MULTI-IMAGE MATCHING USING SEGMENT FEATURES

Juliang SHAO[*], **Roger MOHR**[#] and **Clive FRASER**[*]

*Department of Geomatics, The University of Melbourne
Victoria 3052, Australia
Email: jsha@sli.unimelb.edu.au & c.fraser@eng.unimelb.edu.au

#Xerox Research Centre Europe, 6 chemin de Maupertuis
F- 38240 Meylan, France
Email: Roger.Mohr@xrce.xerox.com

Working Group III/1

KEY WORDS: feature-based matching, computer vision, multi-image matching, quality control, segment features.

ABSTRACT

This paper presents a strategy for matching features in multiple images, which emphasises reliable matching and the recovery of feature extraction errors. The process starts from initial 'good' matches, which are validated in multiple images using multi-image constraints. These initial matches are then filtered through a relaxation procedure and are subsequently used to locally predict additional features that might well be extracted using different thresholds for the feature extraction process. The relaxation labelling is simultaneously performed among multiple images, as opposed to the usual case of just two images. The overall process has been applied to segment matching of both aerial and close-range imagery and an example application is briefly reported.

1 INTRODUCTION

In parallel with advances in digital imaging, the photogrammetric and computer vision communities have witnessed significant developments in image matching over the past two decades (eg Gruen, 1985; Dhond & Aggarwal, 1989; Faugeras, 1993 and Jones, 1997). Nevertheless, image matching remains a bottleneck in both the automation of 3D information extraction from imagery and in the browsing of image databases (Jain and Vailaya, 1996). Within photogrammetry, image matching, and typically feature-based matching, is principally applied in the recovery of surface contour information from stereo pairs of aerial photographs, where the matching can be either on selected, sparsely distributed feature points or on a dense array comprising all available feature points. Feature-based matching can also be readily extended to multiple images (>2) and in this paper we consider the use of line segments as features to support multi-image matching.

The process of matching features in multiple images exhibits a number of fundamental characteristics. Of these, we focus in this paper upon reliable matching and the recovering of feature extraction errors. The operation starts from initial 'good' matches, which are validated within the multiple images using applicable constraints. These initial matches are then filtered through a relaxation procedure and used to locally predict further features which might be extracted using different thresholds for the feature extraction process. A relaxation labeling, which is an extension of the more typical two-image case, is then simultaneously performed among the multiple images. The different steps lead, for a local distribution of features, to an integration of signal (intensity) similarity and geometric constraints.

The proposed multi-step approach has been adopted in order to overcome both ambiguities in matching and errors in feature extraction. As regards the issue of errors, the approach will be illustrated through the use of features comprised of contour segments. In this case, errors in segment extraction are partially a function of both spurious segments and broken line segments.

The first step in the process is to collect an initial set of reliable features, using a strict threshold for the feature extraction algorithm, namely a strong gradient for the segments. These segment features are then matched using two types of information. The first is geometrical information (Gruen, 1985) within the multiple images, which might comprise epipolar lines for stereo images for instance. For the case of three images, however, the position of interest points is further constrained since a match in two images will have a single correspondence in the third. This

correspondence can be better determined when the trifocal tensor (Hartley, 1995; Shashua, 1994) is estimated or the exterior orientation of the camera station configuration is known, as is usually the case in photogrammetry. The second type of information is grey level intensity data. Grey level images are locally similar in corresponding patches and a typical measure of matching quality is the correlation coefficient value, the image distortion being locally approximated by an affine transformation (Lan and Mohr, 1997).

Such an approach proved early on in the present investigation to be successful for practical applications which utilised controlled close-range imagery. The extension to aerial imagery was initially problematic, however, since the segment matching process can inadvertently discard an excessive number of candidate segments. Missing correspondences can be caused by object occlusion, noise or simply poorly chosen feature extraction thresholds. Thus, at the initial step, it is unlikely that all valid matches will be fully obtained. With the proposed approach, however, considerable improvement is achieved in the number of correct multi-image matches for non-noisy segments.

The initial set of recovered matches is used as a starting point in a refined matching process. As the set may contain wrong matches, a relaxation process is applied to filter out unreliable correspondences. In multiple image matching, explicit optimisation strategies using binary compatibility constraints for a removal of ambiguous line segment correspondences are not usually employed. Instead, a more traditional approach considering geometry (eg epipolar line geometry) and intensity information is likely to prove sufficient for the control of matching quality. Here, relaxation labelling is applied at each stage as a post-filtering process.

Also at the initial stage, a sparse set of reliable matches is assumed to be provided. The next step is then an iterative process where at each iteration, already matched features offer 'evidence' for either additional matches or additional features that need to be searched for. No exact prediction of a correspondence can be made without knowing the 3D position of the feature. However, if we accept the concept of a limited disparity gradient, matched features can be employed to predict positions of unmatched features. This operation is performed using a local affine model where once the affine coefficients are determined, a position can be predicted. Given a certain error space, a search can then be carried out for a predicted feature. When a candidate is identified, a final verification step is performed using grey level information and linear segment orientation attributes. If no feature is found at this location, a refinement to the segment extraction is again performed, but with a lower threshold.

The proposed multi-image segment matching scheme has been experimentally evaluated using both aerial and close-range imagery. The paper discusses an experimental application and reports on the analyses of the results obtained.

2 COMPUTING THE INITIAL MATCHES

Consider the relationship between image points and epipolar constraints in a multi-image configuration. Two points in two images are admissible matches if the epipolar constraint is satisfied. For multiple images, this corresponds to a network of constraints, as illustrated in Fig. 1 for the case of three images. Such a network is constructed not only using the epipolar geometry, but also by employing signal similarity which is quantified through cross correlation and gradient direction similarity.

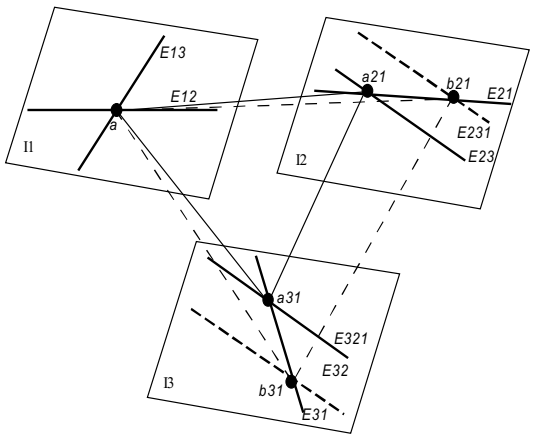
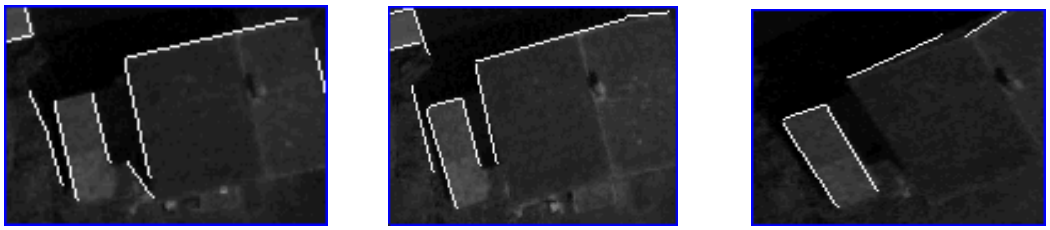


Fig. 1: *E12* and *E21*, *E13* and *E31*, and *E231* and *E321* are the corresponding epipolar lines. Two cliques of admissible matches are possible: *a*, *a21*, *a31* and *a*, *b21* and *b31*.

In order to achieve segment matching at the initial step, for the line segments in Image 1, all segments in the remaining images which satisfy the epipolar constraints need to be considered. Thus, for all possible matched segment pairs, predictions are made in the remaining images and the matching between individual points belonging to theses segments is then checked via both the correlation coefficient and the condition of an admissible similarity in gradient orientation. This process is not symmetrical since a single image (eg Image 1) is used to initialise the segment feature matches, after which the operation is repeated for all other images and image segments.

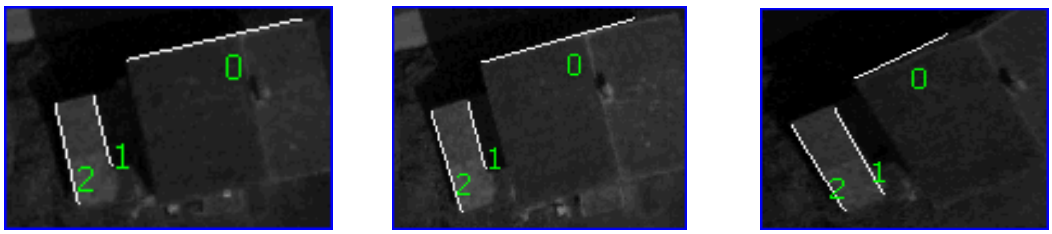
If Images 1 and 2, two points on matched contours *P1* and *P2* satisfy the epipolar constraints, they satisfy a consistency constraint. If *P1*, … , *Pn* form a clique (ie a complete subgraph) for the consistency graph in *n* images, this set can be considered as a set of complete, mutually supported matches. Fig 1 illustrates this for the case of three images, where two concurrent cliques compete: *a*, *a21*, *a31* and *a*, *b21* and *b31*. The subsequent relaxation process will handle this ambiguity.

Figs. 2 and 3 illustrate a case of three original images with segment features, along with the results of an initial segment matching whereby three corresponding segments are pre-matched.



a b c

Fig. 2: Original segments.



a b c

Fig. 3: Initial corresponding segments.

3 EXTENSION OF THE INITIAL MATCHES

It can be assumed at this first matching phase that the relaxation filtering has provided an initial reliable but sparse matching. The second stage now seeks to accommodate missing matches, which may be due to occlusions, image noise or simply poor thresholding in the feature extraction process. A complicating issue is the presence of matching ambiguities, which sometimes can be resolved through relaxation processes if additional features are provided. An effective strategy to counter ambiguities is therefore to extend the number (and density) of both features and initial matches. This extension provides a feedback process in which new matches immediately become supporting evidence for neighbouring feature correspondences. A local affine transformation is used to predict positions of feature candidates in the usual case where exact 3D information for the object to be reconstructed is not known. In the predicted regions, contours can be iteratively searched by changing the threshold of the feature extractor, which is here a contour extractor.

3.1 Line Prediction via a Local Affine Model

A local affine model is applied to the initially matched segments, with the three nearest segments being selected for the parameter determination of the affine transformation. As indicated in Fig. 4, segment pairs 1 - 1’; 2 - 2’; and 3 - 3’ are the one-to-one initial matches. The position of the candidate *L’* can be predicted from the segment *L* via a transformation utilising a line affine model defined in the dual affine space of lines by

$$\begin{bmatrix} A' \\ B' \\ C' \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} \quad (1)$$

where (A', B', C') and (A, B, C) are the coefficients of the corresponding lines in two images, while a, b, c, d, e, and f are the coefficients of the affine dual transform. To solve the six elements of the line affine model, at least three pairs of matched lines are required. It should be noted that to avoid singularity, the three lines in the reference image for the line affine element determination must exclude lines intersecting in a single point, and the case of three parallel lines with an effective intersection at infinity.

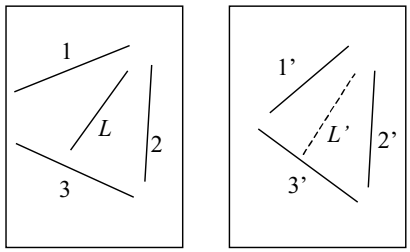


Fig. 4: Candidate prediction (L') in object image derived from L in reference image by the nearest segments: 1 - 1'; 2 - 2'; and 3 - 3' which are the initial and one to one corresponding matches.

3.2 Extension via Prediction

Once the affine coefficients are determined, a feature position can be predicted and a search can be conducted for that feature within specified error bounds. When a candidate is located, a final verification step is performed using grey level intensity values and information on contour orientation. If, on the other hand, no feature is found at this location, a refined segment detection can be iteratively performed through a lowering of threshold values.

Shown in Fig. 5 is the extension of the initial segment matching (Fig. 2) in which additional matched segments have been located. The lines whose numbers are larger than 2 indicate the additional matches obtained via a change of the threshold in the contour extractor for all images.

a b c

Fig. 5: Additional corresponding segments obtained via prediction.

Each segment may have several ambiguous candidates. To eliminate ambiguities, relationships among the neighbouring segments in the same image, and among the candidates in overlapping images, are formulated to enable local consistency checks. This aspect is discussed in the next section.

4 A RELAXATION PROCESS FOR LOCAL CONSISTENCY CHECKS

To eliminate ambiguities in feature-based matching, a relaxation process has been adopted. This technique has previously proved a useful tool in vision processes (Atalay and Yilmaz, 1998; Barnard and Thompson, 1980; Kittler, 1993). Probabilities are employed in the relaxation process. Firstly, the initial probabilities are estimated from similarity measures and they are then updated iteratively so as to impose local consistency conditions. That is, the probability of a point is increased if its compatible neighbours have high probability values. A recursive procedure can be continued either until the probabilities reach a steady state or a pre-set number of iterations has been carried out. This is thus

actually a balancing procedure based on the disparity continuity constraint in local regions (Barnard and Thompson, 1980), which has been been adopted in numerous projects in computer vision.

The constraints of segment matching in the case of two images are generally categorised into three: unitary, binary and 'N-ary' (Jones, 1997). Binary constraints are the most commonly used in relaxation; such a relation can express neighbour compatibility, and it has been used in this way here. Fig. 6 shows the compatibility of the two segments 1 and 2 with corresponding segments 1' and 2' having consistent disparities (ie d ≈ d').

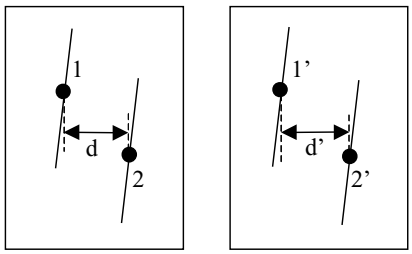


Fig. 6: Compatibility if d ≈ d'.

4.1 Initial Compatibility Measures

Similarity measures are fundamental processes in image correspondence determination. Depending on requirements, which include geometric, radiometric and scalar invariance conditions, similarity measures may be formed from one or more of the following: the correlation coefficient, the absolute value of grey level differences, moments and structural information. In this paper, the correlation coefficient, which has radiometric and partial scale invariance, is employed. The geometric correction is carried out by using epipolar geometry. In the cross correlation, the correlation coefficient is a normalised covariance function, the value of which ranges from -1 to +1. This makes the probability normalisation (0~1) much easier:

$$P = (C + 1)/2 \tag{2}$$

Here, P is the initial probability and C the correlation value, which is the average of two correleation coefficients derived from the two pairs of end points of the corresponding line segments. To pass the grey level constraint, the typical threshold requirement for the correlation coefficent is that the value is positive.

4.2 Expressing the Compatibility

The compatibility criterion in the adopted relaxation labelling is the local similarity disparity between matched features. However, the similarity disparity is usually based on the assumption that only limited rotation exists between the two images. This assumption can be validated via the rotation of a feature point according to epipolar geometry.

As described above, a disparity difference between two neighbouring points is usually employed for a compatibility check. Any imaging scale variation between the images, however, can complicate the comparison. In this paper, scale compensation has been considered. By assuming a scale change of not more than, say, 25%, the improved local disparity consistency is expressed by

$$| d' - (1 - s)d | < t \tag{3}$$

where d and d' are coordinate shifts of two pairs of corresponding points, as illustrated in Fig. 6; s is a disparity scale factor between the two images; and t is a threshold (typically 6 pixels) for the disparity difference. The coordinate shifts express the disparity for the two pairs; therefore d - d' is the disparity difference in the horizontal direction. The measure can apply to any direction. The two neighbouring points are therefore compatible if Eq. 3 is satisfied. Assuming a segment in the reference image and a candidate in the object image are pre-matched, the symbolic expression for the probability is P. As usual, the ambiguity is removed through local consistency checks in the neighbourhood. These might comprise *similarity disparities* via *rotation* and *scale compensation* as stated above. Locally mutual disparity consistency reinforces the probabilities from the compatible neighbouring points:

$$Q_1 = \sum_{l \in R} P \qquad \text{if Point } l \text{ satisfies the condition of local disparity consistency.} \tag{4}$$

The nearest compatible neighbours in a local region R are counted for the contributions to Q_1 for the current segment.

The above formula only illustrates the compatibilities among neighbours in the same image. For a clique constructed in the initial or extended matching, any change in the rejection and acceptation of one or more candidates in any iteration will have an impact on other nodes of the clique. The probabilistic expression for this impact is defined by

$$\begin{cases} Q_2 = (1+P_2)\times(1+P_3)\times\cdots\times(1+P_N), & N>2 \\ Q_2 = 0, & N=2 \end{cases} \tag{5}$$

where Q_2 stands for the contributions from other points in the clique; P_2, P_3, …, P_N are the previous probabilities of the candidates in the images 2, 3, …, N, provided that the current point is in Image 1. If the previous probability of that candidate is zero (ie that candidate was rejected), no contribution is made, whereas if it is 1 (accepted), then most contribution is assigned. When only two image points remain on the clique (ie. $N = 2$), no contribution will be provided. As a result, the current candidate will fail. By this means there is global consideration for all images, instead of just two images.

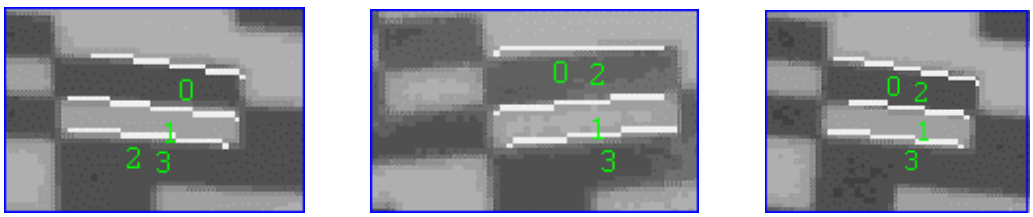
4.3 The Relaxation Process

The essence of relaxation labelling is that compatible neighbours reinforce each other, whereas incompatible neighbours decrease their mutual support, regardless of the similarity functions used. The following section details the relaxation functions utilised for the present investigation. With the two types of contributions from the same image and between images, an iterative computation is performed:

$$P^{k+1} = P^k \times (A + B \times Q_1 \times Q_2) \tag{6}$$

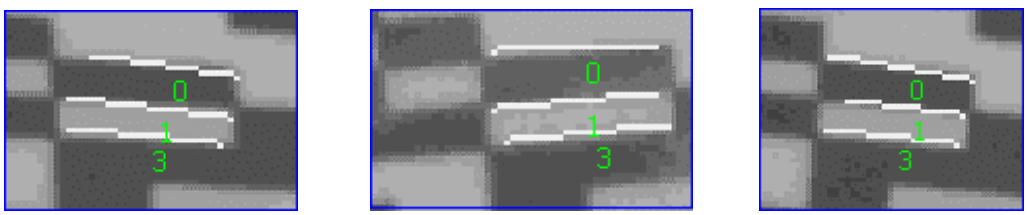
where A and B are coefficients that control convergence speed, and k and $k+1$ are the iteration numbers. A multi-image clique is accepted if all probabilities P_{ij} ($i, j \in N,\ i \neq j$ with N representing the number of overlapped images of the clique) related to the points on this clique reach a given threshold.

Fig. 7 shows the matching results using unrestricted epipolar geometric constraint prior to relaxation processing, while Fig. 8 displays the outcome of this process. Segment 2 in Fig. 7 are rejected through the relaxation process.



a b c

Fig. 7: Corresponding segments before the relaxation processing.



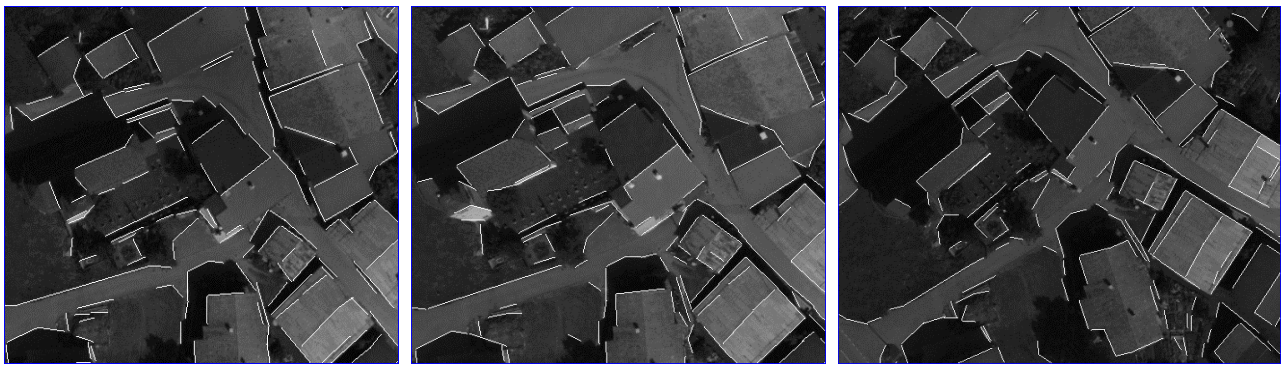
a b c

Fig. 8: Corresponding segments after the relaxation processing.

5 EXPERIMENTAL APPLICATION

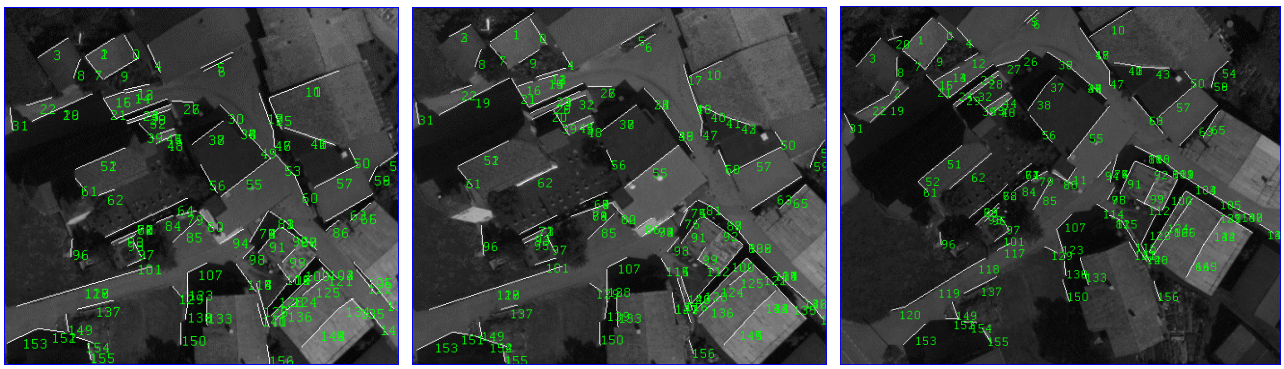
The segment matching process developed has to date been evaluated on four sets of images, including both aerial and close-range image configurations of three overlapping images. One of these experimental applications is reported here

to exemplify the multi-image matching via segment features. This comprised a three-image aerial coverage of part of a town in Europe, as shown with the initially extracted line segments in Fig. 9. At full resolution, there existed a discrepancy in image scale of close to 20% between the Image *a* and the other two involved. There was also a rotation of 25° between Image *c* and Images *a* and *b*. The segment matches displayed in Fig. 10 were derived in the initial matching approach applied to the three images. The matched segments identified at the conclusion of the relaxation process, but prior to extending the initial matches, are shown in Fig. 11. The final outcome obtained via extending the initial matches and applying the relaxation processing is displayed in Fig. 12. The final refinement led to a significant, 14.7% increase in the number of valid segment matches, whereas the rejection rate from the initial matching was 27.7%. This highlights the benefits to be gained by application of a relaxation processing through local consistency checks. The results also indicate that the affine model used was well suited to the prediction of matching candidates. It should be noted that for segment correspondence, one reference segment matching to multiple candidates is reasonable when multiple candidates coincide with the same line and each has a *common part* with the reference segment. The common part stands for the overlapped part restricted by epipolar geometry.



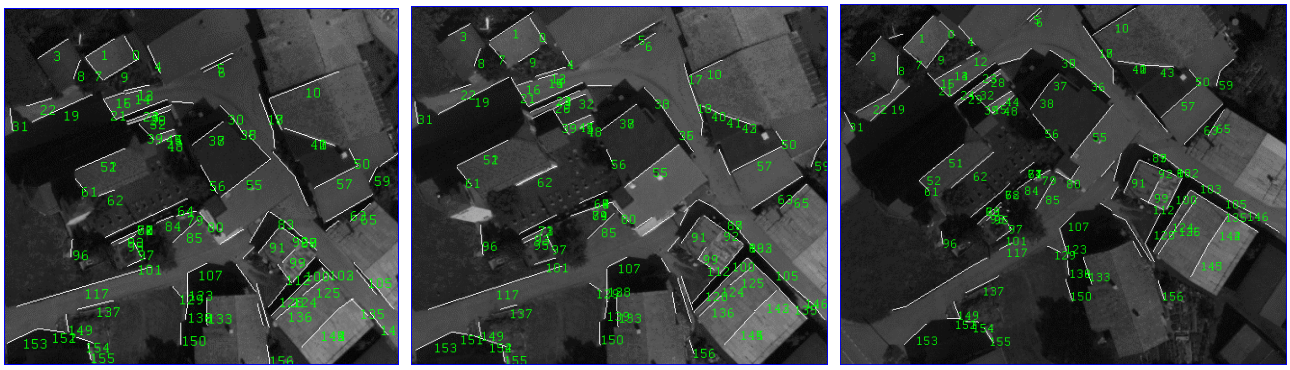
a b c

Fig. 9: Aerial image triplet showing extracted line segments.



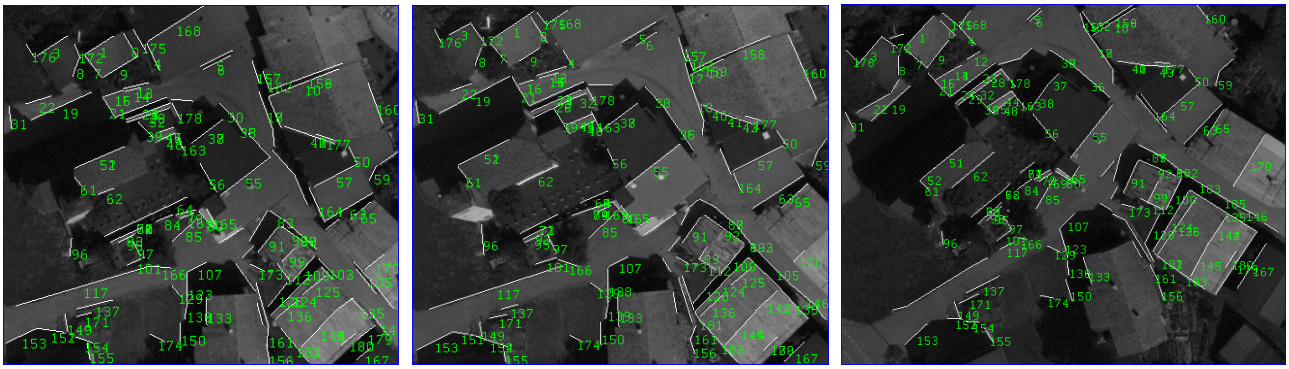
a b c

Fig. 10: Initial segment matches.



a b c

Fig. 11: The results of the relaxation processing to the non-extending matches.



a b c

Fig. 12: The final segments matched by extending the initial matches.

6 CONCLUSIONS

This paper has described an approach for multi-image, feature-based matching of segment features, which comprises three basic phases, namely computation of initial matches, extension of these matches through segment prediction, and finally utilisation of a relaxation process for consistency checking. The results of initial experiments, one of which is briefly touched upon, indicate that the merits of the proposed matching approach include reliable matching and successful recovery from the influence of feature extraction errors.

REFERENCES

Atalay, V. and Yilmaz, M., 1998. A matching algorithm based on linear features, Pattern Recognition Letters, 19: 857-867.

Barnard, S. and Thompson, W., 1980. Disparity analysis of images, IEEE Trans. on Pattern Recognition and Machine Intelligence, 2: 333-340.

Dhond, U. and Aggarwal, J., 1989. Structure from stereo - a review, IEEE Trans. on Pattern Recognition and Machine Intelligence, 19: 1489-1510.

Faugeras, O., 1993. Three-dimensional computer vision - a geometric viewpoint, Artificial Intelligence Series, The MIT Press, Cambridge, MA.

Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique, South Africa Journal of Photogrammetry, Remote Sensing and Cartography, 14: 175-187.

Hartley, R., 1995. A linear method for reconstruction from lines and points, Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, IEEE Computer Society Press, pp. 882 - 887.

Jain, A. and Vailaya, A., 1996. Image retrieval using colour and shape, Pattern Recognition, 29: 1233-1244.

Jones, G., 1997. Constraint, optimization, and hierarchy: reviewing stereoscopic correspondence of complex features, Computer Vision and Image Understanding, 65: 57-78.

Kittler, J., Christmas, W. and Petrou, M., 1993. Probabilistic relaxation for matching problems in computer vision, ICCV, Berlin, pp. 666-673.

Lan, Z. and Mohr, R., 1997. Robust location-based partial correlation, Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, Kiel, Germany, pp. 313 - 320.

Shashua, A., 1994. Trilinearity in visual recognition by alignment, Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden, Springer Verlag, pp. 479 - 484.