTERING

In this framework all processing is done by filtering objects in the spatial domain. This means, that starting from a set of space objects obtained from the set of images as described in the feature extraction section above, different filters are applied yielding increasingly complex space objects. More precisely a spatial filter is an algorithm

$$f : 2^S \rightarrow 2^S$$

that takes a number of space objects as input and generates some different space objects as output. Again a mapping

$$s : S \rightarrow 2^S$$

can be provided, that associates every space object with the source space objects, that were used in the filter to generate it. Therefore every application of a spatial filter generates one more level in a source tree of the space objects. Two filters are proposed to yield the matches of points and lines over multiple views.

3.1 Pairwise Grouping

The first step is a pairwise matching of the objects. In order to do this, a graph

$$G_{PG} = (S_{FE}, E_{PG})$$

induced by the statistical incidence relation (c.f. (Förstner and Heuel, 2000)) is constructed. The vertices of that graph are the space objects and an edge is inserted between two vertices p and q, if and only if there is no reason to reject the statistical hypothesis, that the space objects p and q intersect each other. The edge set is thus denoted by

$$E_{PG} = \{(p, q)|p, q \in S_{FE} \wedge intersect(p, q)\}$$

Every edge in this graph represents a possible match between two image objects, that is not contradictory to the scene geometry. If the image intensity information is to be included in the algorithm, an intensity based distance measure

$$d : I_{FE} \times I_{FE} \rightarrow R$$

must be introduced and those graph edges have to be pruned, that do not comply with the distance measure, i.e. the edge set is adjusted using an intensity distance threshold T as follows

$$E'_{PG} = \{(p, q) \in E_{PG}|d(s_{FE}(p), s_{FE}(q)) < T\}$$

Most matching techniques, including the classical correlation based and least squares approaches, are focused on the development of powerful and robust intensity distance measures (c.f. (Schmid and Mohr, 1997) and (Schmid and Zisserman, 1997)).

As pointed out in the introduction, there are certain conditions, that do not allow any pruning at this stage. Since no possible matches should be lost at this early stage of processing, the full edge set is used here and no pruning is performed. The resulting filtered set is thus obtained by taking every edge of G_{PG} and constructing the intersecting object from its end-vertices space objects. Thus the filter returns the set

$$S_{PG} = \{(c(p, q), \Sigma_{c(p,q)c(p,q)})|(p, q) \in E_{PG}\}$$

where $c(p, q)$ denotes the optimally estimated geometric objects obtained from the intersection of p and q as described in (Heuel, 2001). The source tree is constructed simply as

$$s_{PG}(c(p, q)) = \{p, q\}$$

3.2 Minimum Clique Partition

The second step of the multi-view matching process is the aggregation of the pairwise matches into matches over all images. In order to achieve this, again a graph

$$G_{MC} = (S_{PG}, E_{MC})$$

induced by the statistical equality relation (c.f. (Förstner and Heuel, 2000)) is constructed. The vertices of that graph are again the space objects but this time the edges are inserted between two vertices p and q, if and only if there is no reason to reject the statistical hypothesis, that the space objects p and q are equal. The edge set is denoted by

$$E_{MC} = \{(p, q)|p, q \in S_{PG} \wedge equal(p, q)\}$$

It follows, that a clique in G_{MC} is a set of objects, that share a common position in space. Hence, in order to find the multi-view matches that range over the largest amount of images, i.e. are supported by the largest amount of observations, a minimum clique partition is computed for the graph. To be more precise, a set of sets

$$C = \left\{ C_1, ..., C_N | \begin{array}{l} \bigcap_i C_i = S_{PG} \wedge \\ p, q \in C_i \Rightarrow (p, q) \in E_{MC} \end{array} \right\}$$

is computed, such that the number of matched objects is maximized, i.e. the number of cliques N is minimized. Although the solution might not be unique and this problem is known to be NP-hard (c.f. (Garey, 1979)), it is likely, that the graph G_{MC} consists of many small connected components due to the fact, that reconstructed points and and lines fill only a small portion of space, enabling an efficient computation. The strategy used here to approximate the solution is a classical greedy approach, as can be found for example in (Ausiello et al., 1999). The analysis of the effects of different approximation algorithms on the results is beyond the scope of this paper. Too small cliques, i.e. reconstructions that are supported by to few observations, may be discarded at this stage. Denoting the optimal clique partition with $\tilde{C}$, one may continue with a reduced set

$$\tilde{C}' = \{C_i \in \tilde{C}| \ |C_i| \geq M\}$$

for some minimal supporting observation number M. The resulting set is now given by

$$S_{MC} = \{(c(C_i), \Sigma_{c(C_i)c(C_i)})|C_i \in \tilde{C}\}$$

where again $c(C_i)$ denotes the optimally estimated geometric objects obtained from the objects in C_i as described in (Heuel, 2001). The source tree is analogously to the previous chapter set to

$$s_{MC}(c(C_i)) = C_i$$

3.3 The Overall Procedure

The overall procedure for reconstructing points and lines from multiple oriented views is as follows:

1. The features, i.e. points and lines, together with their uncertainties are extracted using a standard feature extraction algorithm.

2. The locus of the space objects that generated the features, i.e. space line segments for image points and space quads for image lines, are computed together with their uncertainties using the known camera orientation and lower and upper bounds on the object distance.

3. The 3D-objects are filtered pairwise according to the incidence relation yielding new 3D-objects, i.e. space points for image points and space line segments for image lines, that are each observed in two images. Image intensity information may be added at this stage to improve performance, but is not necessary.

4. The 3D-objects are filtered into aggregates according to the equality relation yielding new 3D-objects, i.e. space points and space line segments, that are supported by the maximum number of observations across all views. Reconstructions, that are supported by too few observations may be discarded at this stage.

If one is interested in the matched features itself instead of the 3D reconstruction, the leaves of the source tree have to be computed. To be more precise, for every object o of the reconstruction the generating image features can be retrieved from the source tree as follows:

$$\{s_{FE}(p)|p \in s_{PG}(q), q \in s_{MC}(o)\}$$

4 EXPERIMENTAL RESULTS

In order to evaluate the performance of the introduced procedure on real data a set of seven overlapping aerial images was used. The images were scanned with a ground resolution of 14.3cm and the orientation was established manually. A single building was cut out and the point- and line-features were extracted together with their uncertainties using the approach presented in (Förstner, 1994).

The features were projected into the spatial domain using the known orientation data yielding space line segments and space quads as depicted in figure 2 for a single image.

The next step was to intersect each such line segment with each other and each space quad with each other. The statistical properties of the extracted features were used to test, whether two space objects intersect and also in the construction of the resulting space object. Note that no information from the images is used in this step of processing and the matches are established based only on the known

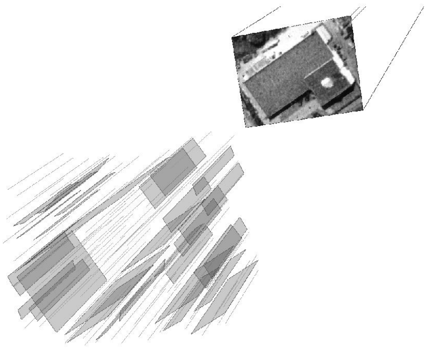

Figure 2: The extracted features of a single image with known orientation in the spatial domain. Each extracted point yields a space line segment and each extracted line yields a space quad.

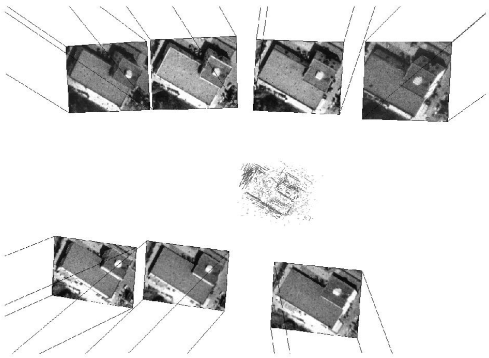

Figure 3: All pairwise epipolar compatible line segments and points between the seven images.

geometric properties. The resulting set of epipolar compatible line segments and points together with their uncertainties are shown in figure 3. As expected the geometric information from only two views is to weak to make an accurate reconstruction with few outliers, thus many outliers have to be removed in the subsequent phase.

The next step was the aggregation of spatial objects into cliques as described above. Again no information from the images other than the geometry is used. Retaining only those cliques with a minimal size of three yields the points and line segments depicted in figure 4. Note that nearly all outliers were removed by this step as expected, because almost only real object corners and edges have produced features in multiple views that generated cliques of sufficient size.

Finally two leaves of the source tree are shown in figures 5 for a point and 6 for a line. Observe, that line chaining takes place in the aggregation process, as can be seen in figure 6, if there is one connecting observation.

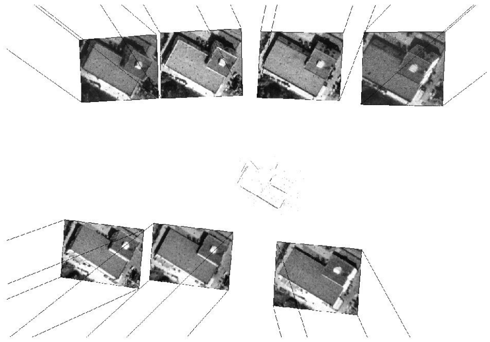

Figure 4: Cliques of line segments and points of size at least three.

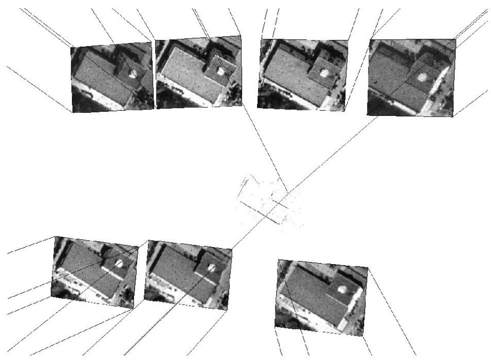

Figure 5: A matching point.

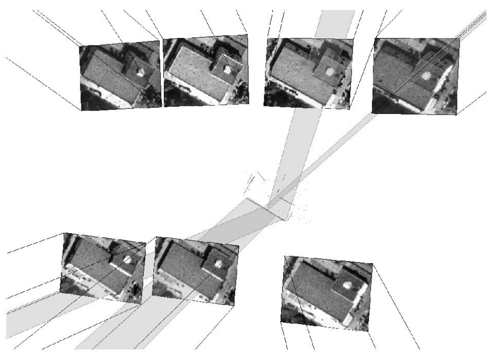

Figure 6: A matching line.

5 CONCLUSION

A unified framework for the fully automatic reconstruction of points and lines from multiple oriented images was presented. Unification was achieved by operating in the spatial domain from the earliest possible stage of processing using statistical geometric properties of the extracted features. The framework uses graphs induced by relational geometric properties, that can be handled in a rigorous statistical manner. It was demonstrated using this framework, that geometric information from the images is sufficient to establish matches of points and line segments over multiple images, thus enabling an accurate scene reconstruction without radiometric information from the images in the matching process at all. This indicates, that the known orientation yields much more information than is used by most other matching methods, that focus on the radiometric properties of the images and use the geometry only to improve robustness and performance. All matching algorithms, that are based on pairwise radiometric distance measures can easily be integrated into the presented framework. As a consequence an improvement of existing feature matching algorithms can be expected due to the extensive and statistically rigorous use of the existing geometric information and the possibility to integrate weak radiometric descriptors into the task of feature matching.

REFERENCES

Ausiello, G., Protasi, M., Marchetti-Spaccamela, A., Gambosi, G., Crescenzi, P. and Kann, V., 1999. Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties. Springer-Verlag New York, Inc.

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic line matching and 3d reconstruction of buildings from multiple views. In: Proc. ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery.

Canny, J. F., 1986. A computational approach to edge detection. IEEE Transaction on Pattern Analysis and Machine Intelligence 8(6), pp. 679–698.

C.G. Harris, M. S., 1988. A combined corner and edge detector. In: Fourth Alvey Vision Conference, pp. 147–151.

Faugeras, O., 1993. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press.

Förstner, W., 1994. A framework for low level feature extraction. In: European Conference on Computer Vision, pp. 383–394.

Förstner, W., B. A. and Heuel, S., 2000. Statistically testing uncertain geometric relations. In: Proc. DAGM 2000, Kiel, Germany, pp. 17–26.

Garey, M.R., J. D. S., 1979. Computers and Intractability - A Guide to the Theory of NP-Completeness. Freeman.

Heuel, S., 2001. Points, lines and planes and their optimal estimation. In: Pattern Recognition, 23rd DAGM Symposium, LNCS, Springer, pp. 92–99.

Heuel, S. and Förstner, W., 2001. Matching, reconstructing and grouping 3d lines from multiple views using uncertain projective geometry. In: CVPR ’01, IEEE.

Horn, B., 1986. Robot Vision. MIT Press.

Jung, F. and Paparoditis, N., 2003. Extracting 3d free-form surface boundaries of man-made objects from multiple calibrated images: A robust, accurate and high resolving power edgel matching and chaining approach. In: Proc. of the ISPRS Conf. Photogrammetric Image Analysis, pp. 39–44.

Schmid, C. and Mohr, R., 1997. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(5), pp. 530–535.

Schmid, C. and Zisserman, A., 1997. Automatic line matching across views. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition.

Smith, S. M. and Brady, J. M., 1997. SUSAN – A new approach to low level image processing. Int. Journal of Computer Vision 23, pp. 45–78.