Automatic Relative Orientation of Aerial Images

Tang Liang and Christian Heipke

Abstract
An approach is described for automatic relative orientation of a stereopair of digital aerial images. The concept and the implementation are based on practical conditions with respect to available a priori knowledge, speed of computation, and obtainable accuracy. Feature-based image matching using point features extracted with a modified version of the Moravec operator and a coarse-to-fine strategy are incorporated into the approach. In higher image pyramid levels, where images are small in size and of low resolution, the entire model area is searched for interest points. In lower levels, window tracking is carried out in order to speed up the entire procedure and to stabilize the final results. In all levels, matching is based on geometric as well as radiometric constraints. The approach was developed as one of the automatic-oriented software components of a digital photogrammetric workstation.

Results obtained from ten aerial image pairs with scales ranging from 1:3,000 to 1:34,000 and scanned with a pixel size of 15 μm, thus yielding some 235 Megabytes per image, are presented. In each case, more than 130 well distributed points were extracted. The obtained root-mean-square standard deviations of the image coordinates consistently lie between 3.2 and 3.6 μm or 0.21 and 0.24 pixels. A human operator checked the resulting models on an analytical plotter. The models were found to be free of y-parallax. The elapsed computing time was approximately 4 minutes per image pair on a Silicon Graphics Iris Indigo workstation with R4000 processor. This means that the procedure runs as fast as, if not faster than, a human operator can carry out the relative orientation while yielding the same level of accuracy. Thus, it could be shown that the presented method for automatic relative orientation is operational for practical applications.

Introduction
With the increasing availability of digital imagery as well as high performance computer hardware and software, the automation of the photogrammetric compilation process has become possible in principle for the first time. Techniques from image processing and computer vision have been successfully employed for the automatic generation of digital terrain models and orthoimages. The emergence of digital photogrammetric workstations (DPWS) and their use by government agencies and private companies signalize the commercialization of these new technologies (Heipke, 1993; Heipke, 1995).

Today's essential research and development issues in digital photogrammetry include the automation of image orientation. Solutions for the reconstruction of the interior orientation parameters using the fiducial marks as identical points are commercially available (Mayr, 1993). The relative and absolute orientation of an image pair remain to be automated for stereo processing.

Absolute orientation includes the identification of ground control features and the measurement of the corresponding image coordinates. Implementations of automatic absolute orientation are rare and work under restricted conditions only (see Haalal and Vosselman (1992), Schickler (1992), and Gurlie (1994) for related work). It should be noted that direct measurements of the elements of exterior orientation can be derived from signals of the global positioning system (GPS) and inertial navigation systems (INS) (see, for example, the January 1993 special issue of Photogrammetric Engineering & Remote Sensing on GPS Photogrammetry). These measurements, in principle, make ground control information as such, and thus also their identification in the images, obsolete. As it stands today, however, at least a small number of control points are still needed to ensure a reliable solution.

In contrast to absolute orientation, relative orientation does not require the recognition of specific features. The conjugate points used only have to be geometrically well distributed in the model area. Relative orientation is a prerequisite in order to provide users from photogrammetry and other disciplines with parallax-free stereo viewing for photogrammetric data collection, interpretation purposes, and a number of other tasks. Relative orientation is also the core for any automatic point transfer system. In this case, the detection of areas covered by more than two images and multi-image matching pose an additional challenge (Schon and Toth, 1995). Moreover, the parameters of relative orientation are needed for epipolar resampling of digital images. Therefore, automatic relative orientation is an essential procedure for the automation of further procedures in photogrammetric stereo processing. Around the beginning of the nineties, investigations began in several institutions on automating the procedure of relative orientation of an aerial stereopair (Hannah, 1989; Schenk et al., 1991; Haalal et al., 1993). However, none of these developments has yet reached photogrammetric practice.

In this paper, we describe a new approach for automatic relative orientation of a stereopair of digital aerial images and report on the obtained results. The approach represents a component of an automation-oriented DPWS software package and was tested in an operational environment. Reports on preliminary versions of the algorithm were given in Tang and Heipke (1993), Tang and Heipke (1994), and Hellwich et al. (1994). In the next section, an overview of the implemented algorithm is given, consisting of feature extraction, feature matching, window tracking, computation of relative orientation parameters, and evaluation of results. Subse-
quirely, the results obtained from ten aerial image pairs with scales ranging from 1:3,000 to 1:34,000 and scanned with a pixel size of 15 μm, thus yielding some 235 Megabytes per image, are presented.

**Description of the Approach**

**General Outline**

Four key points of this approach can be identified:
- it is feature based, using point features;
- it includes a coarse-to-fine strategy;
- it is controlled by geometric as well as radiometric constraints; and
- it uses a large number of features.

Feature-based image matching is more invariant against radiometric changes and generally runs faster than does the area-based approach. Therefore, feature-based matching was chosen for this project. Initially, it was intended to include an area-based algorithm at the last stage of the computations in order to meet high accuracy requirements. It was found, however, that this is not necessary. The features used are points rather than straight or curved lines. Points carry less descriptive information than do lines and are thus less distinct. However, points are invariant with respect to the central projection, and they are easier to describe.

Photogrammetric images show a wealth of detailed information. In order to preserve this information, a digital image must be scanned with pixel sizes ranging from about 10 to 25 μm, and sometimes even smaller. This corresponds to as much as 100 Megabytes per image or more. For processing and managing this large amount of data efficiently, coarse-to-fine strategies based on image pyramids are at the very heart of every modern photogrammetric matching algorithm. These strategies also further reduce the necessity for accurate initial values for the points to be matched. It was therefore straightforward to adopt a coarse-to-fine strategy for our approach.

Due to the lack of background knowledge, an automatic procedure does not have the same capabilities in selecting locally distinct points as a human has. Besides, the image area processed simultaneously normally amounts to a few pixels only. Thus, no contextual information of that area is available to the procedure. In order to reduce the effect of these shortcomings, different geometric and radiometric constraints, namely, epipolar geometry, local plane fitting, and cross correlation, are incorporated into the approach. These constraints also serve to circumvent or recognize wrong correspondences. Point pairs are only considered "conjugate" once they satisfy all these constraints.

The use of a multitude of point features, as compared to an analytical relative orientation with few points, increases the reliability of the results, offering easy ways to eliminate blunders. Moreover, because of the higher redundancy, the individual observations may be less accurate, but still yield the same standard deviations for the orientation parameters. Our approach typically relies on a few hundred conjugate points as compared to 6, 12, or sometimes up to 15 points in the analytical case.

Figure 1 shows the concept of the algorithm for automatic relative orientation of aerial images. We restrict ourselves to aerial images, because the two images are assumed to approximately fulfill the conditions of the normal case of stereo photogrammetry. Along with the image data, some readily available input information must be provided:

- the interior orientation parameters of the camera,
- the relationship between the image and pixel coordinate systems,
- the order of the two images (i.e., which is the left image and which is the right one), and
- approximate values for the overlap in the x- and y-directions.

The relative orientation procedure starts from the highest pyramid level with the smallest image size and the lowest resolution, and ends at the lowest level with the original size and resolution. It is divided into two sub-procedures. The first one runs from the highest level through to a so-called "intermediate pyramid level." In each level, features are ex-
I have prepared these operators has shown that the Moravec and the have chosen to use the Moravec operator. The reasons are thus runs faster than the window of specified size. Given a window for the window center \((i + 0.5(M + 1), j + 0.5(N + 1))\) in the image \(g(i, j); i = 1, 2, \ldots, M; j = 1, 2, \ldots, N\), the four mean values \(\{V_1, V_2, V_3, V_4\}\) and the interest value \(V\) for the window center \((i + 0.5(M + 1), j + 0.5(N + 1))\) in the image \(g(i, j); i = 1, 2, \ldots, M; j = 1, 2, \ldots, N\) can be calculated as follows:

\[
V_i = \frac{1}{M(N-1)} \sum_{k=1}^{M} \sum_{l=1}^{N} \left[ g(i+k,j+l) - g(i+k,j+l+1) \right]^2
\]

\[
V_j = \frac{1}{(M-1)N} \sum_{i=1}^{M-1} \sum_{j=1}^{N} \left[ g(i+k,j+l) - g(i+k+1,j+l) \right]^2
\]

\[
V_s = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left[ g(i+k,j+l) - g(i+k,j+l+1) \right]^2
\]

\[
V_v = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left[ g(i+k,j+l) - g(i+k+1,j+l) \right]^2
\]

\[
V = \text{Min}(V_i, V_j, V_s, V_v)
\]

Using the Moravec operator, a displacement (bias) between a real feature and the window center of the highest \(V\) in the local surrounding can occur (Dreschler, 1981). We partly compensate for this effect by using the center of gravity with respect to the squared grey value differences in the window as the feature position rather than using the window center (see Equation 2). In this way, we also obtain subpixel accuracy in the feature position: i.e.,

\[
i_v = i + \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) (k + 1) + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) (k, l) (k + 1)}{\sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) + g(l, k) + g(k, l)}
\]

\[
i_h = j + \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) (l + 1) + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) (k, l) (l + 1)}{\sum_{i=1}^{M} \sum_{j=1}^{N} g(i, j) + g(k, l) + g(l, k)}
\]

In our algorithm, features are extracted from the two images independently. In each image, the overlap area is subdivided into small adjacent image cells. Within each image cell \(V\) is calculated for all pixels, and only the highest value is retained for further computations (local non-maximum suppression). In order to keep only the best features for matching on the one hand and to guarantee their geometric distribution within the overlap area on the other hand, image cells are grouped into subareas and a threshold is defined for each subarea: only a predefined percentage of the extracted features, i.e., those with the highest interest values, is retained, while the others are discarded (regional or global non-maximum suppression).

Feature Matching

Given the two lists of remaining features from the two images, the task of feature matching is to find the corresponding feature pairs. This is accomplished by using a mathematical model for transforming features from one image to the other and a similarity measure for forming and evaluating the matches. We use the collinearity equations as the mathematical model. For the highest pyramid level, approximate orientation parameters of the two images are derived from the input data, and the terrain surface of the overlap area is assumed to be of a constant elevation. For the lower pyramid

Feature Extraction

Features in an image are distinct points, edges (or lines), and areas. In our approach, we deal with point features only, because points provide the most stable geometry for relative orientation. Interest operators which locate point features in an image can be found in the literature (Moravec, 1977; Nah, 1980; Dreschler, 1981; Förstner, 1986). Work on comparing these operators has shown that the Moravec and the Förstner operators perform best for real images in matching applications (Luhmann and Althoff, 1986). The Förstner operator has some theoretical advantages compared to the classical Moravec operator (e.g., rotation invariance, and the potential for subpixel accuracy). Rotation invariance is less important for aerial images because, in general, the azimuth difference is rather small. Subpixel accuracy, however, is important for photogrammetric applications. Nevertheless, we have chosen to use the Moravec operator. The reasons are that it can also be tuned to subpixel accuracy (see below), and the resulting operator requires less computations and thus runs faster than the subpixel version of the Förstner operator.

The Moravec operator is based on the mean values of squared grey value differences along the horizontal direction, the vertical direction, and the two diagonal directions over a window of specified size. Given a window \((k, l); k = 1, 2, \ldots, M; l = 1, 2, \ldots, N\), the four mean values \(\{V_1, V_2, V_3, V_4\}\) and the interest value \(V\) for the window center \((i + 0.5(M + 1), j + 0.5(N + 1))\) in the image \(g(i, j); i = 1, 2, \ldots, M; j = 1, 2, \ldots, N\) can be calculated as follows:

\[
V_i = \frac{1}{M(N-1)} \sum_{k=1}^{M} \sum_{l=1}^{N} \left[ g(i+k,j+l) - g(i+k,j+l+1) \right]^2
\]

\[
V_j = \frac{1}{(M-1)N} \sum_{i=1}^{M-1} \sum_{j=1}^{N} \left[ g(i+k,j+l) - g(i+k+1,j+l) \right]^2
\]

\[
V_s = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left[ g(i+k,j+l) - g(i+k,j+l+1) \right]^2
\]

\[
V_v = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left[ g(i+k,j+l) - g(i+k+1,j+l) \right]^2
\]

\[
V = \text{Min}(V_i, V_j, V_s, V_v)
\]
levels, both the orientation parameters and the model points obtained from the previous level are available. They are used in the collinearity equations for the transformation of a feature from the left to the right image. The required elevation for the feature is computed from the known model points of the previous level in the following way: first, the model points are projected into the left image. Then, the closest projected model point with respect to the feature under consideration is searched for. Subsequently the elevation of the found model point is used as an approximate height for this feature.

To find matches for the feature, the epipolar line in the right image is determined by using the collinearity equations together with the currently valid orientation parameters. In principle, the conjugate feature in the right image should be found along this line. However, due to the limited accuracy of the available orientation parameters, a search area of a certain size is defined along the epipolar line. All features of the right image within the search area are considered to be potential matches for the feature in the left image. For each pair thus established, the correlation coefficient $Q$ of the surrounding windows is calculated. If $Q$ is larger than a threshold, the match is considered valid in terms of the radiometric constraint, and, for the feature in the right image, the distance to the epipolar line of the left image feature is computed. Among all these possible matches, the $n$ features with the shortest distances are chosen.

**Computation of Relative Orientation Parameters**

After feature matching, the parameters of relative orientation are calculated for each pyramid level in a least-squares bundle adjustment using the candidates for conjugate points as observations. The unknowns are the orientation parameters $Y$, $Z$, $\phi$, $\omega$, and $\kappa$ of the second image and the model coordinates of the conjugate points. Initial values for the unknown orientation parameters are derived from the additional input information and, in the lower pyramid levels, from the previously computed parameters. Initial values for the model coordinates are computed by means of a local forward intersection. In the stochastic model, the observations are considered to be independent and of equal accuracy.

In the course of the adjustment, blunders are removed from the candidate pairs. The residuals of every pair are analyzed, and pairs with residuals larger than the pixel size multiplied by some scale factor are considered as blunders, and are eliminated. This relatively easy method of blunder detection works, because the matching accuracy is related to the pixel size, and because of the high redundancy. After this estimation procedure has converged, the remaining pairs are checked for multiple matches. These are all discarded, and, finally, a least-squares bundle adjustment is carried out, resulting in the relative orientation parameters, the model coordinates of the conjugate points for the current pyramid level, the standard deviation $\sigma_0$ of unit weight, as well as the theoretical standard deviations for all unknowns. $\sigma_0$ can be considered as a measure for the root-mean-square standard deviation of the image coordinates.

It should be noted that, while the use of epipolar geometry prevents blunders in the direction perpendicular to the baseline, it cannot detect blunders parallel to the baseline. These blunders do not have an effect on the parameters of relative orientation. However, they result in model points with grossly incorrect heights. To eliminate these blunders, we assume that the whole overlap area is a plane in the first sub-procedure, ignoring terrain undulations at these small image scales. Parameters for this plane are determined by a least-squares adjustment using the available model coordinates of the conjugate points as observations. Points with residuals larger than three times the resulting standard deviation are considered blunders and are discarded. In general, given rough terrain, the standard deviation will be large, and only points with very large residuals will be eliminated, whereas given smooth terrain, the standard deviation will be small, and points with smaller residuals will be considered blunders. Thus, this blunder elimination technique is adaptive to terrain undulations.

**Evaluation of Different Control Parameter Sets**

As mentioned before, in the first sub-procedure a number of different control parameter sets are used to compute the relative orientation of the two images. These parameters include the size of an image cell and a subarea, and the percentage threshold for feature extraction; the size of the search area and the size of the window to compute the correlation coefficient, the correlation coefficient threshold, and the number of allowed multiple matches for feature extraction; and the scale factor for blunder elimination in the computation of the relative orientation parameters. The determining factors for evaluating the results for each set of control parameters are primarily the number and distribution of the conjugate points. The theoretical standard deviations of the five orientation parameters are effected by both factors; however, they are difficult to interpret. A better measure is the theoretical standard deviation of a point within the model area.

For a parameter set to be accepted as correct, two conditions have to be fulfilled: first, the number of conjugate points must be above a predefined threshold (we have chosen a value of 30); and second, if the model area is divided into three by five adjacent areas of equal size, each of these areas must contain at least one conjugate point. In order to find the best set out of all accepted ones, we subsequently compute the planimetric standard deviations for the four corner points of the model on the ground using the covariance matrix of the five parameters of relative orientation and error propagation. Then we take the maximum of these four values to be the final error measure of a given parameter set. The set of control parameters with the smallest error measure is considered to be the best set, and the corresponding orientation parameters are forward to the next pyramid level.

**Window Tracking**

The idea of window tracking is based on the consideration that a feature from a given level can be only one of the following:

- an indication for a good feature on a lower level.
- a representative of several good features on a lower level.
- a pseudo-feature which disappears later.

Thus, there is a fair chance of finding enough good features on the lower level by window tracking. Window tracking is performed on all levels below the intermediate one. Figure 2 shows the principle. It involves the following steps:

- project a pair of conjugate points onto the next lower level and define a search window around each projected point in the image;
- subdivide each window into image cells and extract features as described in the section on Feature Extraction;
- carry out feature matching as described in the section on Feature Matching;
- compute model coordinates for each pair of candidates for
The described algorithm was tested on a number of stereo pairs showing different image texture and terrain types. Altogether, the relative orientations of ten pairs were determined. All images were scanned using the Zeiss/Intergraph PhotoScan PSI, with a pixel size of 15 μm and 8 bits per pixel. In each case the resulting amount of data was approximately 335 Megabyte per image in the original resolution.

Two Examples Discussed in Detail
First, two examples shall be discussed in some detail in order to provide an idea of the obtainable results. The first example deals with images taken in Finland with an image scale of 1:15,000 and with 60 percent overlap. The depicted area serves partly agricultural purposes and is partly forested. The terrain surface is mostly smooth, but contains some breaklines. The maximum elevation difference is about 50 m. The images of the second example depict an area around Marchetsreut in the Bavarian Forest/Germany. The image scale is also 1:15,000, and the overlap amounts to 80 percent. Again, agricultural and forested areas are depicted. The terrain surface is undulating, having an elevation difference of up to 160 m.

Image pyramids consisting of seven levels (level 0 through level 6) were constructed using a simple low-pass filter. Subsequently, the automatic relative orientation procedure as described in the previous chapter was carried out, using pyramid level 5 as the intermediate level. Tables 1 and 2 show the numerical results for the two examples. For each level, the number of pixels per image, the pixel size in μm, the number of extracted features for the left and the right image, the number of candidates for conjugate points, the number of points eliminated during the computation of the relative orientation parameters, the number of conjugate points, and the standard deviation of the image coordinates in μm and in pixels are given. Figures 3 and 4 show the features and the computed conjugate points extracted from level 5, respectively, for the Finland example. Figure 5 depicts the conjugate points of the original resolution (level 0). In Figures 6, 7, and 8, the same results can be seen for the Marchetsreut example.

The following observations can be made from the obtained results:

- As expected, a large number of conjugate points is computed in each pyramid level. The same holds for the number of extracted features and of candidates for conjugate points.
- The standard deviation of the image coordinates, σ, in level 0 is 3.4 μm and 3.5 μm or 0.23 and 0.23 pixels, respectively. It is comparable to the accuracy of manual measurements. It should be noted that σ expressed in pixels is basically constant for all pyramid levels (except level 6 of the Marchetsreut example), a result which confirms the statement that the accuracy of photogrammetric image matching primarily depends on the pixel size.
- As can be seen from the figures, the distribution of conjugate points is even throughout the whole model area, and the distribution doesn't change between levels 5 and 6.
- The number of points eliminated during the least-squares computation of the orientation parameters is relatively high. The reason is that a large value had been selected for the number of possible multiple matches, i.e., n = 6, and these multiple matches have been eliminated during the subsequent step (see the sections on Feature Matching and on Computation of Relative Orientation Parameters).

In interpreting these results, it should be kept in mind that in the first sub-procedure the control parameters may vary from level to level and from example to example, whereas in the second sub-procedure they are constant for both examples and all levels. This explains some additional observations which can be made from these tables, e.g., the low number of 158 conjugate points in level 6 of the Finland...
TABLE 1. NUMERICAL RESULTS OF THE FINLAND EXAMPLE

<table>
<thead>
<tr>
<th>level</th>
<th>pixels/image</th>
<th>pixel size [μm]</th>
<th>extracted features</th>
<th>candidate pairs</th>
<th>eliminated pairs</th>
<th>conjugate pairs</th>
<th>σ_o [μm]</th>
<th>σ_o [pixel]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>240^2</td>
<td>960</td>
<td>692</td>
<td>158</td>
<td>158</td>
<td>197.7</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>480^2</td>
<td>480</td>
<td>2430</td>
<td>528</td>
<td>490</td>
<td>127.2</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>960^2</td>
<td>240</td>
<td>3989</td>
<td>2909</td>
<td>1367</td>
<td>53.6</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1920^2</td>
<td>120</td>
<td>4262</td>
<td>3472</td>
<td>1942</td>
<td>53.6</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3840^2</td>
<td>60</td>
<td>4383</td>
<td>3051</td>
<td>1327</td>
<td>13.4</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7680^2</td>
<td>30</td>
<td>4538</td>
<td>2518</td>
<td>1046</td>
<td>6.6</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>15360^2</td>
<td>15</td>
<td>4771</td>
<td>1823</td>
<td>603</td>
<td>3.4</td>
<td>0.23</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2. NUMERICAL RESULTS OF THE MARCHETSREUT EXAMPLE

<table>
<thead>
<tr>
<th>level</th>
<th>pixels/image</th>
<th>pixel size [μm]</th>
<th>extracted features</th>
<th>candidate pairs</th>
<th>eliminated pairs</th>
<th>conjugate pairs</th>
<th>σ_o [μm]</th>
<th>σ_o [pixel]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>240^2</td>
<td>960</td>
<td>862</td>
<td>470</td>
<td>459</td>
<td>731.9</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>480^2</td>
<td>480</td>
<td>3924</td>
<td>488</td>
<td>436</td>
<td>155.1</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>960^2</td>
<td>240</td>
<td>3466</td>
<td>2852</td>
<td>1080</td>
<td>52.4</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1920^2</td>
<td>120</td>
<td>3762</td>
<td>2964</td>
<td>1171</td>
<td>25.3</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3840^2</td>
<td>60</td>
<td>3065</td>
<td>3532</td>
<td>1188</td>
<td>13.4</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7680^2</td>
<td>30</td>
<td>4131</td>
<td>4413</td>
<td>1173</td>
<td>6.9</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>15360^2</td>
<td>15</td>
<td>4276</td>
<td>4763</td>
<td>977</td>
<td>3.5</td>
<td>0.23</td>
<td></td>
</tr>
</tbody>
</table>

example as compared to 459 in the Marchetsreut example. In this case, a higher threshold for the cross-correlation coefficient was used in the best parameter set of the Finland example.

Operational Test of the Approach

The obtained results encouraged us to have the algorithm run on another eight data sets, constituting an operational test of the presented approach. It should be noted that the image scales range from 1:3,000 to 1:34,000, and the terrain types from built-up to mountainous.

Table 3 presents some details about the test material and summarizes the results. The relative orientation of all ten examples has been computed totally automatically. While the actual number of conjugate points varies, it can be seen that the results are in accordance with the findings reported for the first two examples. In each case, more than 150 well distributed conjugate points have been extracted from the imagery. The obtained root-mean-square standard deviations of the image coordinates consistently lie between 3.2 and 3.6 μm or 0.21 and 0.24 pixels, which, as stated above, is comparable to the accuracy of manual measurements. A human operator checked the resulting models on an analytical plotter. The models were all found to be free of y-parallaxes. In order to reach smaller root-mean-square standard deviations, the most promising direction to take is to work with smaller
pixel sizes because, as stated above, the matching accuracy is intrinsically related to the pixel size.

All tests were carried out on a Silicon Graphics Iris Indigo R4000 workstation. The elapsed computing time was approximately 4 minutes per image pair, given that the image pyramid already exists. The latter is also needed for other tasks like automatic interior orientation, and should be generated immediately after scanning the images. Thus, the procedure runs as fast as, if not faster than, a human operator can measure the conjugate points for relative orientation while yielding the same level of accuracy.

Conclusions

Automatic relative orientation is an essential procedure for the automation of further procedures in photogrammetric stereo processing. It is the core of any system for automatic aerial triangulation, and is also a prerequisite for epipolar resampling of digital images, and for stereoviewing.

The approach to automatic relative orientation described in this paper can be characterized by:

- being based on the condition of aerial photogrammetry;
- being practice-oriented with respect to available a priori knowledge, computational speed, and obtained accuracy;
being fully automated;
- using image pyramids for coarse-to-fine processing;
- using point features for image matching;
- using proper mathematical models and similarity measures for feature matching; and
- using window tracking to speed up the procedure.

The reported results for ten aerial image pairs of varying image scales and terrain characteristics are very satisfactory, and show that the presented method for automatic relative orientation is operational for practical applications. The computations take less time than a human operator needs in order to carry out a relative orientation, and the attained level of accuracy is similar to that of manual measurements as checked by means of the standard deviation of the image coordinates and by independent visual inspection of the resulting stereo models.

In the near future, the algorithm will be implemented into a DPWS environment. Thus, interactive measurement capabilities will be available in cases where the automatic procedure fails. Further tests will be conducted with large scale imagery, color imagery, and degraded imagery, resulting, e.g., from poor scanning or lossy image compression schemes like JPEG. Investigations will also be carried out in order to find out the sensitivity of the algorithm to inaccurate values for overlap.

Acknowledgments
This work would not have been possible without the help of many people. We want to explicitly thank Heinrich Ebner, Full Professor and Head of the Chair for Photogrammetry and Remote Sensing, Technical University Munich; our colleagues and former colleagues Günter Strunz, Timm Ohlhof, Olaf Hellwich, and Konrad Eder; and student co-worker Joachim Batscheider. Part of the reported research was funded by the Deutsche Agentur für Raumfahrtangelegenheiten (DARA) under project grant 50 QM 9205, while other parts were funded by Carl Zeiss, Oberkochen, Germany. We would like to express our appreciation for this support.
Table 3. Description of and Results for All Test Areas

<table>
<thead>
<tr>
<th>Project</th>
<th>Description of Terrain</th>
<th>Image Scale</th>
<th>No. of Conjugate Points</th>
<th>( \sigma_a ) [m]</th>
<th>( \sigma_r ) [Pixel]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayer. Zell</td>
<td>mountainous, forested</td>
<td>1:34,000</td>
<td>220</td>
<td>3.6</td>
<td>0.24</td>
</tr>
<tr>
<td>Bopfingen</td>
<td>hilly</td>
<td>1:25,000</td>
<td>183</td>
<td>3.3</td>
<td>0.22</td>
</tr>
<tr>
<td>CZ1</td>
<td>agricultural</td>
<td>1:12,000</td>
<td>206</td>
<td>3.4</td>
<td>0.21</td>
</tr>
<tr>
<td>Eichstätt</td>
<td>steep, agricultural, partly industrial</td>
<td>1:17,000</td>
<td>713</td>
<td>3.2</td>
<td>0.22</td>
</tr>
<tr>
<td>Finland</td>
<td>flat, agricultural, forested</td>
<td>1:15,000</td>
<td>603</td>
<td>3.4</td>
<td>0.23</td>
</tr>
<tr>
<td>Marchetsreut</td>
<td>undulating, agricultural</td>
<td>1:15,000</td>
<td>977</td>
<td>3.3</td>
<td>0.23</td>
</tr>
<tr>
<td>Norway</td>
<td>hilly, coastal, agricultural</td>
<td>1:6,000</td>
<td>686</td>
<td>3.3</td>
<td>0.22</td>
</tr>
<tr>
<td>Pomona</td>
<td>densely built-up</td>
<td>1:25,000</td>
<td>457</td>
<td>3.5</td>
<td>0.23</td>
</tr>
<tr>
<td>Tasmanian</td>
<td>hilly, residential</td>
<td>1:12,000</td>
<td>962</td>
<td>3.5</td>
<td>0.23</td>
</tr>
<tr>
<td>Weilheim</td>
<td>steep, open</td>
<td>1:3,000</td>
<td>260</td>
<td>3.2</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Figure 8. Conjugate points from the original imagery (level 0), example Marchetsreut.

References


(Received 15 March 1995; accepted 20 May 1994)