TOWARDS AUTOMATIC RELATIVE ORIENTATION FOR ARCHITECTURAL PHOTOGRAMMETRY

Frank A. van den Heuvel
Delft University of Technology
Department of Geodesy
Thijsseweg 11, 2629JA Delft, The Netherlands
E-mail: F.A.vandenHeuvel@geo.tudelft.nl

Commission V, WG V/2

KEY WORDS: relative orientation, wide-baseline stereo, vanishing point detection, feature-based matching, architecture

ABSTRACT

A major challenge in close-range photogrammetry and computer vision is the automation of model acquisition from imagery. Determining the relative position and orientation of close-range imagery is usually the first step in a procedure for modelling, assuming that the camera has been calibrated beforehand. An essential part of an orientation procedure based on image content is the establishment of correspondence between the images. The problem at hand is to find correspondence between two convergent images and their relative orientation simultaneously and automatically. In computer vision this is called the wide-baseline stereo problem. For the approach presented in this paper, a new feature-based matching procedure is designed that exploits the characteristics of the application by applying generic knowledge of the construction of the building. The procedure relies on rigorous statistical testing of constraints on the observations. The outline of the procedure is presented, as well as the results of experiments. Relative orientation was successfully detected for two images with an angle of 65 degree between the optical axes. It is shown that the procedure is robust with respect to unfavourable characteristics of the application, such as occlusions and repetitive structures in the building facades.

1. INTRODUCTION

Determination of the relative orientation of images is a prerequisite for object modelling. Although relative orientation of two images includes relative position, it however excludes the distance between the images and thus the scale of the model remains undetermined. Camera calibration, i.e. the determination of interior orientation (intrinsic) parameters, is also essential, however, in photogrammetry cameras are usually pre-calibrated. In this paper, the camera is assumed to be calibrated, although imagery of unknown interior orientation can be handled by applying a method for the estimation of interior orientation parameters from vanishing points (van den Heuvel, 1999).

Without using external measurement methods, such as the Global Positioning System (GPS), relative orientation of two images is based on corresponding features in the images. With only two images, we have to rely on point features because there is no information on relative orientation in corresponding image lines. With at least seven corresponding points available, the five parameters of relative orientation (in computer vision; the parameters of the *essential matrix*) can be determined by a direct solution (Förstner, 2000). Automatic relative orientation, therefore, is equivalent to solving the correspondence problem automatically. This topic has been extensively studied in photogrammetry, especially aerial photogrammetry (Heipke, 1997), and in computer vision (Pritchett and Zisserman, 1998), (Matas et al., 2001).

Our goal is the automatic relative orientation of two widely separated views, i.e. (strongly) convergent imagery. In computer vision this problem is known as *wide-baseline stereo*. Generic knowledge of the architectural application field is applied in the method and this facilitates a robust solution of the correspondence problem. For the method proposed here, prerequisites for the image acquisition are limited to:

- No major image rotation around the optical axis (κ less than 45 degree).
- Image tilt (rotation around x-axis, ω) less than 45 degree.
- Overlap between the two images.

The automatic procedure consists of three main steps (Figure 1). In each step different types of a priori object knowledge, i.e. knowledge on the construction of the building, are applied. Firstly, straight lines are extracted from the two images. This implies that the majority of building edges be assumed straight, and that lens distortion is limited or removed in advanced by resampling the image. Secondly, vanishing points are detected for each image. The building is assumed to be built along three orthogonal axes. The detection of at least two vanishing points related to these object orientations, results in an ambiguous orientation of the image relative to the building, and a set of edges of which the object orientation is known. In the third and last step, edges are intersected to points in image space, and finally point correspondence and relative orientation are detected simultaneously. The paper concentrates on this last step that relies on the assumption of planar facades, in which the object points recede.

The remainder of the paper is structured as follows. After discussing related research in the next section, the developed procedure for automatic relative orientation is described in section 3. In section 4, results from several experiments are analysed. Conclusions are drawn in section 5.

2. RELATED RESEARCH

In aerial photogrammetry the automatic relative orientation (as a part of automatic triangulation) is commonly available in digital photogrammetric workstations (Heipke, 1997). More and more, external measurement devices – usually an integration of GPS and INS – are used. For a range of aerial applications they make triangulation even superfluous. External measurement of image orientation is also used in close-range photogrammetry (Teller, 2001). For finding correspondence from the image itself, usually a feature-based matching technique is applied at different levels of the image pyramid. Wang (Wang, 1998) presents a structural matching method that requires an image pyramid. The method is applied to aerial images, as well as to terrestrial images of a building. In (Habib and Kelley, 2001) a robust feature-based method is described that does not require an image pyramid. This method is based on a Hough transform and relaxes the required quality of orientation angles to the order of 10 degrees. The method presented in this paper is feature-based, but differs from the approaches of Habib & Kelly and Wang in the more extensive use of generic object knowledge. Choices for parameterisation required by the Hough transform are avoided by the use of rigorous statistical testing of constraints on the observations. Furthermore, the proposed method does not rely on image segmentation required by structural matching.

In the use of image sequences without external measurement of camera orientation, approximate orientation values of an image are derived from the determined values of previous images in the sequence. In these *short-baseline* applications area-based matching is a commonly used technique to establish image correspondence at sub-pixel level (Pollefeys et al., 2000). With increasing baseline length, feature-based matching techniques are expected to be more successful, especially in applications where occlusions are frequent. The general approach of a feature-based matching procedure is described by Matas (Matas et al., 2001) and can be summarised as follows:

- Features that have viewpoint invariant characteristics are extracted in both images.
- Based on their characteristics, the features of two images are compared and a list of possible matches is established.
- A geometrically consistent set of matches is searched for. In this step the RANSAC algorithm is often applied (Fischler and Bolles, 1981).

Examples of this general approach are found in (Pritchett and Zisserman, 1998), (Tuytelaars and Van Gool, 2000), and (Baumberg, 2000). The method presented in this paper differs in the following aspects:

- Object information, as described in the previous section, is exploited at several stages. This makes the method robust, but limits its applicability to images of buildings or man-made structures with similar characteristics.
- The features are straight image lines – projections of the building edges – and not (invariant) image regions as in the approaches above. The intersection of two straight lines and a number of viewpoint invariant characteristics are used in the matching.
- Vanishing point detection is applied for the initial estimation of the orientation of the images relative to the object. Apart from a remaining ambiguity in the rotation matrix of each image, the vanishing point detection reduces the relative orientation problem to a relative position problem.
- The emphasis is on the use of geometric constraints for the selection of the set of correct matches, and not on photometric information. In the current application the photometric content in a region around for instance the corner of a window strongly depends on the viewpoint of the image as a result of discontinuities in the facade of the building. Furthermore, buildings often show repetitive patterns, such as identical windows. As a result, photometry is not a strong clue for detecting correct matches.
- Clustering is based on the results of statistical testing of geometric constraints. All possible combinations of tentative matches are tested. The number of tests is of the order m^2 (with m the number of tentative matches). This is in contrast to the RANSAC procedure in which the selection of the final solution is based on a randomly chosen subset of possible matches. In the method presented here the computational burden is reduced to an acceptable level by keeping the number of possible matches low.

3. PROCEDURE FOR RELATIVE ORIENTATION

As stated in the introduction, the procedure consists of three steps. An overview is depicted in Figure 1. If an uncalibrated camera is used, an additional step that follows the vanishing point detection is required (van den Heuvel, 1999). The detection of the three main vanishing points is more reliable in case of a calibrated camera. Then, after the detection of the first vanishing point, the search space for the other two is reduced considerably (van den Heuvel, 1998). In order not to complicate the description of the procedure, use of a calibrated camera is assumed.

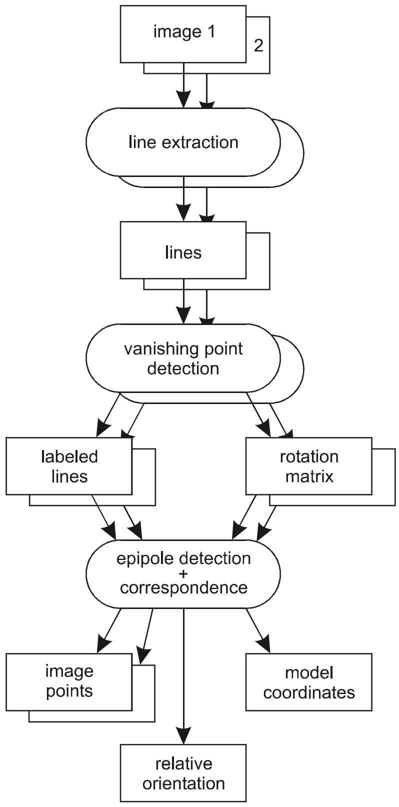

Figure 1. Overview of the procedure

3.1 Line feature extraction

Edges are automatically extracted by applying a line-growing algorithm (van den Heuvel, 2001). The coordinates of the endpoints represent the image lines. The interpretation plane is the plane in which both the image line and object edge recede (Figure 2). The image coordinates (x,y) are assumed to be corrected for lens and image plane distortions. Then the spatial vector ($\mathbf{x}$) related to an endpoint can be written as:

$$\mathbf{x} = (x, y, -f)\ , \ f\text{: focal length} \qquad (1)$$

The normal to an interpretation plane ($\mathbf{n}$) is computed from the rays ($\mathbf{x}$) to the endpoints of the line:

$$\mathbf{n} = \mathbf{x}_1 \times \mathbf{x}_2 \qquad (2)$$

This normal vector plays a major role in the vanishing point detection procedure that is summarised in the next section.

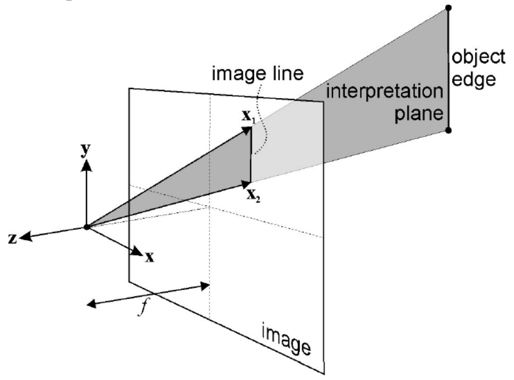

Figure 2: The interpretation plane

3.2 Vanishing point detection and image orientation

The extracted straight lines are input to a vanishing point detection procedure. This procedure starts with the selection of the longest extracted line which is assumed to intersect with other image lines in one of the three main vanishing points. In fact, the procedure is based on the analysis of intersection constraints on interpretation planes and thus does not require the lines to actually intersect in image space. Image lines are grouped, based on accepted statistical tests on the intersection constraint of three interpretation planes:

$$0 = (\mathbf{n}_1 \times \mathbf{n}_2) \cdot \mathbf{n}_3 \qquad (3)$$

A complete set of independent constraints is adjusted. The object orientations (**v**) that relates to the vanishing points are computed from the adjusted observations:

$$\mathbf{v} = \mathbf{n}_1 \times \mathbf{n}_2 \qquad (4)$$

Here we concentrate on the results, as the procedure has been described in detail in (van den Heuvel, 1998).

As a by-product of the vanishing point detection, the orientation of the image relative to the object is found. The rotation matrix can be constructed when at least two of the three vanishing points – associated with orthogonal object orientations – have been detected (Förstner and Gülch, 1999):

$$\mathbf{R} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \qquad (5)$$

However, this rotation matrix is ambiguous because there is no unique relation between the object orientations (**v**) from the vanishing points and the object coordinate system. To reduce the ambiguity in this rotation matrix it is assumed that the object orientation that is closest to the y-axis of the camera system corresponds to the Z-axis of the object system. Now four options for the rotation matrix remain, corresponding to four 90 degree rotations of the object system around the Z-axis. Note that – up to this ambiguity – the orientation of the images is found by the vanishing point detection, and thus only relative position remains to be determined in order to complete the relative orientation.

In the sequel of the procedure only those image lines are used that have been uniquely grouped to one vanishing point. Especially lines on or near the connecting line between two vanishing points (a so-called horizon) cannot be uniquely grouped and therefore these lines are not used for the final and main step of the procedure for automatic relative orientation described in the next section.

3.3 Correspondence and relative position

With the orientation of the images relative to the building known, the relative position and correspondence problem shows many similarities with the vanishing point detection problem. The line in space that connects the two projection centres intersects the images in a point that is called the epipole (Figure 3). The spatial orientation associated with a vanishing point is found as the intersection of interpretation planes, while the relative position vector is found as the intersection of epipolar planes. An interpretation plane is constructed from the two endpoints of an image line. An epipolar plane also needs two (corresponding) image points, one from each image. Because of these similarities the procedures for the detection of the epipole is also similar to the one for the vanishing point detection. However, there are some important differences. First, correspondence between the two images is unknown, while in the vanishing point detection an image line links the two endpoints. On the other hand there exist only one epipole (per image), while an image of a building usually shows two or more vanishing points.

Like for the vanishing point detection, statistical tests on the intersection of planes (now epipolar instead of interpretation planes) can be grouped for detecting correspondences that support the same epipole. However, the number of possible correspondences, and consequently the number of statistical tests to evaluate would explode without the use of additional object knowledge. With n points per image, there are n^2 correspondence hypotheses, and thus n^2 possible epipolar planes. As an intersection constraint involves three planes, the number of tests is of the order n^6.

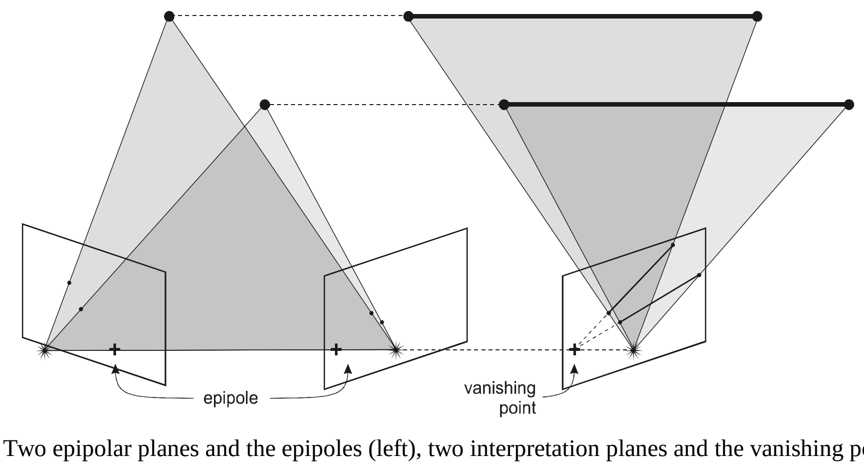

Figure 3 : Two epipolar planes and the epipoles (left), two interpretation planes and the vanishing point (right)

Introducing a new constraint reduces the number of tests to investigate. This constraint on the coplanarity of two object points involves two instead of three correspondences, reducing the order of the number of tests to n^4. Further significant reduction of the number of tests is achieved by computing a number of attributes for each constructed image point. A correspondence hypothesis is only set up when the attributes of the two points match to a sufficient degree.

The procedure for deriving relative position, *epipole detection* in short, consists of the following five steps:

1. Compute points in each image through line intersection.
2. Establish correspondence hypotheses between the computed points.
3. Compute statistical tests for coplanarity of all combinations of two correspondences.
4. Grouping of correspondences based on the test results and epipolar plane intersection constraints.
5. Refinement of grouped correspondences by overall adjustment and testing.

Each step is repeated for each of the four possible permutations of the rotation matrix of the second image. The exterior orientation of the first image determines the object co-ordinate system.

3.3.1 From lines to points

Points are created in each image as intersection of two image lines being projections of edges with a different (perpendicular) orientation in object space. This orientation is determined by the vanishing point detection procedure. Two lines are intersected when an endpoint of a line is within a preset distance (commonly set to 5 pixels) from the other line. Furthermore, the following attributes are computed and registered for each point:

- The orientation of the plane passing through the object point (perpendicular to the orientations of the two intersected lines)
- The type of junction (T-junction or L-junction)
- The orientation of the junction (4 options: Figure 4)
- The ratio of the lengths of the two edges

These attributes are all evaluated in object space, i.e. after projecting the lines onto a plane of which the orientation is known, again through the vanishing point detection.

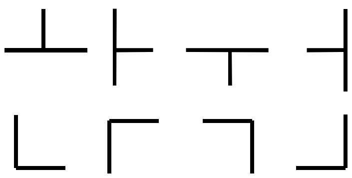

Figure 4: Four T-junctions (top) and four L-junctions

3.3.2 The correspondence hypotheses

In the next step of the procedure a list of possible correponding image points is set up. A correspondence is added to the list only when the attributes of points have been compared. Both points have to have the same (plane) orientation, the same junction type, and the same junction orientation. The ratio of the line lengths should be similar for both points with a typical maximum difference of 50%. The correspondence search turned out not to be sensitive to this criterion.

The creation of image points, their attributes, and consequently the correspondence hypotheses depend on the current permutation of the rotation matrix of the second image. With each permutation the vanishing point (i.e. object orientation) line labels for X and Y are exchanged.

3.3.3 Statistical testing of coplanarity

With the image orientations known, the epipole is defined by a minimum of only two correspondences (the intersection of the two related epipolar planes; Figure 3). Furthermore, because each image point is constructed from the intersection of two lines of which the spatial orientation is known from the vanishing point detection, the orientation of the object plane in that point is known. With the epipole derived from the two correspondences, the location of the two object points in model space can be computed by forward intersection, and thus the position of the two parallel planes through those points and the distance between the planes is known. The coplanarity constraint requires this distance to be zero.

For the formulation of the coplanarity constraint the ray of an image point is rotated to the object co-ordinate system. After replacement of equation (1) by

$$\mathbf{x}^i = \mathbf{R}^i (x, y, -f) \tag{6}$$

in which i refers to one of the two images, equations (2) and (4) are used to compute the orientation vector of the epipole ($\mathbf{v}$) from two correspondences. Knowing the epipole, forward intersection results in the distance (d) from the projection center to an object point. This distance is projected onto the normal $\mathbf{n}_p$ of the object plane. For image i and correspondence j:

$$\bar{d}^i_j = d^i_j \frac{\mathbf{x}^i \cdot \mathbf{n}_p}{\left|\mathbf{x}^i\right|} \tag{7}$$

Normal $\mathbf{n}_p$ is a unit vector in the object X or Y direction, dependent on the orientation of the façade. For the formulation of the coplanarity constraints the first image is used (in principle, using the second image would yield identical results). The constraint for coplanarity of the two object points can now be written as:

$$0 = \bar{d}^1_1 - \bar{d}^1_2 \tag{8}$$

The statistical testing of the hypothesis is explained in (van den Heuvel, 1998). The coefficients of the linearised form of equation (8) are derived numerically.

3.3.4 Clustering of correspondences

For all combinations of two possible correspondences the statistical test based on constraint (8) is evaluated. Each test links two correspondences that define an epipole. The clustering procedure aims at grouping of correspondences of which the inter-correspondence tests are accepted, and at the same time support the same epipole. In order to verify the latter, a statistical test is used that is based on the intersection of epipolar planes. This constraint is identical to the intersection of interpretation planes constraint (3).

The clustering procedure includes the following steps:

- For each accepted test (8) it is checked whether one of the two correspondences is present in an existing cluster. If this is not the case, a new cluster is established.
- If a correspondence of an accepted test is present in an existing cluster, the other correspondence becomes a candidate for that cluster.

- The candidate correspondence becomes a member of the cluster if all tests of constraints (3) are accepted. These tests involve three correspondences: the candidate and two correspondences already in the cluster.

This procedure is repeated as long as new clusters are created. The result is that overlap between clusters can be considerable. Many of these clusters contain a set of correct correspondences with minor differences. The clusters that contain more than a minimum number of correspondences are analysed in more detail by applying an integrated adjustment of all constraints.

3.3.5 Overall adjustment and testing

The clustering procedure results in groups of correspondences that support the same relative orientation, while the related object points are expected to recede in the same object plane. An overall adjustment is set up for all clusters with more than a minimum number of correspondences. The functional model contains n-1 condition equations for coplanarity (equation (8), n is the number of correspondences), and n-2 condition equations for intersection of epipolar planes (equation (3)). A set of independent equations results. The adjusted epipole is computed from adjusted observations. Two different types of statistical tests are applied. First an overall test or Fisher test is applied. The second test examines the alternative hypothesis of an error in a single correspondence. If such a test is rejected, the correspondence is removed from the cluster and the model is built again. This iterative testing procedure stops if all correspondence tests are accepted. The cluster with the largest number of correspondences of which no correspondence tests are rejected is selected. This cluster is expected to contain corresponding image points of which the related object points are in (or near to) a plane. More details on the statistical testing can be found in (van den Heuvel, 1998).

3.3.6 Towards automatic reconstruction

Apart from the primary cluster detected as described above, other clusters can also contain correct correspondences of which the related object points are in a different plane. In order to detect such clusters, all clusters that meet the following two requirements are incorporated in the overall adjustment:

1. The cluster does not have a correspondence (nor an image point) in common with the primary cluster.
2. The correspondences of the cluster confirm the epipole of the primary cluster.

For all correspondences – those of the primary as well as those of the clusters selected by the criteria above – the epipolar plane intersection constraints are set up. Of course, the coplanarity constraints are only applied to correspondences of the same cluster.

With the integrated adjustment of more than only the primary cluster, there is not only additional evidence gathered for the epipole, but at the same time different object planes are being detected. Preliminary faces can be created by a bounding box around the object points of a cluster. Object planes are then to be intersected to find the edges of the building. The detection of object planes is a by-product of the proposed procedure for automatic relative orientation, but also a first step towards automatic reconstruction. The experiments described in the next section aim at the detection of the primary object plane only.

4. EXPERIMENTAL RESULTS

In this section an experiment is discussed in which the procedure for automatic relative orientation is applied to three images of a historic building. The images are taken from ground level with a handheld calibrated digital camera (1536x1024 pixels). They were taken from the south-south-west (SSW), south-east (SE), and east (E) approximately (Figure 5). The number of extracted straight lines can be found in Table 1.

Figure 5: The three images (labelled SSW, SE, and E)

The a priori precision of the endpoints of the lines was set to 1 pixel standard deviation in the vanishing point and the epipole detection. The "vanishing lines" are the lines that were uniquely grouped to one of the three vanishing points. These lines are displayed in Figure 6. Especially near the horizon line of image SE a considerable number of lines is lost because the vanishing point detection cannot distinguish between the left and right façade. As a result, for this image most of the intersection points are created in the upper part of the facades. Some statistics of the epipole detection are listed in Table 2.

	SSW	SE	E
# lines	339	457	202
# vanishing lines	286	276	143
# intersection points	164	78	30

Table 1: Numbers of extracted lines and points

	SSW - SE	SE – E
# correspondence hypotheses	1109	144
# coplanarity tests	182899	3308
# accepted tests	15825	1102
# clusters	5638	424
max. # points in a cluster	16	9
# clusters accepted	97	85
max. # points in a cluster	15	7
Fisher-test / critical value	4.4	1.0
Deviation manually measured	1.7 deg	1.6 deg

Table 2: Statistics of the epipole detection

The detected correspondences are displayed in Figure 7. Note that for the first image pair only those corresponding points are detected that are on the central part of the façade because this part is in a different plane from the rest of the façade. In fact, all possible correspondences are detected. However, looking at the location of the image points in detail, two corresponding points are often not at exactly the same location on the building. The reason is that many edges border occlusions. When there are several points created close together – which often is the case in the corners of the windows – the statistical testing cannot distinguish between different possible correspondences. Indeed, many of the accepted clusters are very similar in the correspondences they contain and their Fisher-test. Furthermore, as a result of a large number of "imperfect" correspondences (the image points are not projections of exactly the same point on the building), the overall Fisher-test is often rejected, while all tests of individual correspondences are accepted. The lack of "perfect" correspondences reduces the precision that can be reached. To check this precision, relative orientation of both pairs was determined through least-squares adjustment of 20 manually measured points. The deviations in the orientation of the detected epipole are below 2 degree (Table 2). As expected the computational burden of the proposed method is considerable. For the first pair computational time is in the order of 10 minutes on a modern PC.

Figure 6: The "vanishing lines" (left) and created points.

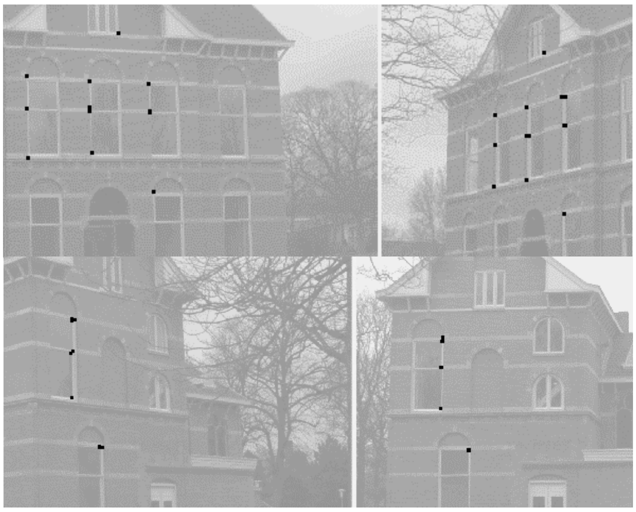

Figure 7: The detected correspondences of both pairs.

5. CONCLUSIONS

A new method for automatic relative orientation has been presented. It relies on the extraction of straight image lines and their vanishing point labelling. Vanishing point detection is a crucial step in the procedure that results in an ambiguous orientation of the images relative to the building. The epipole detection shows many similarities with the vanishing point detection. Both are based on clustering of rigorous statistical tests and adjustment of constraints on the observations.

Experiments show that relative orientation can be detected successfully between two images with an angle of 65 degree between the optical axes (see section 4, first image pair), while the difference in orientation with a manually determined relative position vector was less than 2 degree.

The proposed method can be regarded as a first step towards automated reconstruction because the model coordinates of the corresponding points and the parameters of the plane in which they recede become available as a by-product.

REFERENCES

Baumberg, A., 2000. Reliable feature matching across widely separated views, Computer Vision & Pattern Recognition (CVPR) 2000, pp. 774-781.

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, Vol. 24(6), pp. 381-395.

Förstner, W., 2000. New orientation procedures. International Archives of Photogrammetry and Remote Sensing, Vol. 33 part 3, pp. 297-304.

Förstner, W. and Gülch, E., 1999. Automatic orientation and recognition in highly structured scenes. J. of Photo-grammetry and Remote Sensing, Vol. 54, pp. 23-34.

Habib, A. and Kelley, D., 2001. Automatic relative orientation of large scale imager over urban areas using modified iterated Hough transform. J. of Photogrammetry and Remote Sensing, Vol. 56, pp. 29-41.

Heipke, C., 1997. Automation of interior, relative, and absolute orientation. ISPRS J. of Photogrammetry and Remote Sensing, Vol. 52, pp. 1-19.

Matas, J., Urban, M. and Pajdla, T., 2001. Unifying view for wide-baseline stereo matching. In: B. Likar (Editor), Computer Vision Winter Workshop. Slovenian Pattern Recorgnition Society, pp. 214-222.

Pollefeys, M., Koch, R., Vergauwen, M. and Van Gool, L., 2000. Automated reconstruction of 3D scenes from sequences of images. ISPRS J. of Photogrammetry and Remote Sensing, Vol. 55(4), pp. 251-267.

Pritchett, P. and Zisserman, A., 1998. Wide baseline stereo matching, Sixth International Conference on Computer Vision, pp. 754 -760.

Teller, S., 2001. Scalable, controlled imagery capture in urban environments. Report 825, MIT Laboratory for Computer Science.

Tuytelaars, T. and Van Gool, L., 2000. Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions, British Machine Vision Conference BMVC'2000, pp. 412-425.

van den Heuvel, F.A., 1998. Vanishing point detection for architectural photogrammetry. In: H. Chikatsu and E. Shimizu (Editors). International Archives of Photogrammetry and Remote Sensing, Vol. 32 part 5, pp. 652-659.

van den Heuvel, F.A., 1999. Estimation of interior orientation parameters from constraints on line measurements in a single image. In: P. Patias (Editor). International Archives of Photogrammetry and Remote Sensing, Vol. 32 part 5W11, pp. 81-88.

van den Heuvel, F.A., 2001. Object reconstruction from a single architectural image taken with an uncalibrated camera. Photogrammetrie, Fernerkundung, Geoinformation, Vol. 2001(4), pp. 247-260.

Wang, Y., 1998. Principles and applications of structural image matching. ISPRS J. of Photogrammetry and Remote Sensing, Vol. 53, pp. 154-165.