Trinocular Vision for Automatic and Robust Three-Dimensional Determination of the Trajectories of Moving Objects*

Emmanuel P. Baltsavias and *Dirk Stallmann*
Institute of Geodesy and Photogrammetry, ETH - Hoenggerberg, 8093 Zurich, Switzerland

ABSTRACT: An automatic procedure for tracking signalized points on multiple, differently moving surfaces will be described. The practical tests involve tracking of points on a person walking within a calibration frame. Three simultaneously operating CCD cameras, which are oriented using control points, provide the input data. For implementation, the hardware (H/W) and software (S/W) of a digital photogrammetric system have been used. The procedure includes extraction of the signalized points and finding corresponding points in all three images, thus providing two-dimensional (2-D) and three-dimensional (3-D) coordinates and tracking using the information in both image and object space. Reliability aspects such as occlusions, multiple solutions, and different backgrounds will be analyzed. The effect of a third image on reliability will be analyzed.

INTRODUCTION

THIS RESEARCH is part of an on-going project with title "Automatic Object Tracking in Stereo Vision Systems," supported by the Swiss National Science Foundation (NF). It is a continuation of previous work on this topic (Novak *et al.*, 1990; Zhou, 1990). The aim was to localize and track signalized points in a rather complicated environment, before progressing with tracking of natural features. Some motives for the specific experiment were given by doctors who wanted to analyze the human motion for gait studies. However, the developed methods are not tied to any specific biomedical application. The task includes blob detection and localization, correspondence and derivation of three-dimensional (3-D) coordinates, image and object tracking. Each of these topics has extensive applications, exceeding by far the area of object tracking. There exist numerous references on these topics but due to lack of space they will not be mentioned, apart from those that were actually used.

The aim of this research was to gain experience on this topic, to develop new algorithms and test existing ones, and to determine failures, their causes, and possible solutions. The low-level processes involved in the blob detection were carefully analyzed, as they have a great influence on the results of the subsequent steps. The use of three cameras was intended to prove the superiority of this approach over mono or even stereo sequence processing. At this stage, no specific object motion model was used for tracking. However, integration of object motion models can significantly improve the results, as will be explained in the conclusions in more details.

EXPERIMENT SET-UP, DATA ACQUISITION, AND CAMERA CALIBRATION

The experiment aimed at tracking signalized points mounted on differently moving discontinuous surfaces (in this case a walking person; see Figure 1). The signalized points were 30 paper balls with a diameter of 2.6 cm, onto which stripes of retroreflective material (increasing the reflected light by a factor of 1000) were attached manually. The balls were slightly cut at one side to provide for a flat bases, and were attached to the person's clothing (hence the surfaces were also not rigid). The balls were mounted not only on the part of the body facing the cameras but also on the front, on the back, and on the left arm and leg. To provide control over the accuracy, the walking person held a ruler in his hand. Two retroreflective circular targets with a diameter of 2.5 cm were attached to the two end points of the aluminium ruler (approx 0.5 m in length). The distance between the centers of the two retroreflective targets was measured with a laser interferometer to an accuracy better than 0.05 mm. The person walked within a testfield calibration frame so that the CCDs could be calibrated. The occlusions caused by the calibration frame can be avoided in an operational set-up, but for this particular experiment they were actually useful as they provided a test of how the algorithm reacted to occlusions. The geometric configuration is shown in Figure 2. The three camera stations formed a triangle with equal sides (1.2 m) parallel to the movement of the person. The focal length of each camera was 8.5 mm, the average distance from the sensors to the object was 4.15 m, and the average scale was 1:485.

The illumination consisted of rather weak fluorescent strip lamps approximately centered behind each of the CCDs. The illumination could have been chosen such that, practically speaking, only the retroreflective targets were visible and everything else was black. This would permit a very simple, robust, and fast blob detection and localization. However, this possibility was intentionally not used as one of the experiment aims was to detect and track blobs in more complicated environments, using possibly natural features or weaker projected patterns. Another consideration was that, due to saturation of the blobs, their size and shape changed, leading to a less precise localization and a fusion of neighboring blobs. With the illumination and geometric configuration that have been used, the size, shape, and contrast of the blobs varied greatly (see Figure 3). Some blobs were still saturated while others were hardly visible. Shadows of the calibration frame on the moving object created radiometric differences between corresponding images. Some other problems such as synchronization problems (see top of images in Figure 1) and echoes further degraded the image quality.

For image acquisition three Aqua TV HR 480 frame transfer CCDs were used with computer pixel sizes of 11.56 μm in x and 7.8 μm in y. The CCDs were modified to transmit pixelclock and were synchronized by using the Mapvision system. The images were stored on three Sony U-matic video recorders. A monitor connected to the video recorders permitted visual control of the

*Modified version of a presented paper, Symposium of ISPRS Commission V, "Close-Range Photogrammetry Meets Machine Vision," Zurich, Switzerland, 3–7 September 1990. Proceedings of the Symposium published by SPIE, Vol. 1395.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 57, No. 8, August 1991, pp. 1079-1086.

0099-1112/91/5708–1079$03.00/0

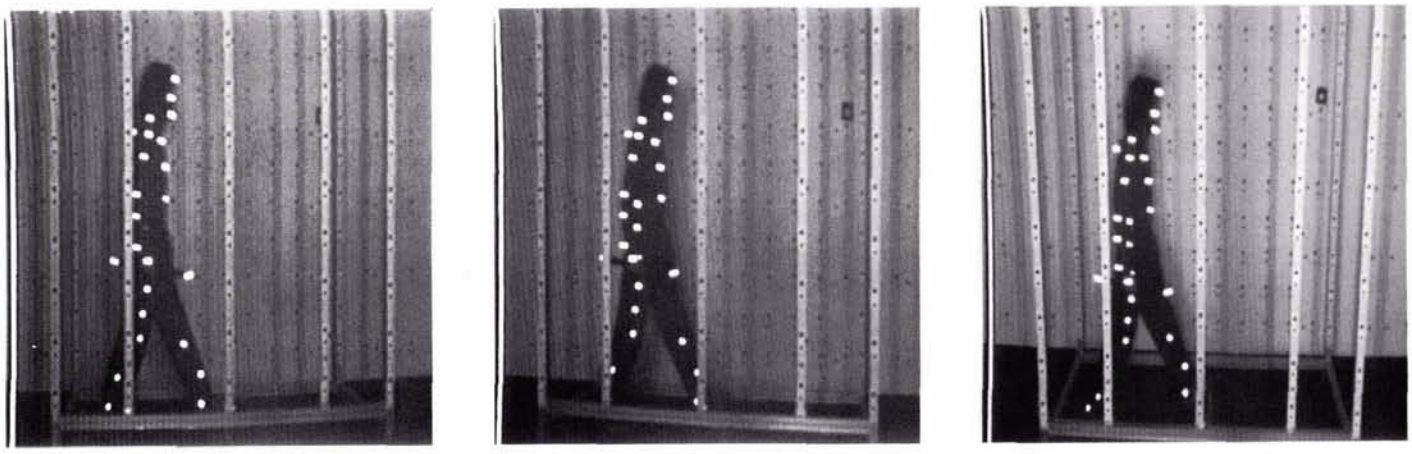

FIG. 1. An original (not interpolated) triplet of the image sequences. From left to right cameras 1, 2, and 3.

image quality. One electronic device between the cameras and the video recorders added a vertical line of high contrast and a binary pattern (see left and top left in the images in Figure 1) to the video signal. One of the cameras provided the pixelclock to the electronic device for the synchronization of the generated pattern with the video signal. The vertical line can be used to determine and correct for line jitter (Beyer, 1988) but it was not used in this project. The binary pattern is a code of the sequence of each image during recording and permits a safe identification of each image during the digitization. The images were digitized from the video tapes by a Matrox MVP framegrabber on a PC. The digitization is automatic with the help of a VES card installed on the PC which permits access to the remote control of the video recorder. Altogether, 56 images (512 by 512 pixels) from each of the three sequences were digitized. The images were transferred to a Sun local network and the processing was performed on a Sun SPARCstation 1.

Each digitized image was divided into two fields, each 512 (H) by 256 (V) pixels. This increased the temporal resolution by a factor of two (50 images / sec), thus making tracking easier, and resulted in uniformly imaged signalized points without a horizontal shift between successive lines. Additionally, it decreased the resolution in the y direction by a factor of two and changed the shape of the blobs. Small (2 to 3 lines), weak contrast blobs are reduced to one-line faint blobs which are very difficult to distinguish from noise. To reduce these negative effects, the fields were linearly interpolated in the y direction. This, of course, leads to an increase in the existing noise and may change the position of a blob, especially if its upper and lower horizontal edges belong to different fields. These images (3 by 112) were treated as the original input images and were further processed. An example of this process is shown in Figure 4.

A 3-D close-range testfield was used for the determination of the interior and exterior orientation of the sensors and of additional parameters to model systematic errors. The testfield consists of a metal frame (which is not absolutely stable) and a wall with about 250 signalized points. The calibration was performed with a bundle adjustment in two steps (for details, see Beyer (1987)). First, five images of the testfield from different stations were used for each camera. These were averaged images and were acquired and digitized in the same way as the image sequences. Only the illumination was changed to increase the contrast of the testfield points. Their pixel coordinates were measured with least-squares template matching. The template matching had some problems because several points on the wall (columnwise) were in the shadow of the testfield

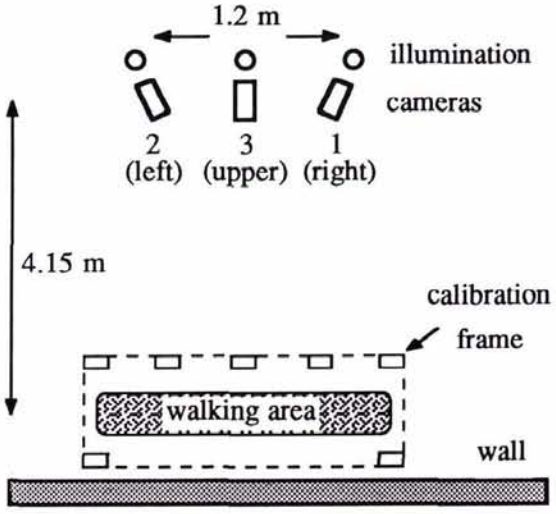

FIG. 2. Geometric configuration of the image acquisition.

frame or imaged close to the frame. This effect caused columnwise systematic residuals in the x direction with opposite sign in the different image regions and reduced the precision of the interior orientation. The principal point coordinates, the focal length, and the additional parameters (ten-parameter set of Duane Brown) for each camera were determined in the adjustment. Only three additional parameters, a scale in the x direction and the first two symmetric radial distortion terms were significant. The image residual RMS values were 1.3 μm in x and 0.4 μm in y. The values of the interior orientation and the additional parameters were used as fixed quantities in a second adjustment for the determination of the exterior orientation. Only three averaged background images were used, one for each sequence. The standard deviations of the exterior orientation parameters are given in Table 1.

As can be seen, the results in X are significantly worse than in the other two directions. The calibration results were used as known quantities in the methods for the determination of 3-D object coordinates. To estimate the positional accuracy that could be achieved by these methods using the calibration results, a few good signalized points on the walking person in the middle of the image were used. Their object coordinates were determined by using the calibration results and an intersection. Thus, the exterior orientation was not treated as error

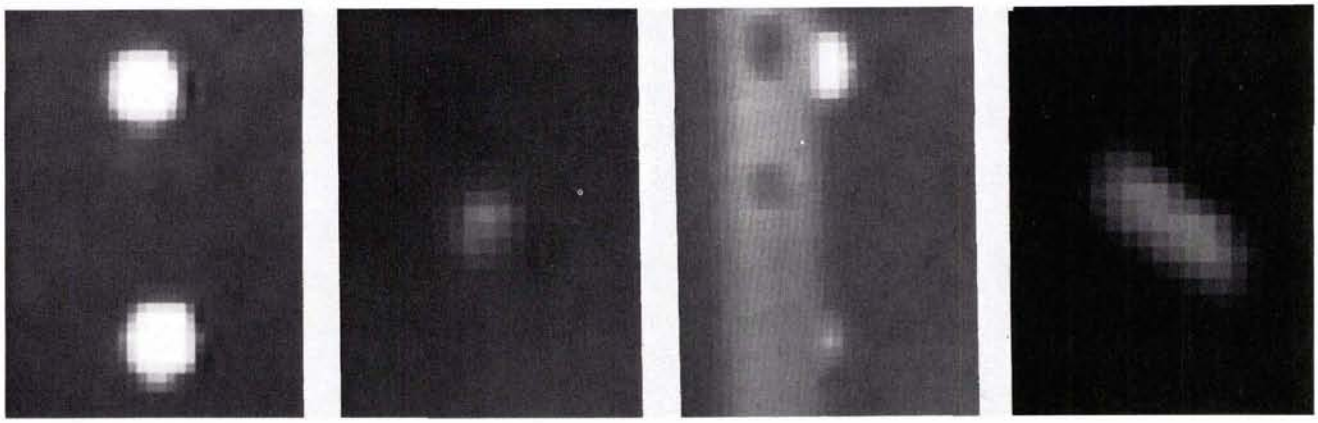

FIG. 3. From left to right: big blobs with good contrast ; small, faint blob ; big bright blob and small faint blob occluded by the calibration frame ; blob with medium contrast and image smear due to motion.

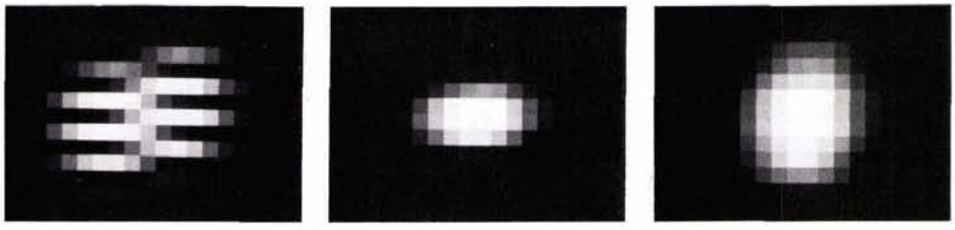

FIG. 4. A blob : original image (left), one field (middle), interpolated field (right).

TABLE 1. STANDARD DEVIATIONS OF THE EXTERIOR ORIENTATION ELEMENTS.

	translations [mm]			rotations [rad/100,000]		
Camera	X	Y	Z	omega	phi	kappa
1	0.77	0.41	0.38	9.5	17.8	7.9
2	0.65	0.39	0.35	9.1	14.5	7.4
3	0.70	0.48	0.41	11.0	15.4	9.3

free but was assigned the covariance matrix that resulted from the second calibration step. The standard deviations of the object coordinates were 0.83 mm in X, 0.61 mm in Y, and 3.5 mm in Z, indicating that the accuracy that could be achieved by the methods for the determination of 3-D object coordinates would be rather low. The significant systematic errors that were not modeled by the calibration were verified by choosing good points in one image and checking their epipolar lines in the other two images of the sequence. The epipolar lines did not pass through the corresponding points but were offset up to one pixel, especially in the x direction.

DETECTION AND LOCALIZATION OF BLOBS IN IMAGES

The first step was the subtraction of the background. Small sequences of the background images for each camera were averaged, divided into fields, interpolated to full images, subtracted from the images of the respective sequence, and the difference images thresholded (Figure 5). To determine the threshold automatically, grey level differences due to motion and noise should be discriminated. To determine the noise level, the following procedure was followed. Because it was known that the moving object did not cover the whole image format, the background was subtracted from the first, middle, and last image of each sequence. In these difference images rectangular regions at the four corners of each image were selected and the

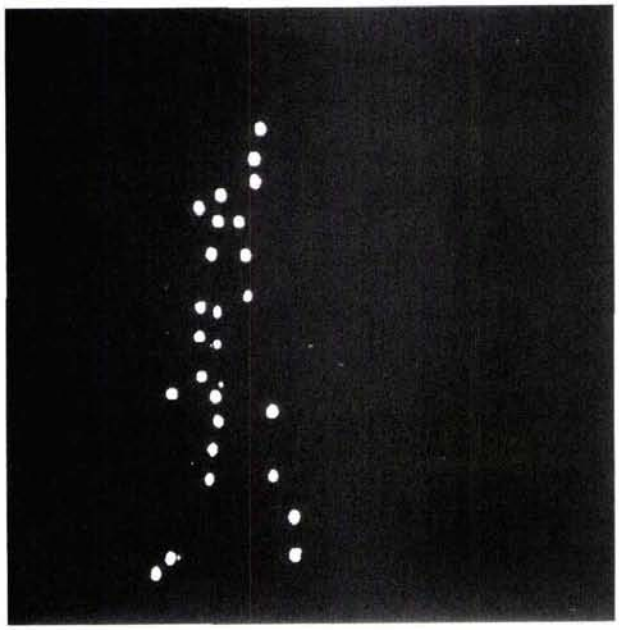

FIG. 5. Thresholded difference image (refers to 2nd field of camera 3 in Figure 4).

statistics of the grey level differences were computed. By using the maximum, average, and standard deviation of the differences, the regions including moving objects could be eliminated and the noise level could be computed from the remaining regions. The maximum difference often exceeded 20 grey levels, an indication of poor image quality, and was not used as a threshold as it might exclude a part of the moving region. The

threshold was selected as a multiple of the standard deviation added to the average and was computed separately for positive and negative differences. In this case, where the blobs were known to be lighter than the background, another one-sided threshold could also be used (and this was the case). This would result in an even smaller moving area/region of interest (ROI). Both methods require that the difference between moving area and background is higher than the noise level (threshold). Only a region of each image was processed except for the first image. This region was determined by the x, y range of the ROI in the previous image plus a maximum x, y image displacement due to motion plus an uncertainty due to occluded blobs at the border of the moving area which could change the x, y range. Because the thresholds were moderately chosen to ensure detection of all blobs and because of other bright blobs caused by the highly reflecting aluminium calibration frame, a few wrong blobs were detected.

The second step was aimed at refining the previous results. It is also useful in cases when no background image is available. Its aim is to detect edge pixels having certain characteristics. Using the binary image of the previous step as a mask, the differences of each pixel from its eight neighbors were computed using the original grey level images. A pixel was selected if n_1 differences exceeded a threshold t_1 and n_2 exceeded a threshold t_2 (whereby $t_1 > t_2$). The thresholds can be chosen based on the standard deviation of the grey levels of the original images or only the moving area. Pixels that were only just rejected were also selected, given that one of their four-connected neighbors was also selected, and this was repeated until no other rejected pixel was found. The result of this operation was a binary image.

The third step used these binary images and the original images, detected each blob (blob coloring), derived different attributes for each one of them, computed the coordinates of their centers, and chose only those blobs whose attributes were within a range defined by *a priori* knowledge. Because the output of the previous step was the outline (edges) of the blobs, the edges were first filled-in to permit an easier selection of a range for each of the blob attributes. For blob coloring, four-connectivity was used to avoid merging together eight-connected neighboring blobs. The following attributes may be used for blob selection (their permissible range of values for this case is listed in brackets):

- minimum and maximum perimeter (6 , 60).
- minimum and maximum size (2 , 225).
- minimum and maximum for x and y range of blob pixel coordinates (1 , 15).
- maximum of ratio max(x range,y range) / min(x range,y range) (3).
- lower and upper limit of ratio blob size / (x range * y range) (0.5, 1).
- similarly for the four quadrants of the rectangle including the blob (0.5 , 1). The number of quadrants that must pass the test can be chosen (2) and an upper limit for the maximum of the quadrant ratios over the minimum one (2).
- lower and upper limit of grey level mean (65 , 255).
- lower and upper limit of grey level standard deviation (1 , 127).
- minimum, maximum, and mean grey level values of the blob and a neighborhood (larger or smaller than the rectangle including the blob). Depending on the type of blobs (bright on dark background or the opposite) and the neighborhood size, certain relations between the minimum, maximum, and mean grey level values of these two areas must be fulfilled.

Most of these thresholds can be derived by knowing the physical size and shape of the signalized points and the imaging geometry and, hence, also their size and shape in image space. For an average ideal case, the blobs should be 4.5 by 6.6 pixels in x and y, respectively, and very bright on a dark background. As can be seen, the selected thresholds are very relaxed and no strict requirements for their selection is required. This permitted the selection of almost all blobs ranging from a size of 1 by 2 pixels up to 15 by 15 pixels and a brightness of 65 grey levels, hardly distinguishable from their background, up to 255 grey levels. A nice property was the fact that the method worked with the same thresholds for all sequences, although their radiometric properties differed. The computation of the center pixel coordinates can be computed by the following methods:

- center of gravity using the binary image (set 1 of pixel coordinates).
- center of gravity by using grey values as weights. These grey values are the difference of the original grey values from a given grey level or from a grey level with a given distance from the minimum grey level of the blob (set 2).
- center of gravity by using the grey level gradients as weights. Only gradients above a certain threshold were considered (in this case, 4 to 5 grey levels) (set 3).
- center computed from the two positions (ascending and descending) of the maximum gradient in each x and y profile of the blob. This option was not actually used.

The pairwise differences between the first three methods were also computed. With good targets, their range was 0 to 0.2 pixels. Greater differences, up to 0.9 pixels, appeared when the blobs were close to or partly occluded by the bright calibration frame. In this case, the second method shifted the center towards the frame, while the third one shifted it in the opposite direction. Hence, the average of the second and third methods was also computed (set 4 of pixel coordinates).

The full information about all blobs, accepted and rejected, was stored to permit an updating of the results by subsequent processing steps. This information included a binary image with all detected blobs and a file with the attribute values and the four sets of center pixel coordinates for each blob. For each rejected blob, the reason for rejection was also given. A visual control of the results showed that the great majority of the 8,695 detected blobs were correct. A few wrong blobs were also selected due to the very relaxed thresholds. On the other hand, some real blobs were not detected because they were very small and faint. Overlapping blobs were handled as one blob, because no method was incorporated at this stage to permit blob splitting. Out of the 32 signalized points, 17 to 34 blobs/image were detected in the second step and 17 to 31 were selected in the third one, the big range indicating severe occlusion problems.

The blob center pixel coordinates were further refined by means of least-squares template matching (set 5 of pixel coordinates) (Gruen and Baltsavias, 1985). The fourth set of the previously determined pixel coordinates was used as approximate values for the template matching. With a preselected synthetic template, problems could arise because of the big scale/shear and brightness/contrast differences among the blobs. However, the affine geometric and the two radiometric parameters proved enough to transform the blobs such that they fitted to the template. The selected template was small (13^2 pixels) and included little more than the circular target, to make matching insensitive to the background. The varying background (including the case when two blobs touch each other) was the greatest problem; and it is similar to the problems encountered when using template matching to measure signalized control points in aerial images. Investigations showed that the best performance can be achieved by selecting a template as described above, computing the grey level gradients from the average of template and search signal, updating the radiometric corrections in each iteration, and weighting the grey level observations with the corresponding template grey level gradients (instead of the identity matrix as is usually done). The latter also results in more stable shaping parameters. For each point the following information was saved: size of affine parameters, change from approximate center coordinates, the *a posteriori* variance of unit weight from the adjustment, the grey level crosscorrelation

coefficient, the standard deviations of the two shifts, and the number of iterations. These statistics can be used for a qualitative analysis of the results. In this case the last five measures were used, and points with values outside the range (mean ± 3 standard deviations) were marked as candidate bad points. Some statistics of the matching results are listed in Table 2. Successful points converged within the predefined maximum number of iterations and their patches were not transformed outside the search window.

Finally, the results of the different methods were combined to create the final pixel coordinates. The results are sorted based on their quality. The points are divided into four groups:

- Group 1 : Results of matching and step 3 are satisfactory.
- Group 2 : Only the result of matching is satisfactory.
- Group 3 : Only the result of step 3 is satisfactory.
- Group 4 : Both steps rejected the point.

The points of group 1 are considered the best, then those of group 2, etc. Within each group different quality criteria are used for sorting and the final pixel coordinates are chosen differently.

Group 1:

Quality criteria:
- −average of absolute differences of x, y pixel coordinates of sets 4 and 5
- −maximum of x, y-shift standard deviation from matching
- −size of blob (actually 1/size is used)

Choice of final pixel coordinates (PC):
if |set 5 − set 4| > threshold t_1 and |set 3 − set 2| > threshold t_2 then PC = (set 5 + set 4)/2
otherwise PC = set 5

Group 2:

Quality criteria:
- −maximum of x, y-shift standard deviation from matching
- −size of blob

Choice of final pixel coordinates: PC = set 5

Groups 3 and 4:

Quality criteria:
- −average of absolute differences of x, y pixel coordinates of sets 2 and 3
- −size of blob

Choice of final pixel coordinates:
if |set 3 − set 2| < threshold t_2 then PC = set 4
otherwise PC = set 2

The criteria values were transformed in the range [0, 1] so that they have the same numerical contribution. The smaller the value, the better the point is. The final quality value is the weighted sum of all criteria values. In this case, equal weights were used. The thresholds t_1 and t_2 were 0.25 and 0.3 pixels, respectively. For assigning each point to one of the groups, a point was considered rejected by step 3 only if the perimeter or size of the blob was outside the range. The results of all groups were stored but marked appropriately.

DETERMINATION OF 3-D OBJECT COORDINATES

Using the pixel coordinates of each image triplet, the 3-D object coordinates were determined. Two methods were used. The first is matching based on intersections of multiple epipolar lines (Maas, 1990) and will not be described in detail. Due to software communication problems, the previously determined pixel coordinates could not be used. They were derived from the binary images of step 3 (including the rejected blobs too) and as such they were neither very precise nor free of errors. The principle of the method is the following. One image is selected and its blobs sequentially processed. By knowing the interior and exterior orientation of the cameras and the additional parameters of the bundle adjustment, the epipolar lines of each blob in the remaining two images were computed and candidate corresponding blobs in the second and third image along these lines were identified. Because the calibration is not perfect, the search for blobs is within a band along the epipolar lines. When more than one candidate exist in the second image, their epipolar lines are computed in the third one and checked whether they pass through one of the selected candidates. Thus, through redundant information, many multiple solutions can be disambiguated. After initial candidates are selected for each pair of images, the final correspondence is made based on a combinatorial algorithm, and, thus, consistent triplets or pairs are selected. Finally, the 3-D coordinates are computed by intersection by means of a least-squares adjustment (Table 3). The values of Table 3 are rather pessimistic. The precision of vectors is higher than the precision of single points, due to high correlations between the 3-D coordinates of successive triplets (verified by the bundle adjustment mentioned in the section on quality analysis). The *a posteriori* standard deviation of unit weight is high due to the binary determination of the center of gravity, and the standard deviations of the object coordinates are high due to calibration imprecisions that forced the use of a wide epipolar band.

The method requires only very coarse approximate values but, if good approximations exist, they can further restrict the search space. Because of the use of a 54-μm wide epipolar band, and the density of blobs in certain image regions, wrong correspondences were chosen in some cases (particularly when only pairs could be found; see Table 4). An option of the method, which was used in this case, permits the selection of one image blob for multiple object blobs to account for occlusions. This happened also in our case but the multiple selection did not correspond to occlusions in most of the cases, but rather to blobs identified in only two images (the blob in one image had two corresponding points in the second one).

The second method is based on multiphoto geometrically constrained matching (Baltsavias, 1988). The results of the first method were used as approximations. The images of camera 3 (see Figure 2) were selected as reference images (obviously, blobs existing only in the other two images were not processed).

TABLE 3. INTERSECTION RESULTS (GLOBAL VALUES FROM TRIPLETS).

S_o [μm]	standard deviation [mm] X	Y	Z
4.33	1.55	1.66	7.28

TABLE 2. RESULTS OF TEMPLATE MATCHING FOR BLOB LOCALIZATION.

Number of points	Successful matchings	Points/ sec	Iterations/ sec	Iterations/ point
8,695 (100%)	8,556 (98.4%)	7.1	24.3	3.4

TABLE 4. NUMBER AND TYPE OF DETERMINED OBJECT POINTS.

Maximum possible object blobs	Detected blobs	Detected triplets	Detected pairs
3,584 (=32×112)	2,873 (80.2%)	2,169 (75.5%)	704 (24.5%)

A test for automatic detection of blunders was used but with quite relaxed criteria (see Table 5). A point was accepted if at least two of the three rays passed the test. To account for the calibration imprecisions, a first convergence was performed with large weight for the collinearity conditions and then a second convergence with reduced weight to permit finding corresponding points that are a few pixels away from the epipolar lines.

The above two methods were repeated with only two image sequences (cameras 1 and 2). The second method was also repeated with cameras 2 and 3. For the second method, the images of camera 2 and 3 were used as reference for the two two-ray versions, respectively. The epipolar band used in the first method had a width of 58 μm. A comparison between two and three images for both methods showed the advantages arising from the use of a third camera. The number of detected blobs increased considerably, disambiguation of multiple solutions became easier, occlusions in only one image could be accommodated by the other two, and the precision of the object coordinates increased. Thus, by a limited increase of H/W and processing time, the results became more complete, precise, and robust.

DETERMINATION OF OBJECT TRAJECTORIES

Tracking of the blobs in each image sequence was performed by two methods. The first method is described by Papantoniou and Dracos (1989) and in more detail by Papantoniou and Maas (1990). The method operates on three successive data sets (images) to derive the connections between the first two. To resolve disambiguities, criteria such as maximum displacement between adjacent images, Lagrangian acceleration, local correlation of the velocity vectors, and kinetic energy are used. The local correlations within each group of neighboring blobs were not considered in this experiment. The method has been used with less complex motions than in this case and very high, uniform target density (on the order of 1000 blobs/image), while in our case the target density is medium. A disadvantage of the method is that it will not continue tracking an object after an interruption of the track (e.g., occlusion). Occlusions were inevitable and, thus, only partial trajectories could be determined (see Figure 6).

Tracking was sometimes interrupted when the motion changed direction abruptly (typical case with hands and feet). Some statistics of the image motion from field to field are listed in Table 6.

The second method consisted of three steps. First, in each sequence the following was done for all images except the last one. For each blob (reference image), its position in the next image was determined by least-squares matching. The position of the blob in the current image was used as an approximation for its position in the next one. This was justified by the fact that image motion was limited to a few pixels from image to image. The parameters for the least-squares matching were similar to those used for the blob localization (the weight matrix of the grey level observations was the identity matrix). Because now no synthetic template was used and the template dimensions were fixed and chosen such that they include the biggest blob, problems occurred with small blobs when their background changed from image to image, e.g., close to the calibration frame. These signal disturbances sometimes led to an instability and a wrong estimation of the shaping parameters. Therefore, a second option using only the two shifts was tried. Although the geometric adaptation was insufficient and the results less precise, the solution was more robust. Precision is not important in this case, because both 2-D and 3-D coordinates have already been determined, but tracking is. After this process, pairs of correspondences within each image sequence have been established.

The second step was the combination of corresponding pairs in partial trajectories, i.e., trajectories without any interruption. The procedure will be explained by an example. Assume that the blobs of the first image have been tracked in the second one. These coordinates (list 1) were compared to the coordinates of the extracted blobs in the second image as explained in the section on blob detection and localization (list 2). Each of the coordinates of list 1 was compared to all coordinates of list 2. If their absolute difference was less than a threshold, the blobs were considered identical (the case that the difference was less than the threshold for more than one blob of list 2 never occurred). If no difference less than the threshold was found, it was assumed that either the trajectory was interrupted due to occlusion or that the tracking (matching) was wrong. After all blobs of list 1 had been processed, list 2 was checked for unassigned blobs. These were considered as the beginning of new partial trajectories. By this method the pairs of images 1-2 and 2-3 were combined in triplets 1-2-3 and the process was continued up to the last image. The partial trajectories were very similar to those determined by the first method, and thus will not be presented. The first and the second step can be combined. In this case, the approximate position of each blob in the next image (used for matching in step one) can be derived from the blob position in the current image plus the known displacement from the previous image to the current one.

The third step connected the partial trajectories to form full trajectories. For the last point of each partial trajectory, candidate start points of other partial trajectories were searched. Only partial trajectories starting within a certain range of fields after the field of the last point were considered. The position of the last point in the field of the start point of the candidate trajectory was extrapolated by using its known position in the previous three fields. The extrapolation accounted for the possibility that the motion direction changed abruptly by considering the acceleration of the movement. The distance of the extrapolated last point from the candidate start point was computed. The

TABLE 5. NUMBER OF AUTOMATICALLY DETERMINED BLUNDERS.

version	number of points	number of blunders
matching, 2 rays (cameras 2 and 3)	2417 (100%)	122 (5.05%)
matching, 3 rays	2872 (100%)	71 (2.47%)

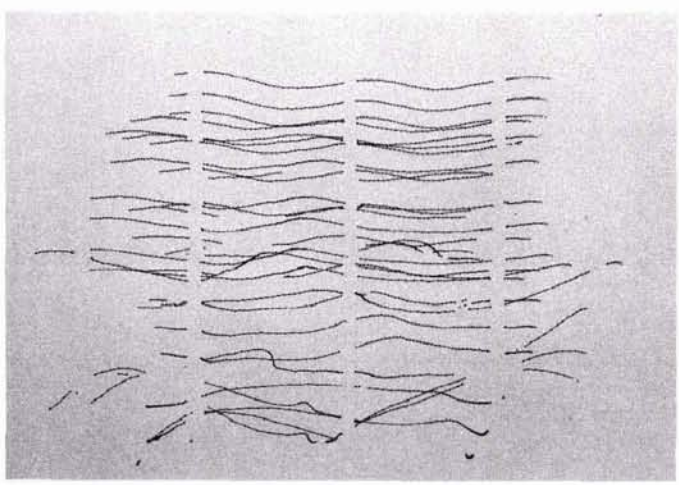

FIG. 6. Partial trajectories determined by the first method (sequence of camera 1).

TABLE 6. STATISTICS OF THE IMAGE DISPLACEMENT DUE TO MOTION (IN PIXELS).

	2-D displacement			x-displacement				y-displacement			
	Mean	St. Dev.	Max	Mean	St. dev.	Max	Min	Mean	St. Dev.	Max	Min
Camera 1	2.97	1.54	11.58	2.84	1.48	9.54	−0.90	0.00	0.96	6.91	−6.46
Camera 2	2.90	1.52	11.59	2.77	1.45	9.52	−1.09	0.00	0.96	7.73	−5.94
Camera 3	2.86	1.51	12.21	2.72	1.48	10.00	−0.51	0.00	0.94	6.82	−7.00

same extrapolation was done for the candidate start point in the field of the last point, and a second distance and the average of the two distances were computed. This was done for each candidate start point. The best candidate start point was selected based on the average distance and the time (field) distance between last and start point. An additional requirement was that the selected start point did not fit better to another partial trajectory whose last point was within a time range from the field of the last point under consideration. The results are shown in Figure 7. Some partial trajectories (especially those of the blobs on the left arm and leg) could not be connected because the occlusion time, and thus the distance between the partial trajectories, was too long and thus the extrapolation of the magnitude and the direction of displacement was not reliable. Tracking in image space caused problems in some cases because the distance between neighboring blobs was much smaller than the average displacement between fields. Such ambiguities can be avoided in object space.

Although the above methods do not use the available information in object space, the latter can be used at least as a post-processing for verification of the results and elimination of errors. Assume that blob A^1_1 (superscript indicating image number and subscript image sequence number) was tracked in the second image, i.e., A^2_1. The corresponding blobs A^1_2, A^1_3 and A^2_2, A^2_3, respectively, are known from the derivation of the 3-D coordinates. By tracking in image space A^1_2 and A^1_3, blobs A^{*2}_2 and A^{*2}_3 are detected. If the solution is correct, they should coincide with A^2_2 and A^2_3. Ambiguities can be resolved by using the quality criteria which are available for both image tracking and 3-D object coordinate determination.

QUALITY ANALYSIS

Limited space does not permit an analysis of side aspects which might seem secondary but are still important. Here, only some aspects of accuracy will be discussed. The circular targets on the ruler were manually identified and their determined object coordinates were used to calculate the distance between them. This distance was compared to the known distance after a correction for room temperature. Points that were automatically determined by matching as blunders were not used. The statistics of the differences are listed in Table 7. Because the results contained some blunders which strongly influenced the global statistics, differences outside the range median ± 3 standard deviations were excluded. These empirical accuracy measures were compared to the theoretical accuracy obtained by the bundle adjustment. By determining from three rays some of the ruler points as new points in a bundle adjustment and using the covariance matrix of their object coordinates and error propagation, a theoretical accuracy for the distance was derived. It was 1.97 mm.

As can be seen from Table 7, the empirical and theoretical accuracy estimates are consistent. In some cases, the empirical accuracy is slightly worse than the theoretical accuracy mainly because of the partially occluded targets which were also used in these computations (occlusions account for a shift of the x-pixel coordinates up to two pixels). This comparison was performed for both 3-D object coordinate determination methods. As expected, matching is slightly more accurate than the first

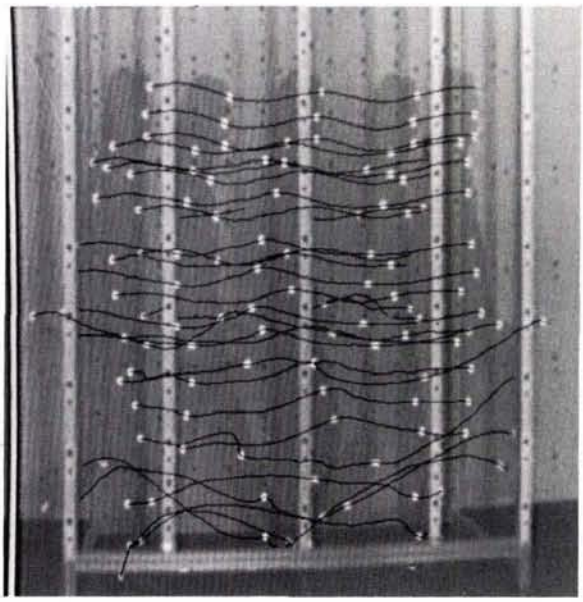

FIG. 7. Full trajectories by the second method (sequence of camera 1).

method for both two- and three-ray versions. The first method is less accurate partly because of the binary determination of the center of gravity. The three-ray versions for both methods are generally more accurate than the two-ray versions. The only exception is the camera 2-3 version of matching as compared to the three-ray version of matching. There are two possible explanations for this. First, in the first version more blunders were detected and removed as seen from the Table 5, and thus the magnitude of the remaining ones was smaller as seen from the maximum error in Table 7. Second, matching was performed with a small weight for the collinearity constraints, because the corresponding points did not lie on the epipolar lines, and hence the image rays from which the object coordinates were estimated did not intersect. The distance of the corresponding points from the epipolar line for camera 1 was bigger than that for camera 2 and thus the intersection error was bigger for the three-ray version. The systematic negative bias in all versions implies a scale difference possibly due to the calibration inaccuracies.

CONCLUSIONS

A method for tracking of signalized points was presented consisting of independent modules for blob detection and localization, correspondence and 3-D coordinate determination, and object tracking. This modularity permits an easy replacement of the first module by another one using natural features. Different approaches for performing each one of the three previously mentioned tasks were used and compared to each other. Although many difficult cases had to be handled, such as many and long occlusions, multiple solutions, small and hardly visible targets, and differently moving objects, and the image quality

TABLE 7. DIFFERENCES OF KNOWN DISTANCE MINUS ESTIMATED (IN MM).

Version	number of distances	number of blunders	maximum absolute	average	average absolute	RMS
matching, 3 rays	93	10	4.91	−0.83	1.39	1.74
matching, 2 rays (cameras 2, 3)	73	4	3.26	−0.75	0.95	1.17
matching, 2 rays (cameras 1, 2)	72	4	5.91	−1.26	1.83	2.21
method 1, 3 rays	96	9	6.11	−1.70	1.97	2.42
method 1, 2 rays (cameras 1, 2)	75	6	6.60	−2.16	2.42	2.89

and camera calibration were rather poor, the results are satisfactory. Errors in tracking appear only when big gaps exist in the image sequences. The achieved accuracy, as checked by the known distance, was similar or even better than the theoretical accuracy. Least-squares matching with its extensions proved to be a quite general and flexible technique, being able to handle blob localization, 3-D measurement, and image tracking. The low-level processing is often underestimated. Thus, its poor results adversely deteriorate the subsequent steps. Efforts to improve blob detection and localization pay off, as proven by this experiment. The use of a third camera adds minor costs but considerably improves completeness, precision, and robustness of the results. The illumination, although it was not an important factor for this experiment, greatly influences the results and is a very inexpensive way to improve them.

Some weaknesses which will be investigated as a part of future research include the following. The lack of feed-back between the different modules is a major drawback. Errors from each module propagate to the next one and they are often amplified. Thus, it is necessary to perform all steps simultaneously or with a feed-forward, feed-back loop. This not only will diminish the amplification of errors, but will also decrease the errors of each module by the use of global information. For example, missing (undetected or rejected) blobs could be identified by another, less strict search or revision of rejected blobs, respectively, within the epipolar band. Although very often quality measures and sorting of the results were available, this information was not used to resolve ambiguities between the results of different methods (e.g., different correspondences from image tracking and 3-D measurements) or guide the process, e.g., in a best-first strategy scheme, which will greatly improve the robustness. Such hierarchical strategies can be used in multiple levels such as use first the biggest blobs, the most precise image and object points, the longest trajectories, etc. Methods for precise blob center determination, in case of partial occlusions and blob overlapping, must be developed. Apart from blob splitting methods, the information from image tracking (e.g., two blobs approaching each other) and the 3-D information from correspondence (where no overlapping is possible) can play a very important role in solving this problem. Slightly overlapping blobs can be separated by morphologic operations (opening and closing). To determine correctly the center pixel coordinates of a partially occluded blob, the center must be determined only from its partially visible perimeter (knowledge about the size and shape of the blob in the previous images where no occlusion occurs can be used to extrapolate its size and shape in case of occlusion). Knowledge about the object motion can be incorporated in all three processing modules, especially in object tracking (e.g., in this case knowledge about the position of blobs on the human body and thus the motion, i.e., translation or translation and rotation, could be exploited). Topological relations can also be used for image or object tracking, e.g., the top blob (on the head) in one image should correspond to the top blob in the next image. A correct calibration with complete modeling of the systematic errors and the analysis of its stability over time is an essential factor, especially if high precision is required. The interpolation of the image fields improves the blob detection and tracking but apparently contributed to measurement errors. Thus, either an edge-guided interpolation should be used or the 2-D and 3-D measurements should be performed in the original images or fields.

ACKNOWLEDGMENTS

The authors want to express their gratitude to Dipl. Ing. H.-G. Maas for supporting the image acquisition and digitization and providing software for the detection of 3-D coordinates, Dipl. Ing. H. Beyer for providing software and performing the calibration of the CCDs and a bundle adjustment for 3-D quality analysis, D. Wilkins for modifying his template matching program to permit a fast processing of image sequences, and Dr. D. Papantoniou for providing software for tracking and for engaging in stimulating discussions on tracking problems.

REFERENCES

Baltsavias, E. P., 1988. Hierarchical Multiphoto Matching and DTM Generation. *Proceedings of 16th ISPRS Congress*, Kyoto, Japan, 1 - 10 July. *International Archives of Photogrammetry and Remote Sensing*, Vol. 27/B11, pp. III:476–492.

Beyer, H. A., 1987. Some Aspects of the Geometric Calibration of the CCD Cameras. *Proceedings of the International Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, 2 - 4 June,pp.68–81.

———, 1988. Linejitter and Geometric Calibration of CCD-Cameras. *Proceedings of 16th ISPRS Congress*, Kyoto, Japan, July 1 - 10. *International Archives of Photogrammetry and Remote Sensing*, Vol.27/B10, pp. V:315–324.

Gruen, A., and E. P. Baltsavias, 1985. Adaptive Least Squares Correlation with Geometrical Constraints. *Proceedings of SPIE*, Vol. 595, pp. 72–82.

Maas, H.-G., 1990. Digital Photogrammetry for Determination of Tracer Particle Coordinates in Turbulent Flow Research. *Proceedings of SPIE*, Vol. 1395.

Novak, K., E. P. Baltsavias, and A. Gruen, 1990. *Automatische Objektverfolgung in Stereovisionssystemen*. Report No. 170, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, January 1990.

Papantoniou, D., and T. Dracos, 1989. Analysing 3-D Turbulent Motions in Open Channel Flow by Use of Stereoscopy and Particle Tracking. *Advances in Turbulence* 2, Springer Verlag, Berlin, pp. 278–285.

Papantoniou, D., and H.-G. Maas, 1990. Recent Advances in 3-D Particle Tracking Velocimetry. *Proceedings of the 5th International Symposium on Applications of Laser Techniques to Fluid Mechanics*, Lisbon, Portugal, 9 - 12 July.

Zhou, H. B., 1990. Object Points Detection in a Photogrammetric Test Field. *Proceedings of SPIE*, Vol. 1395.

(Received 24 July 1990; accepted 17 August 1990)