The OEEPE-ISPRS test „Performance of tie point extraction in automatic aerial triangulation" - Status and first results

Christian Heipke, Konrad Eder, Rüdiger Brand, Roland Winkler
Chair for Photogrammetry and Remote Sensing
Technische Universität München
D-80290 München
Germany
email: {chris,konni,ruediger,pinki}@photo.verm.tu-muenchen.de

Commission II, Working Group II/8

Key words: AAT, automatic, multi-site test, practical investigation

Abstract

The European Organisation for Experimental Photogrammetric Research (OEEPE) and the International Society for Photogrammetry and Remote Sensing (ISPRS) have launched a common test on the performance of tie point extraction in automatic aerial triangulation. The aims of the test are to investigate the geometrical block stability, the accuracy of the tie points and the derived orientation parameters, and the limitations of existing commercial and experimental software systems. In order to separate the essentially new aspect of digital processing, namely automation, from conventional issues of aerial triangulation, control information is not assessed, and the test blocks to be processed have an arbitrary block datum.
The Chair for Photogrammetry and Remote Sensing, Technische Universität München acts as pilot centre of the test. In early 1997 various small blocks of different scene content have been distributed to interested participants. Their task was to generate tie points in an automatic way. The results of 21 participants, including all major software vendors of AAT and users of their systems, have been analysed and are presented in this paper. Given a large number of tie points per image has been extracted, the blocks were found to mostly stable. Under good conditions (open, flat terrain) an accuracy for the tie points of 3-4 µm could be reached, while under less favourable conditions, result was 4-9 µm. These figures were found to be very similar for the different systems. In mountainous and forested areas, some participants had difficulties to produce acceptable results. Reliable self control is a feature missing in all systems to date. Also, it seems that considerable experience is required to properly run the systems

1 INTRODUCTION

Automatic aerial triangulation (AAT) has been an increasingly interesting topic of research and development in digital photogrammetry over the last couple of years (see Schenk 1997 for an excellent review of the subject). The two tasks of measuring the image coordinates of tie points and of computing the orientation parameters, which were well separated in analytical photogrammetry, are more and more being merged into a single task, carried out in a hierarchical fashion using image pyramids. At the same time a shift of focus concerning the results of aerial triangulation can be observed. While in earlier times point densification was the primary goal, currently the orientation parameters themselves are of growing importance, since they can be directly used for subsequent tasks such as DTM generation, orthoprojection or vector data capture.
Today, a number of AAT software systems with different degree of automation are commercially available, either as stand-alone packages or as part of a Digital Photogrammetric Workstation, and are being introduced into practice (de Venecia et al. 1996; Ackermann, Krzystek 1997; Tang et al. 1997). Recently, users have started to report on their experience with these systems (e.g. Kersten, O`Sullivan 1996; Kersten, Häring 1997). However, at present a comprehensive comparison between the systems and also with analytical aerial triangulation, does not exist. At the same time a number of questions remain open, from the theoretical side (multi-image matching vs. matching only two images at a time, area based vs. feature based matching, the influence of local image texture etc.) as well as from the practical side (how many tie points should be available per image, which degree of automation can be reached and what does it depend on, what is the effect of image compression, how to implement an efficient procedure for quality control etc.).
In this situation the European Organisation for Experimental Photogrammetric Research (OEEPE) and the International Society for Photogrammetry and Remote Sensing (ISPRS) launched a common test on the performance of tie point extraction in automatic aerial triangulation (Heipke, Eder 1996). „Tie point extraction" is meant to include the selection, transfer and image coordinate measurement of block tie points. The test is primarily aimed towards the commercial software development and the user community of AAT systems. The results are intended to serve as a guide for assessing the capabilities of available systems and to give some hints towards improving them.
It may be asked, why a test about the indirect determination of the image orientation parameters by means of aerial triangulation should be conducted in a time, when these parameters are more and more being measured directly using GPS and INS. Ultimately, such indirect methods might become obsolete, but it seems save to predict that they will remain to be used for some time to come (see also the results of the OEEPE test on GPS, Ackermann 1996).
This paper reports about the test status. The analysis of the results is still in progress. Therefore, only first results can be presented. The next chapter briefly outlines the test goals, the test organisation, and the used data sets. Chapter 3 contains a list of the test participants and shortly discusses some features of the used software systems. The analysis procedure is described in chapter 4. The test results along with discussions are presented in chapter 5. Conclusions and open issues are presented in chapter 6.

2 THE OEEPE-ISPRS TEST

2.1 Test goals

In preliminary discussions with a number of potential test participants, a significant interest was expressed in assessing and comparing available AAT algorithms and strategies in terms of the achievable accuracy of the conjugate points as well as for the orientation parameters, and in terms of the software limitations. An operational AAT test including a large number of images and issues related to ground control, while important in everyday practice, were generally considered to be less important. Also issues related to the development status of individual AAT programs (user interface, program stability, computing time etc.) were described as secondary in importance. Throughout the test tie point extraction was considered to be a totally **autonomous** process to be carried out without any user interaction. In particular, any interaction during the tie point generation process as well as manual editing of the automatically obtained results in order to improve the measurement precision, to eliminate blunders and/or to introduce new measurements in areas where the automatic process failed to determine tie points, was not allowed within the test. In this way the essentially new aspect of digital imagery, namely automation, could be investigated separately from the issues which basically remain constant in the transition from analytical to digital photogrammetry (control information, block configuration, accuracy propagation, etc.). The aims of the test were to investigate

- the geometrical stability of the resulting block,
- the accuracy of the image coordinates of the tie points, and
- the limitations of existing commercial and experimental software systems.

2.2 Test organisation

The test was set up as a multi-site comparative test. The Chair for Photogrammetry and Remote Sensing, Technische Universität München, acted as pilot centre which selected, prepared and distributed the test data, and subsequently collected and analysed the results.

The test participants received various sets of image data together with appropriate additional information (see section 2.3 for details). They were then asked to automatically extract as many tie points from the images as they thought appropriate using their experimental or commercial software. If possible a common set of control parameters for the individual programs was to be used. The resulting image coordinates were subsequently communicated back to the pilot centre together with a report detailing the hard- and software and the algorithm used, the workflow adopted, necessary human interaction before and after the actual matching process, computational times, and a general assessment of the obtained results and problems encountered. This information was finally analysed by the pilot centre (see chapter 4 for a description of the analysis).

2.3 Test data sets

Guidelines for the selection of the test data were

- the need for a representative test data set covering different standard applications in photogrammetry,
- small blocks/strips and manageable data volumes,
- a fair chance for success for existing AAT systems,
- use of high quality images and scanners only.

The first point inspired the use of different scene contents, topography, cameras, scales, film material, and overlap configurations. As far as image scales are concerned, preference was given to larger scales, because in these cases, potential matching problems due to occlusions and relief displacement are more pronounced. The second point led to the selection of blocks with 3x2 and 3x3 images, strips with 3 images and pixel sizes of 20-30 µm (although for some data sets higher resolution images were available). While operational problems cannot be detected with such small blocks, the geometrical block stability and the accuracy of the tie points can be assessed. Taking the third point into account imagery with different scale within the block, with large rotations and non-topographic imagery was excluded. As for the last point, only first generation film products were scanned and all employed scanners are especially designed for photogrammetric applications.

According to these guidelines four blocks and two strips had been selected as test data sets. Since one of them was only processed by very few participants the analysis has been postponed, and results are not yet available. Table 1 shows some general information about the other five data sets. Due to lack of space the images themselves are not depicted in this paper. A short description of each test data set follows:

Echallens: This scene near Lausanne, Switzerland is rather flat and shows mainly open terrain. The black and white images are rich in texture, the overlap configuration corresponds to the standard values of aerial photogrammetry. As the imagery was used in earlier tests on aerial triangulation (Kölbl 1983), a number of signalised points is visible. The negatives were scanned directly ensuring a high radiometric image quality. Computations with this block can be considered as a sort of base line test for the AAT system.

Kapellen: The imagery was taken over a coal mining area in Germany and is rather flat. Most of the scene contains residential houses with some open spaces, in some areas mining gear is visible. Again, the black and white negatives have been scanned directly. In most parts there is rich texture, however, on the right part of the two strips, rather dark and homogeneous areas exist, and the overall image quality is not very good.

Montserrat: The scene is partly covered with forest, is rather mountainous in the northern part and includes the city of Montserrat in Catalunya, Spain in the southern part of the block. The block is the only one of medium image scale. The black and white imagery was selected from a standard flight of the Institut Cartografic de Catalunya (ICC), Barcelona. The negatives were scanned and were converted to positives during scanning.

OSU: This scene shows part of the campus of The Ohio State University, Columbus, USA. It is predominantly flat, the depicted buildings are rather large, and mostly separated by a fair distance. In the central part of the block where all 9 images overlap, a number of tennis courts can be seen. This is significant, since these courts can pose problems to matching algorithms because of the highly repetitive texture. In the western part the Scioto River can be seen. The photo flight was carried out in early September, while the tree still had leaves. The film material is false colour infrared, and thus the image quality is not as high as for a panchromatic film. The red channel, corresponding to the infrared scene reflection was selected for the test.

München: This strip of large scale colour images depicts the city centre of Munich, Germany, a densely built-up area. The large building visible in the images 55, and 56 is the famous Frauenkirche. The red channel was selected for the test. The images were taken some time ago, however, scanning was carried out recently.

The test participants received the image data on DAT or Exabyte together with information on the camera calibration, initial values for the exterior orientation parameters accurate to about 50 m for the projection centre and 2 grad for the angles, and an average terrain height for each project. Ground control points (GCP) was not generally distributed, but a limited set of three points per project were available upon special request, because some AAT program systems need GCP as input.

3 TEST PARTICIPANTS AND SOFTWARE SYSTEMS

After announcing the test 39 interested groups requested the test data. 21 of them (more than 50 %) actually participated in the test, processed at least some of the test images, and sent back the results. Among those 21 groups four major commercial photogrammetric software providers of AAT (Carl Zeiss, Intergraph, inpho, LH Systems) were present, together with five national or regional mapping organisations, four private companies and three research institutes employing commercial

products. In addition five research institutes who have developed their own AAT software took part in the test.
Table 2 gives an overview of the participants, the employed software, and the processed test data sets. Four groups can be distinguished, namely users of the commercial systems HATS from LH Systems (de Venecia et a. 1996; 7 users), Match AT from inpho (Ackermann, Krzystek 1997; 5 users), and Phodis AT from Carl Zeiss (Tang et al. 1997; 4 users), and the five participants having developed their own software software (Honkavaara, Hoghoen 1996; Wang 1996; Brand, Heipke 1998; Forlani et al. 1998; Paszotta 1998). Table 2 is organised accordingly. Altogether more than 80 sets of image coordinates were actually processed in the test. Some of the received results contained obvious gross errors. After consultation with the participants these results were deleted. They are not shown in table 2.

Neither the commercial products nor the developments of the research institutes can be presented in detail in this paper. However, some aspects shall be mentioned. All approaches use image pyramids in order to solve the problem of obtaining initial values for the unknown orientation parameters. With the exception of the system developed at IPI, only points are used as matching primitives. Points are in most cases selected using the Förstner-Operator (Förstner 1991). Some differences in the approaches are:

- Some approaches include a full bundle adjustment together with the matching component. Examples are Match AT, and the developments of TUM.
- Some systems such as Match AT and the FGI development need ground control points as input for the computation.
- In some systems (Match AT is an example) a rough DTM can be generated along with the tie point coordinates.
- Some participants (e.g. DIIAR) have found that their system is very sensitive to the quality of the initial values of exterior orientation and have therefore changed the provided values prior to running their AAT software.
- Most approaches use a combination of feature based and area based matching. Often least squares matching is used in the final computations (an exception is the TUM development which at the current stage only relies on FBM). However, it is not always clear which technique is actually employed at which stage.
- In some approaches areas around the standard positions for tie points from analytical photogrammetry are searched for points to be matched (e.g. HATS, Match AT, FGI), other systems such as Phodis AT, the one from IPI and from DIIAR try to match points in the whole images, at least in the upper pyramid levels.
- Some approaches such as Match AT, TUM and that of OUAT provide for automatic blunder detection, others such as HATS use an interactive scheme. One user (Swissphoto) has developed its own automatic blunder detection around a commercial system (HATS in this case).
- Some systems are designed as autonomous systems without any operator control (such as Match AT and Phodis AT), other approaches are more flexible and usually call upon the operator in order to manually measure additional points or eliminate blunders such as HATS. It should be noted that this possibility was not used by the test participants.
- Some systems such as Match AT and HATS have a list of free parameters, sometimes collected in a parameter file, which can be used to tune the results. The effect of these parameters, however, is not always clearly documented. For other systems such as Phodis AT only a minimum number of parameters must be provided by the user. While most participants used a standard parameter set for all test images, some did optimise the values in order to achieve a better result.

Given these numerous differences in the approaches it is virtually impossible within this test to link a certain result to a particular design feature. What makes the situation more complicated is the fact that different participants used different versions of the same software (see table 2). Nevertheless, as will be seen the results show some distinct trends.

4 ANALYSIS PROCEDURE

Usually, nine tie points per image are measured in analytical aerial triangulation. When evaluating the results, one (very tedious) way consists in revisiting and checking each observation. AAT systems, on the other hand, can deliver a few hundred tie points per image, and sometimes more than 1000. Given the expected number of test participants revisiting and checking each observation was considered impossible within the test. Therefore, alternative ways of evaluating the results were developed.

The analysis procedure consists of different steps. In the **first step** for each received set of image coordinates a robust bundle adjustment was carried out at the pilot centre. The image coordinates were assumed to be uncorrelated and of an accuracy of $\sigma_{o,a\ priori}$ = 1/3 of a pixel, excluding the blunders. The value chosen for $\sigma_{o,a\ priori}$ is important for the analysis, since it influences the results of the robust adjustment. 1/3 of a pixel is the accuracy generally attributed to digital image matching and was therefore selected here. In further tests this value will be varied as appropriate in order to find out whether a higher accuracy for the tie points is obtainable.
The block datum was fixed by introducing the minimum of seven orientation parameters (six parameters of one image and one base line) as constant values. In this way it could be ensured that the resulting block would not be influenced by ground control information. Rather, the potential of the purely automatic tie point extraction could be assessed. For the bundle adjustment the program package CLIC developed at the pilot centre over the last 15 years was used. In this way a number of results were obtained:

- the number and percentage of detected blunders.
- the average number of tie points per image,
- the number of multi-ray points,
- plots showing the distribution of the tie points in image and in object space, before and after the robust bundle adjustment,
- plots showing the distribution of the points eliminated during the robust adjustment,
- plots showing the distribution of only those tie points which connect strips.

This information was used in order to obtain a first impression of the quality of the received results. Additional information obtained from the robust bundle adjustment consisted in the standard deviation σ_o of the image coordinates, the exterior orientation parameters for each image (to be used in the second analysis step described below), and the covariance matrix Σ of the computed orientation parameters.
Within automatic tie point extraction, care has to be taken that the tie points are well distributed, and the images and strips are appropriately connected with enough multi-ray points. If checks to this end are not employed by the program system a block with rather weak geometric stability can be the result. Such a weak block geometry cannot be detected using σ_o as an indicator. However, it leads to an unfavourable covariance matrix Σ for the orientation parameters. An analysis of Σ, e.g. based on criterion matrices (Baarda 1973; Förstner 1995) thus can give insight into the block stability. Such an analysis is being carried out for this test, but the results are not yet available.

A **second analysis step** was carried out for each set of image coordinates in order to independently assess the accuracy of the obtained orientation parameters. At the pilot centre image coordinates of tie points were measured interactively in the test images, and a classical aerial triangulation was computed. In four of the five projects the measurements were carried out using the digital images, in two cases (Montserrat and München) with images of higher resolution than those used in the test. Only for OSU the analogue images were used for the reference measurements, because in this way the tie points could be better identified. The results of the reference measurements are shown in table 3. Also the quality of the interior orientation from the digital images in terms of the standard deviation $\sigma_{o,int}$ of one fiducial measurement after an affine transformation between pixel and image coordinates is given. $\sigma_{o,int}$ is interesting, because it contains possible film deformations and deformations due to geometric scanner errors. The results for Echallens and München can be considered excellent, those for Kapellen and Montserrat still agree with the expectations. A value for $\sigma_{o,int}$ was also determined for the digital OSU images and amounted to approximately 13.6 µm or 0.54 pixel. This value is rather large and an indication for problems with the geometry of the digital images. Since this effect was not observed in the film images the large $\sigma_{o,int}$ of OSU is an indication that the used scanner was not sufficiently well calibrated (see also further discussion in chapter 5).

The quality of the reference measurements is summarised in the standard deviation $\sigma_{o,i}$ of the bundle adjustment. Also, these figures fulfil the expectations. The best results were reached for Echallens (0.16 pixel or 3.2 µm). The reason is that many of the signalised points were used in the interactive measurements, and additional parameters could be significantly determined from the corresponding ground control coordinates. For Kapellen and Montserrat the results (0.25 and 0.23 pixels or 6.1 and 3.4 µm, respectively) reflect the attainable accuracy for interactive measurements with digital images using natural tie points. The same holds for the OSU result obtained from the analogue FIR images. The value for München is somewhat larger (0.33 pixels or 4.9 µm), possibly due to the age of the images. For these images the calibration protocol does not show image coordinates of the fiducial marks, and thus standard values had to be used.

In the following the interactively measured image coordinates serve as reference for the results of the participants. Subsequently, forward intersections were computed from these coordinates, introducing the exterior orientation parameters from the participants obtained in the robust adjustment of the first analysis step as constant values. For Echallens, the additional parameters were introduced as well. This computation also results in a value for the accuracy of the image coordinates termed σ_{FI} for "forward intersection". σ_{FI} can be considered as a measure of quality for the orientation parameters determined from the image coordinates of the participants. Since $\sigma_{o,i}$ represents the optimal accuracy for the image coordinates with the given measurements (in the least squares sense), σ_{FI} cannot be smaller than $\sigma_{o,i}$.

When comparing σ_{FI}, $\sigma_{o,i}$, and σ_o (the latter from the robust bundle adjustment of the first analysis step) two different cases should be distinguished:

- $\sigma_{o,i}$ and σ_{FI} are of the same magnitude. In this case the exterior orientation parameters computed from the participants´ results coincide with those from the interactive measurements. σ_o can be larger, in the order of, or smaller than $\sigma_{o,i}$ and can be considered as the accuracy of the automatically determined tie points.
- σ_{FI} is considerably larger than $\sigma_{o,i}$. In this case deformations in the blocks of the participants are assumed, regardless of the value of σ_o. A small value for σ_o (possibly smaller than $\sigma_{o,i}$) can indicate that the distribution of the tie points is not appropriate and thus the images and/or strips are not appropriately connected.

These arguments assume that the transformations from pixel to image coordinates are identical for the results of the participants and the interactively measured image coordinates. Preliminary studies suggest that the remaining differences between the two transformations do not influence the interpretation of σ_{FI}. They can, however, account for differences between $\sigma_{o,i}$ and σ_{FI} in the order of 1-2 µm.

5 RESULTS AND DISCUSSION

The results are presented in the form of tables. Table 4 contains the results for Echallens: the number and the percentage of the eliminated blunders, the average number of tie points per image, and the total number and the number of the multi-ray points in object space as well as the accuracy figures σ_o (both in pixels and in µm) and σ_{FI}.

Echallens is the base line data set for the test, as mentioned above. Concentrating first on these results a number of observations can be made:

- It can be seen (and comes at no surprise) that within AAT a robust adjustment is absolutely necessary. In the systems which do not include a blunder detection scheme up to 22% of the measurements needed to be eliminated.
- The number of observations per image, the average number of detected blunders, the remaining number of observations per image and the resulting number of object points differ considerably between the participants and systems. Whereas UNSW used only 12 points per image and 41 points in object space TUM delivered 468 points per image and 1591 object points. Most participants generated an average of between 60 and 250 tie points per image.
- A closer look at the number of rays per object point reveals that only Match AT and the TUM and FGI developments seem to focus on obtaining a large number of multi-ray points. HATS can apparently be tuned in the same way, see the SWPH result.
- The standard deviation σ_o of the tie point coordinates generally lies between 0.15 and 0.20 pixels or 3 and 4 µm with the exception of the TUM development which in the version used for the test relies uniquely on FBM. The independent check using σ_{FI} confirms this block accuracy. Due to the large number of tie points and the resulting high redundancy, the σ_{FI} value for TUM is also in this range. Exceptions exist where either only a small amount of tie points was measured (BKG, NLS-SWE, UNSW) or very few multi-ray points were extracted (DIIAR). The first group of participants all used HATS. As mentioned above HATS calls upon the operator in case of problems, but this feature was deliberately ignored within the test, since only automatic results were to be generated. As exemplified by the results from LHS and SWPH HATS can be turned into a fully automatic system. A prerequisite, however, seems to be a large number of observations. Otherwise an appropriate distribution of the points across the images cannot be guaranteed and the resulting block can be severely deformed. In these cases σ_o cannot be used as an indicator for the block stability as mentioned before.
  As for the DIIAR result it should be noted that the system was still under development while the test data were processed.

Montserrat (see results in table 5) is a more difficult data set. The scene is more mountainous and contains forest, especially in the mountainous area in the upper part of the scene between the first and the second strip, leading to rather unfavourable conditions for image matching. As for the descriptive results it can be seen that the average number of tie points per image is somewhat larger than for Echallens, especially for the Phodis AT results. The distribution of multi-ray points is similar. Also the OUAT approach seems to focus on obtaining as many six-ray points as possible.

Different conclusions must be drawn from the accuracy figures of Montserrat. While the σ_o column seems to suggest homogeneous results an inspection of σ_{FI} reveals the opposite. In a number of cases σ_{FI} is significantly larger than σ_o. As already found for Echallens, σ_o alone cannot be used to characterise the quality of the block. Blocks generated from rather few tie points (BKG, NLS-SF, NLS-SWE, UNSW) or from an overwhelming number of 2-ray points (DIIAR) were again found to be deformed. But the discrepancy between σ_o and σ_{FI} also exist in other cases (EPFL, inpho, HL, OUAT). Analysing the results of the forward intersection for these four participants, it was found that points in the overlapping area between the first and the second image strip showed large residuals in the flight direction. The same phenomenon, although less pronounced, was subsequently also observed in some of the other results. Due to the nature of the test, reasons for these large residuals cannot be determined unambiguously. The interior orientation is a possible source of error, especially since in contrast to Echallens the flight direction between the first strip the other two differs by 200 grad. Differences in the interior orientation might explain the inpho result, since inpho had also measured tie points interactively which were found to be consistent with their automatic result, but inconsistent with the manual measurements of the pilot centre. Also, the quality of the reference measurements was considered not to be beyond any doubt, although the good results suggest the reference measurements are indeed correct. In order to double check them, the complete reference measurement was repeated. An error could not be detected.

The most plausible explanation seems to be that for the participants exhibiting large residuals in the forward intersection, the first and the second strip are not correctly connected, probably due to the unfavourable matching conditions.

Despite these problems it should also be mentioned that the participants who delivered a correct result reached an accuracy of around 0.2 pixels or 6 µm.

The results for OSU (see table 6) add some more insight :

- The number of eliminated blunders is considerably larger. For the Phodis AT results it reaches 31% and for DIIAR 45%. Also, in the systems with blunder detection some additional gross errors needed to be eliminated. An analysis of the blunder distribution revealed that most of them were related to the tennis courts (see chapter 2) and thus to the problem of repetitive texture. In all but one data set (DIIAR) the number and distribution of the remaining observations were sufficient to generate a stable block, as can be seen from the σ_{FI} values.
- The lack of multi-ray points is very much apparent. With the exception of the results from inpho and TUM especially the number of 7-ray, 8-ray, and 9-ray points is very small. The relatively large amount of 4- ray and 5-ray points and the resulting high redundancy in the block adjustment seems to compensate for this lack. The DIIAR result, however, shows the limits of such a compensation.
- The accuracy level is somewhat worse than for Echallens and Montserrat. This is true for σ_o and also for σ_{FI}. The reasons are probably twofold: first, the FIR film material does not show the same image quality as the panchromatic material used for Echallens and Montserrat. Second, the scanner used for the OSU images was apparently somewhat decalibrated when the images were scanned. This can be deduced from the large $\sigma_{o,int}$ (see chapter 2). Thus, only an accuracy of approximately 0.3 pixels or 7 – 8 µm was reached.
- As already found earlier if too few observations are generated the resulting block can be severely deformed, which is again visible when comparing σ_o and σ_{FI} (see results from NLS-SWE, UNSW, and to some extent EPFL).

The results of Kapellen and München (tables 7 and 8) generally confirm the previously discussed results. The σ_o values are in the same range as for OSU (around 0.3 pixel) and thus higher than for Echallens and Montserrat. The reason is probably the somewhat lower image quality as judged from visual inspection. In the Kapellen data set, this is mainly visible in the right part of the scene. As for München, besides the fact that no image coordinates for the fiducials were available, it must also be kept in mind that due to the high buildings in the scene and the large scale of the images perspective deformations play a much greater role than for the other data sets.

6 CONCLUSIONS

Compared to the test goals (see section 2) the following conclusions can be drawn (it should be emphasised again, that we consider the point extraction to be a totally **autonomous** process within AAT):

- A good **geometric block stability** can be guaranteed, if and only if a sufficiently large number of tie points (say 100 to 300 per image) is measured. The reason is that local matching procedures, as they are employed in the tested systems in order to achieve an acceptable level of accuracy, are subject to a large number of blunders which need to be eliminated at a later stage. Robust block adjustment is an appropriate tool for this task, although in some cases the amount of blunders was found to be extremely high. If too few points are measured the resulting block can be heavily deformed. Within the test this problem occurred mainly for results generated with HATS. As mentioned before, that HATS calls upon the operator if points are missing or need to be remeasured. This feature was not used in the test.
- Especially in larger blocks the stability also depends on the number and distribution of the available GCP and/or the quality of the direct measurements for the orientation parameters from GPS and/or INS. Such information can lead to a somewhat reduced number of necessary tie points per image, especially when it can be used as input for the matching procedure. As mentioned before, however, no such effects were investigated within the test.
- The high redundancy in the adjustment leads to a smaller theoretical standard deviation and an improved reliability for the exterior orientation parameters as compared to analytical photogrammetry. These parameters, of course, must be regarded as the prime result of AAT. This point will be further analysed within the test in order to quantitatively demonstrate the advantages of AAT.
- While the significance of a large number of multi-ray points is not as high as in analytical photogrammetry neglecting this aspect too much can also lead to severe block deformations. In the test all commercial systems generated enough multi-ray points, but it seems save to predict that more emphasise should be concentrated on this point, in particular within the Phodis AT system.
- Under favourable conditions (open and flat terrain, good texture; see Echallens) the **accuracy of the tie point coordinates** can reach 0.15-0.2 pixels or 3-4 µm using only

natural tie points if least squares matching is employed for coordinate refinement. In analytical photogrammetry a comparable accuracy has only been achieved using signalised points.

- Taking all test results into account a realistic values for σ_0 lies in the range of 0.2-0.3 pixels or 4-9 μm (again with only natural tie points and least squares matching), at least when the images were scanned with a pixel size of 20-30 μm. The values are rather similar across the different systems. Since most of them use least square matching in the final coordinate measurement this result seems plausible. In this test the effect of pixel size was not separately investigated. Experience and the literature suggest that smaller pixel sizes will not increase the accuracy of the tie points accordingly.
- **Limitations of existing systems** showed up in the Montserrat example. Some participants failed to produce correct and accurate results. The strip connection seems to be the weak point, at least in mountainous and forested terrain.
- Failure to produce an acceptable result is not indicated by the systems (with the partial exception of HATS, see above), because internal self control is not sufficiently accounted for. Elements of self control are the individual matching results, the distribution of the tie points and the number of multi-ray points within the block, the measurement accuracy, and the covariance matrix of the unknowns. As was shown in a number of cases the σ_0 of the block adjustment is by itself not a valid indicator of errors or deformations within the block. It seems indispensable to combine matching and bundle adjustment in order to realise a reliable self control mechanism. While this combination has been realised for Match AT and HATS, it is missing in the Phodis AT design.
- Due to the large amount of required observations (see above) the self control mechanism needs to be automatic. A human operator is simply not able to revisit or remeasure individual tie points, if he/she is to work at an economical pace.
- It is interesting that both, success and failure, in the Montserrat example occurred partly with one and the same system. This suggests that an extensive amount of experience in handling the software is necessary in order to appropriately tune any available free parameters. Taking also the results into account which due to gross errors are not contained in the presented tables this experience seems to be especially necessary for using Match AT and for HATS. If the number of free parameters cannot be significantly reduced additional effort should be focused on training of the AAT operators.

As already mentioned in the introduction within this test not all topics related to a complete system analysis were investigated. For instance, issues related to an economical use (e.g. the time and cost needed for preparation, computation, and post processing) have not been considered. Furthermore, the behaviour of AAT systems for larger and non-regular blocks and the influence of control information was outside the scope of the investigations. From the obtained results, it can be concluded that the current AAT systems after only a few years of market presence, show a remarkable level of performance. A number of details, however, need further refinement. In summary, it can be predicted, that in a production environment fully autonomous tie point extraction while feasible in many cases, will be followed by a verification and editing stage carried out by a human operator. Software development should be concentrated on creating more reliable self control mechanisms and on designing user friendly interfaces for an efficient verification and editing of the AAT results including a stereo measurement capability for high accuracy requirements.

ACKNOWLEDGEMENTS

We wish to express our gratitude to OEEPE and ISPRS for supporting the described test, and in particular to Prof. Ackermann and Prof. Förstner for initiating the test. We are also grateful to the organisations and individuals having provided the test data. Special thanks go to Prof. Kölbl. At his institute the first author was able to spend a sabbatical in which much of the test analysis could be carried out. Finally, this test would not have been possible without the enthusiasm of the test participants. We were very surprised and exited by the large interest in the test and by the number of groups participating. They all put a considerable effort into processing the data sets while keeping up with a rather tight time schedule. More often than not processing had to be done in addition to the every day work load. We hope that we could fulfil their expectations.

REFERENCES

Ackermann F., 1996: Experimental tests on fast ambiguity solutions for airborne kinematic GPS positioning, IntArchPhRS (31) B6, 3rd section, 1-6.

Ackermann F., Krzystek P., 1997: Complete automation of digital aerial triangulation, Photogrammetric Record (15) 89, 645-656.

Baarda W., 1973: S-transformations and criterion matrices, Netherlands Geodetic Commission New Series (5) 1, Delft.

Brand R., Heipke C., 1998: A system for automatic aerial triangulation, IntArchPhRS (32) 2.

Förstner W., 1991: Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und –vermessung, DGK C, 370.

Förstner W., 1995: A metric for comparing symmetric positive definite matrices, Note, Institute for Photogrammetry, Bonn University.

Forlani G., Pinto L., Scaioni M., 1998: TRIADIGIT: one more AAT program, IntArchPhRS (32) 2.

Heipke C., Eder K., 1996: Performance of tie point extraction in automatic aerial triangulation, description of the OEEPE-ISPRS test, Technische Universität München.

Honkavaara E., Hogholen A, 1996: Automatic tie point measurement in aerial triangulation, IntArchPhRS (31) B3, 337-342.

Kersten T., Häring S., 1997: Efficient automated digital aerial triangulaion through customisation of a commercial photogrammetric system, IntArchPhRS (32) 3-2W3, 72-79.

Kersten T., O´Sullivan W., 1996: Experiences with the Helava Automated Triangulation System, IntArchPhRS (31) B3, 591-596.

Kölbl O., 1983: Augmentation de la performance de la photogrammétrie en mensuration cadastrale, Vermessung, Photogrammetrie, Kulturtechnik (81) 1, 1-8.

Paszotta Z. 1998: Report on the OEEPE-ISPRS test on "Performance of tie point extraction in automatic aerial triangulation", unpublished.

Schenk T., 1997: Towards automatic aerial triangulation, ISPRS Journal of Photogrammetry & Remote Sensing (52) 3, 110-121.

Tang L., Braun J., Debitsch R., 1997: Automatic aerial triangulation – concept, realisation and results, , ISPRS Journal of Photogrammetry & Remote Sensing (52) 3, 122-131.

de Venecia K., Miller S., Pacey R., Walker S. 1997: Experiences with a commercial package for automated aerial triangulation, ASPRS/ACSM Annual Convention, (1) 548-557.

Wang Y. 1996: Structural matching and its applications for photogrammetric automation, IntArchPhRS (31) B3, 918-923.

Project name	Echallens	Kapellen	Montserrat	München	OSU
Scene content	open, partly forest	settlement, partly open	forest, partly built-up	city centre	built-up, partly trees
Scene topography	flat	flat	hilly	buildings	flat, buildings
Image scale	1:5.000	1:4.000	1:15.000	1:2.000	1:4.000
Camera	Wild RC 10	Zeiss RMK A	Zeiss RMK TOP	Zeiss RMK A	Wild RC 10
Focal length [mm]	150	150	150	300	150
Flight datum	September 1982	April 1992	May 1995	May 1975	September 1995
Film material	black and white	black and white	black and white	colour	FIR
Number of images	3 x 3	2 x 3	3 x 3	3	3 x 3
Overlap	l = 60 %, q = 30 %	l = 60 %, q = 60 %	l = 60 %, q = 30 %	l = 60 %	l = 60 %, q = 60 %
Scanner used	LH DSW 200	Wehrli RM1	Zeiss PS1	Zeiss PS1	LH DSW 200
Pixel size for test	20 µm	24 µm	30 µm	30 µm	25 µm
Scanned material	negative, original	negative, original	negative, original	positive, original	positive, original
Scanned channel	pan	pan	pan	red	red (= infrared)
Scan datum	January 1996	June 1996	November 1996	December 1996	October 1995
Source	EPFL, Lausanne	Hannover University	ICC, Barcelona	Technische Universität (TU) München	The Ohio State University / TU München

Table 1: Description of the test data sets

Full name and abbreviation of participant		Software and Version No.	Echallens	Kapellen	Montserrat	München	OSU
LH Systems, San Diego	LHS	HATS, 3.2.1.1	X	X	X	X	
Bundesamt für Geod. u. Kart., Frankfurt/M.	BKG	HATS, 3.1.1.2	X	X	X	X	
Institute for Photogrammetry, EPFL Lausanne	EPFL	HATS, 3.1.3k	X		X		X
National Land Survey of Finland, Helsinki	NLS-SF	HATS, 3.2.1.2			X		
National Land Survey of Sweden, Gävle	NLS-SWE	HATS, 4.0.8	X	X	X		X
School of Geomatics, UNSW, Sydney	UNSW	HATS, 3.2.1	X		X	X	X
Swissphoto, Regensdorf	SWPH	HATS, 3.2.1.2*	X	X	X	X	X
Inpho GmbH, Stuttgart	Inpho	Match AT, 2.1.0	X	X	X	X	X
Intergraph, Huntsville	I-graph	Match AT, 2.1.1	X	X	X	X	X
Compagnia Generale Ripreseaeree, Parma	CGR	Match AT, 2.1.1			X	X	X
Hansa Luftbild, Münster	HL	Match AT, 2.1.1			X		
Photogrammetrie GmbH, München	Ph GmbH	Match AT, 2.1.1			X	X	
Carl Zeiss, Oberkochen	CZ	Phodis AT, 2.0.1	X	X	X	X	X
Bayerisches Landesvermessungsamt, München	B-LVA	Phodis AT, 2.0.0	X	X	X	X	X
General Command of Mapping, Ankara	GCM	Phodis AT, 2.0.0		X	X	X	X
Landesvermessung + Geobasisdaten, Hannover	LGN	Phodis AT, 2.0.0	X		X		
Chair for Photogram. & Rem. Sensing, TU München	TUM	own dev.	X			X	X
Dip. Ing. e Idraul. Amb. e del Rilev., Politec. di Milano	DIIAR	own dev.	X		X	X	X
Finnish Geodetic Institute, Masala	FGI	own dev.	X		X		
Institute of Photogrammetry, Hannover	IPI	own dev.		X		X	X
Chair of Ph & RS, Olsztyn Univ. of Agricul. a. Techn.	OUAT	own dev.			X		

Table 2: List of test participants (*: SWPH combined HATS with customised software, see Kersten, Häring 1997)

Project name	Echallens	Kapellen	Montserrat	München	OSU
Image scale	1:5.000	1:4.000	1:15.000	1:2.000	1:4.000
Av. flying height	850 m	600 m	2350 m	650 m	600 m
Pixel size for ref. meas.	20 µm	24 µm	15 µm	15 µm	(analogue)
No. of fiducials	4	4	8	4	4
$\sigma_{o,int}$ of dig. images	4.8 µm / 0.24 pel	8.5 µm / 0.35 pel	7.3 µm / 0.48 pel	3.4 µm / 0.23 pel	(13.6 µm / 0.54 pel)
Av. no. of tie pts. p. image	41	55	65	58	27
$\sigma_{o,i}$ of bundle adjustment	3.2 µm / 0.16 pel	6.1 µm / 0.25 pel	3.4 µm / 0.23 pel	4.9 µm / 0.33 pel	4.8 µm

Table 3: Results of interactive reference measurements

Participant	elim. blunders		av. no. of tie pts. per image	no. of multi ray points in object space						σ_o		σ_{FI}
	no.	%		total	2 ray pts.	3 ray pts.	4 ray pts.	5 ray pts.	6 ray pts.	[pel]	[µm]	[µm]
LHS	86	14	59	195	99	63	21	5	7	0.21	4.2	4.5
BKG	24	16	14	48	31	6	7	2	2	0.14	2.7	8.2
EPFL	65	22	25	88	57	16	11	2	2	0.18	3.6	4.8
NLS-SWE	15	6	26	93	63	17	6	4	3	0.18	3.5	6.0
UNSW	6	5	12	41	22	9	8	1	1	0.15	3.0	5.5
SWPH	0	0	73	235	128	55	33	4	15	0.17	3.4	3.9
Inpho	0	0	182	496	180	123	115	11	67	0.17	3.3	3.7
I-graph	0	0	98	327	194	70	44	0	19	0.18	3.6	4.9
CZ	179	8	250	906	538	318	35	10	5	0.20	4.0	4.7
B-LVA	208	9	245	895	549	293	42	6	5	0.18	3.5	4.6
LGN	234	9	275	988	555	381	41	7	4	0.17	3.4	4.5
TUM	46	1	468	1591	839	534	163	43	12	0.33	6.6	4.2
DIIAR	200	16	115	479	404	70	4	1	0	0.28	5.6	6.2
FGI	0	0	379	1433	1097	175	133	6	22	0.17	3.4	4.4

Table 4: Results for test data set Echallens

Participant	elim. blunders		av. no. of tie pts. per image	no. of multi ray points in object space						σ_o		σ_{FI}
	no.	%		total	2 ray pts.	3 ray pts.	4 ray pts.	5 ray pts.	6 ray pts.	[pel]	[µm]	[µm]
LHS	43	7	62	209	125	44	26	8	6	0.19	5.8	4.9
BKG	10	6	18	66	48	9	5	3	1	0.10	3.1	9.7
EPFL	60	12	49	168	103	36	19	6	4	0.20	6.0	13.4
NLS-SF	21	12	17	60	36	16	6	1	1	0.22	6.5	12.2
NLS-SWE	23	10	22	81	56	16	6	2	1	0.25	7.4	18.3
UNSW	18	17	10	32	17	8	4	2	1	0.14	4.3	17.6
SWPH	0	0	69	243	165	40	26	2	10	0.21	6.4	5.4
Inpho	0	0	184	574	265	182	82	13	32	0.11	3.3	11.4
I-graph	0	0	148	508	286	154	49	11	8	0.20	6.0	7.2
CGR	0	0	160	550	334	138	54	2	22	0.14	4.3	5.9
HL	0	0	102	337	182	89	51	5	10	0.15	4.6	10.6
Ph GmbH	0	0	98	352	227	76	41	5	3	0.17	5.2	6.3
CZ	371	10	358	1315	824	413	56	15	7	0.22	6.7	6.4
B-LVA	373	12	330	1245	841	335	58	7	4	0.21	6.2	5.0
GCM	573	11	495	1969	1523	384	51	7	4	0.19	5.7	5.2
LGN	429	12	349	1307	849	396	52	6	4	0.20	5.9	4.4
DIIAR	354	24	123	524	475	39	9	1	0	0.25	7.4	20.1
FGI	0	0	395	1506	1112	286	74	26	8	0.18	5.4	5.5
OUAT	0	0	147	493	285	122	62	0	24	0.25	7.4	13.6

Table 5: Results for test data set Montserrat

Participant	elim. blunders		av. no. of tie pts. per image	no. of multi ray points in object space									σ_o		σ_{FI} [µm]
	no.	%		total	2 ray pts.	3 ray pts.	4 ray pts.	5 ray pts.	6 ray pts.	7 ray pts.	8 ray pts.	9 ray pts.	[pel]	[µm]	
EPFL	28	12	22	67	34	14	9	3	6	1	0	0	0.28	7.0	9.3
NLS-SWE	66	27	19	61	34	13	6	4	3	0	1	0	0.27	6.7	10.8
UNSW	32	28	9	26	13	5	4	2	1	0	0	1	0.25	6.2	21.5
SWPH	42	7	61	184	97	34	29	13	7	2	1	1	0.28	6.9	7.6
Inpho	126	5	242	631	234	150	123	38	60	3	6	17	0.30	7.6	8.4
I-graph	94	8	121	398	235	90	42	7	22	0	0	2	0.30	7.5	7.2
CGR	176	9	195	613	324	146	89	17	31	2	1	3	0.30	7.4	7.5
CZ	961	30	247	906	591	230	70	11	4	0	0	0	0.33	8.2	8.2
B-LVA	981	31	247	909	607	221	63	14	3	0	1	0	0.31	7.7	7.6
GCM	1281	30	337	1319	1021	223	58	11	5	1	0	0	0.31	7.7	8.8
TUM	127	4	361	1055	485	262	160	76	46	7	15	4	0.37	9.3	9.7
DIIAR	392	45	53	220	186	31	3	0	0	0	0	0	0.33	8.3	22.7
IPI	1169	27	351	1429	1242	123	31	17	12	2	1	1	0.28	7.1	8.1

Table 6: Results for test data set OSU

Participant	elim. blunders		av. no. of tie pts. per image	no. of multi ray points in object space						σ_o		σ_{FI}
	no.	%		total	2 ray pts.	3 ray pts.	4 ray pts.	5 ray pts.	6 ray pts.	[pel]	[μm]	[μm]
LHS	113	25	57	128	76	26	18	7	1	0.29	7.0	8.7
BKG	49	38	13	33	25	3	4	1	0	0.23	5.5	27.4
NLS-SWE	40	20	26	58	31	17	7	1	2	0.25	6.1	12.4
SWPH	31	8	56	120	64	22	29	2	3	0.25	6.0	8.1
Inpho	0	0	181	353	148	79	94	11	21	0.33	7.8	8.0
I-graph	58	8	113	251	133	75	33	2	8	0.30	7.2	7.7
CZ	467	28	195	492	345	108	37	2	0	0.36	8.3	9.7
B-LVA	536	32	188	479	342	105	30	1	1	0.31	7.5	9.4
GCM	726	34	236	632	515	83	30	4	0	0.29	7.0	8.4
IPI	483	40	118	347	331	13	3	0	0	0.28	6.7	12.6

Table 7: Results for test data set Kapellen

Participant	elim. blunders		av. no. of tie pts. per image	no. of multi ray points in object space			σ_o		σ_{FI} [μm]
	no.	%		total	2 ray pts.	3 ray pts.	[pel]	[μm]	
LHS	35	21	45	63	54	9	0.23	7.0	8.6
BKG	8	18	12	17	15	2	0.16	4.8	24.9
UNSW	4	14	8	11	8	3	0.22	6.5	28.9
SWPH	0	0	71	104	99	5	0.28	8.4	7.6
Inpho	0	0	159	199	118	81	0.26	7.9	7.0
I-graph	0	0	147	189	126	63	0.28	8.3	8.1
CGR	34	11	90	128	112	16	0.22	6.6	8.2
Ph GmbH	15	9	50	64	41	23	0.24	7.3	8.3
CZ	121	17	202	274	214	60	0.27	8.1	6.7
B-LVA	357	19	493	712	657	52	0.25	7.6	8.3
GCM	188	17	301	424	369	55	0.25	7.5	9.2
TUM	11	2	201	272	213	59	0.33	9.8	7.5
DIIAR	164	41	77	112	103	9	0.26	7.7	28.4
IPI	143	13	323	483	479	4	0.32	9.6	8.5

Table 8: Results for test data set München