AUTOMATED MEASURING OF TRUNK SHAPE FROM A SEQUENCE OF IMAGE PAIRS

Olli Jokinen

Department of Surveying
Helsinki University of Technology
P.O. Box 1200, 02015 TKK, Finland
ollie.jokinen@tkk.fi

ABSTRACT

The paper deals with digitizing the 3-D shape of a trunk from a sequence of image pairs when a harvester is approaching the tree. The stereo correspondence problem is solved hierarchically using natural features, image correlation, and epipolar constraints. False matches are removed by analyzing the distribution of reconstructed 3-D points and applying a novel similarity criterion. Two methods for solving the motion between successive frames are considered. The first one is based on tracking the stereo correspondences to the next frame using image correlation. The second one is based on registering 3-D point sets of successive frames using surface matching techniques. In the first case, the final surface model is triangulated from points successfully tracked through several frames while in the second case, it is triangulated from points where the correspondences are compatible in surface matching. Test results verify the performance of the methods in a pine forest.

1. INTRODUCTION

In mechanical felling, a harvester chops down a tree, cuts off the branches, and bucks the trunk into pieces. Optimal cutting positions are selected according to the dimensions and shape of the trunk, which are currently estimated using mechanical sensors mounted in the logging unit. These sensors provide near mm-level accuracy for the tapering of the trunk while all the other qualitative attributes are evaluated by the harvester operator. Optimal cutting of the trunk depends on the demand, the trunk at hand, the stocks of already cut trees and logs locally and within the wood procurement organization. A price matrix provides the bucking computer with information on how to prioritize various log grades and diameter-length combinations within the same grade, while a demand matrix specifies the desired proportion for each combination (Kivinen, 2004). On the whole, well optimized cutting brings the forest owner increased economical profit.

An improvement for the cutting process would be if the shape of the trunk could be measured already when the harvester is approaching the tree. This would require installing additional non-contact sensors, such as cameras or a laser scanner, on-board the harvester. The performance requirement of such a measuring system is high, namely, mm-level accuracy in near real-time. Moreover, the system should be able to measure the trunk up till six meters.

In this paper, it is investigated if a stereo vision system with cameras mounted above the harvester grab could be used for digitizing the 3-D shape of the trunk when the harvester is approaching the tree. The main focus is on obtaining a surface model for the lower trunk while subsequent processing for extraction of geometric information of the trunk is left for future
research. We consider the case of calibrated cameras where the interior and relative orientations of the cameras have been solved in advance in a calibration test field. The stereo correspondence problem is solved using natural features extracted from the images. These yield a set of 3-D object points which are tracked in the images to the next image pair of the sequence. The motion between subsequent image pairs is obtained using the tracked 3-D points. Alternatively, a method based on surface matching is considered for registering the 3-D data sets reconstructed from subsequent stereo frames. Although the methods are mostly well-known, the application is new, unexplored, and challenging in a difficult forest environment. Moreover, the similarity criterion introduced for removing false stereo matches is a new contribution.

The paper is organized as follows. Related works on object reconstruction from a stereo sequence are discussed in Section 2. Section 3 presents the methods for interest point detection, hierarchical stereo correspondence matching, and motion estimation. The methods are explored with real data captured with cameras on-board an allterrain vehicle in the first test case and on-board a harvester in the second test case in Section 4. Conclusions and ideas for future research are presented in Section 5.

2. RELATED WORKS

3-D structure can be derived from corresponding features between the left and right images. Motion of the cameras brings in new data that can be used to improve the structure estimation. The main questions are how to establish correspondences between the left and right images and between successive frames, and how to process the successive frames for the improvement of the structure.

Moyung and Fieguth (2000) propose an iterative algorithm where epipolar constraints and rigidity constraints are used to resolve ambiguous stereo and motion correspondences, respectively. The reconstruction and motion estimation are iterated. The accuracy of the motion estimate improves as the number and accuracy of reconstructed points increase.

Molton and Brady (2000) combine structure and motion estimation by allowing multiple hypotheses in the stereo matches until a motion estimate is available. Features which move consistently with a global motion are considered as correct stereo matches. The system is able to recover large motion steps between successive views.

In (Gilbert et al., 2005), feature points identified on a rigid object are tracked separately in the left and right images using correlation. 3-D points are reconstructed and the point clouds are registered yielding a new position of the cameras. The accumulated error in the computed positions is compensated for by detecting loops in the movement.

In (Badino et al., 2004), stereo matching is computed for every frame and optical flow is applied to find correspondences between points in successive frames. Stabilization and a better estimation are achieved computing the observed motion between the current frame and many frames in the past in a multi-frame fashion.

In (Comport et al., 2007), the tracking of a stereo rig is based on matching the current image pair against a reference image pair that contains dense stereo correspondences.

Salgado and Sánchez (2005) compute optical flow for the left and right images, and then use the
optical flow information to constrain the computation of disparity maps. Large displacements are solved using a coarse-to-fine approach.

Zhang et al. (2003) match a pair of video streams simultaneously instead of matching pairs of images frame by frame. For moving objects, the matching is based on oriented spacetime windows that allow the matching pixels to shift linearly over time.

Our approach for tracking-based motion estimation is similar to Gilbert et al. (2005) in that we track corresponding points separately in the left and right images using correlation and then compute the motion using the reconstructed 3-D points.

There appear also many studies (Hopkinson et al., 2004; Pfeifer et al., 2004; Thies et al., 2004) where terrestrial laser scanning is considered for determining geometrical aspects of trees such as tree height and trunk diameter at breast height. Terrestrial laser scanning is also of our interest as our colleagues have been developing a 3-D laser scanner system intended for standing tree trunk measurements and mounted to the harvester head. The first results of that work can be found in (Miettinen et al., 2009).

3. METHODS

The processing chain for the trunk measurement is described in a simplified flow chart in Figure 1. As an input, we have a sequence of image pairs taken with calibrated cameras on-board a harvester approaching a tree, and the output is a surface model for the lower trunk. We are confined to pine forests where the trunks are clearly visible.

![Figure 1. A simplified flow chart of the measuring process.](image-url)
3.1 Area of interest and interest points

It is assumed that the user points out the trunk that is to be measured so that one pixel on the trunk is known. Moreover, it is assumed that the orientation of the sensor is obtained from an inertial navigation system. The image is then rotated around its center so that the trunk is vertical in the image. Vertical edges are then detected using the Sobel operator. The number of edge pixels in each column of the image is studied. This number is high at trunk boundaries and also on the trunk in the case of a pine. There usually appear also many unwanted vertical edges that result from other trees than the one in question.

There are two alternatives to proceed. The first approach is applicable when the images have been taken so that the lower part of the image contains vertical edges mainly from the tree of interest located in the foreground while the upper part of the image includes also edges from other trees in the background. The edge image is first opened morphologically in order to remove edge pixels not belonging to the tree of interest. The histogram of vertical edges at each column in the lower part of the image is then thresholded, where the threshold is selected as $k$ times the maximum of the histogram with $k$ chosen experimentally (0.5-0.7 in the experiments). The first and last column left after thresholding are regarded as approximate trunk boundaries.

The second approach is applicable when most of the edge pixels result from trunk boundaries and not so much from the trunk itself. The edge image is opened morphologically and the histogram of edge pixels at each column is thresholded using an appropriate percentile increasing from 95 to 99 percent as the vehicle is approaching the tree. The columns, which are left after thresholding, are morphologically closed and weighted centers of groups of columns next to the trunk pixel given by the user in the beginning are calculated. The columns at these weighted centers are regarded as approximate trunk boundaries.

In both approaches, an area of interest (AOI) is selected as between two vertical straight lines located at the approximate trunk boundaries. The image is finally rotated back to the original orientation. An optimal AOI would contain only the trunk area and no background.

Interest points are determined within the AOI using the Canny edge detector. Points, which are vertical edges according to the Sobel operator, are not considered as they may be located at occluding boundaries of the trunk. Interest points are determined in the left images of each image pair.

3.2 Stereo matching

The stereo correspondence problem is solved hierarchically starting from interest points which are strong edges. Once the corresponding points have been found for strong edges, a dense disparity map is generated by interpolation and extrapolation for the whole image. More interest points including also weaker edges are then processed limiting the search space of corresponding points around the values given by the disparity map. This reduces the computing time considerably.

At each level of hierarchy, the stereo correspondence problem is solved using the standard method based on image correlation and epipolar constraints. For each interest point in the left image, an epipolar line in the right image is calculated. Normalized image correlation is then calculated between a neighborhood around the interest point and a neighborhood around each pixel within a narrow band around the epipolar line. The size of the correlation window is
selected manually according to the variation of image texture (21 x 21 pixels in the test cases). The pixel which gives maximum normalized correlation is chosen as an initial corresponding point if the correlation exceeds a given threshold (0.7 in the test cases). The corresponding point is refined to sub-pixel accuracy by fitting a polynomial surface of second degree to the correlation values within a neighborhood around the initial point and finding the maximum point of this surface (Pan et al., 2006).

False matches are rejected during postprocessing as follows. Let the coordinate system be fixed to the left camera so that the x-axis points to the right, the y-axis upwards, and the z-axis towards the viewer. The 3-D object coordinates of the matches are calculated by intersection. First, only those matches are accepted for which the distance of the 3-D points from their median in the approximately horizontal \(xz\)-plane is less than the sample mean of the distances plus \(n\) times the standard deviation of the distances where \(n\) increases from zero to two depending on the level of the hierarchy. A straight line is then fitted to the 3-D points left. A histogram of distances from the straight line is calculated. Second, those matches are accepted for which the distance from the straight line is less than a threshold that is selected as a distance at which the histogram has decreased below \(h\) times the maximum of the histogram where \(h\) is selected experimentally and manually (0.1 or 0.3 in the test cases). A new straight line is then fitted to the 3-D points left and the same process is repeated. Finally, the 3-D points left are triangulated into a TIN (triangulated irregular network) surface. The length of the longest side of each triangle in the TIN model is calculated. Only those points are accepted which belong to a triangle the longest side of which is less than the sample mean of longest sides of triangles.

The sets of corresponding points should constitute similar shapes in the left and right images (cf. Figures 4, 5, 11), i.e., the corresponding points should satisfy a similarity criterion to be defined as follows. Let \(p\) be a point in the left image and \(q\) the corresponding point of \(p\) in the right image. The locations of other corresponding points within neighborhoods of \(p\) and \(q\) are compared. Binary image windows centered at points \(p\) and \(q\) are constructed, which contain ones at locations where there appear corresponding points and zeros elsewhere. The image windows are multiplied together (binary multiplication of images) and the number of ones is counted. If this number is bigger than half times the number of ones in the left image window, then the corresponding points in the neighborhoods of \(p\) and \(q\) constitute shapes similar enough so that \(p\) and \(q\) can be considered as correct correspondences and satisfy the similarity criterion. Otherwise, the corresponding points \(p\) and \(q\) are rejected. This evidently removes also correct correspondences but the purpose is to filter out areas where the patterns of correspondences are clearly different in the left and right images. Small differences in the patterns are allowed to occur since the left and right cameras view the scene from different viewpoints and this causes differences to the patterns.

### 3.3 Motion estimation by tracking

The first method for motion estimation between successive image pairs is based on tracking corresponding points in the left and right images using image correlation with sub-pixel accuracy similarly as with stereo matching. The tracking to the next frame is first performed separately for the left and right images. It is then checked whether the right point candidate in the next frame equals the corresponding point of the left point candidate in the next frame. In other words, only those candidate pairs are accepted that are also corresponding points in the next frame as determined previously by stereo matching. A moderate sized search window (91 x 91 pixels in the test cases) is used and a threshold (0.7 in the test cases) is set for the correlation to be acceptable in the tracking.
The corresponding points in the left and right images yield 3-D object points for each frame. The 3-D motion between successive frames is calculated from the tracked object points using the method of singular value decomposition for rotation fitting (Kanatani, 1994).

Once the motion has been estimated, the 3-D points are transformed to the same coordinate system and triangulated into a TIN model. It appears, however, that outliers still exist in the TIN model if all the points found by stereo matching are used. The solution is to use only those points which have been successfully tracked to the next frame or the next after that. This efficiently removes false matches and produces a reliable surface.

3.4 Motion estimation by surface matching

The second method for motion estimation is based on surface matching techniques. For each frame, the stereo matching step yields a 3-D point set triangulated into a TIN model. The TIN models of subsequent frames are registered sequentially into the same coordinate system using a fast surface matching algorithm introduced by Jokinen (1998, 2000). This algorithm handles the TIN models as surfaces $z = z(x,y)$, which is a natural representation keeping in mind the orientation of the coordinate system fixed to the left camera (see Section 3.2). At each iteration, the corresponding points are determined directly on the $xy$-domain and the registration parameters are updated so that the mean of the squares of weighted distances in $z$ direction between compatible corresponding points is minimized. The weighting includes adaptive thresholds for the direction of the surface normal and the distance between corresponding points. The adaptive thresholds become tighter as the iteration proceeds according to the idea proposed by Zhang (1994).

The surface matching algorithm provides also an efficient way to eliminate unwanted observations from the original data sets. The correspondences which have non-zero weights in matching result most likely from points located on the trunk surface while correspondences with zero weight are removed when a final TIN model is triangulated from the points of the registered data sets.

4. FIELD TESTS

4.1 Allterrain vehicle sequence

A test sequence of image pairs was recorded with Basler scA1400-17fm cameras on-board an allterrain vehicle approaching a tree in a pine forest. The length of the baseline was 67.5 cm and about 100 image pairs were captured at a rate of 17 frames per second during the approach. Figure 2 shows an example image pair. A rotated edge image around the tree of interest and the same image opened with a structuring element of 15 x 1 pixels are illustrated in Figures 3a-b. Figure 3c shows the AOI determined by semi-automated outlining of the trunk edges using the first alternative approach for determination of trunk boundaries.
Figure 2. An example image pair of the sequence taken in a pine forest. Radial and tangential image distortions have been corrected.

Figure 3. a) A rotated edge image around the tree of interest in the left image of Figure 2. b) An opened edge image. c) The AOI shown as non-black area.
The result of stereo matching at the first level of the hierarchy is illustrated in Figure 4. The interest points in the left image are shown in red in Figure 4a and the corresponding points found in the right image before postprocessing are shown in red in Figure 4b. The same images after postprocessing are shown in Figures 4c and 4d, respectively. The false matches that appear in Figure 4b have been correctly removed in Figure 4d. The red points constitute similar shapes in Figures 4c and 4d, which proves that the corresponding points are most likely correct. The images were then processed at two levels of denser sets of interest points. The result at the last level of the hierarchy is illustrated in Figure 5. Figures 5a and 5b are before applying the similarity criterion and Figures 5c and 5d after applying the similarity criterion. The red points constitute similar shapes in Figures 5c and 5d as false matches especially in the lower part of the image have been correctly rejected although some true matches have been lost too. The size of the neighborhood for the similarity criterion was here 21 x 21 pixels. A smaller neighborhood would prevent losing true matches but some false matches would appear. The final surface model triangulated from the calculated 3-D points is shown in Figure 6. The point density is highest in the area towards the cameras and decreases to the sides of the trunk. The surface model covers a sector that is roughly 60 degrees of the circumference of the trunk.
Figure 5. Enlargement of the trunk at the last level of the hierarchy. a) Interest points in the left image after postprocessing without applying the similarity criterion. b) Corresponding points in the right image after postprocessing without applying the similarity criterion. c) Interest points in the left image after postprocessing including the similarity criterion. d) Corresponding points in the right image after postprocessing including the similarity criterion.

Figure 6. A TIN model of the trunk reconstructed from an image pair. The color is proportional to the $z$ coordinate.
Stereo matching between left and right images and tracking of points between successive frames are illustrated further in Figure 7. Figures 7a and 7b show a patch of the left and right images of the current frame and Figures 7c and 7d show a patch of the left and right images of the next frame. The size of the patches (91 x 91 pixels) equals the size of the search window used in tracking. The size of the matching window used in computing the image correlation is 21 x 21 pixels. The red points in the images are corresponding points found by stereo matching or tracking between successive frames. The red point in Figure 7a is an interest point. The corresponding point found by stereo matching is shown as a red point in Figure 7b. The results of tracking in the left and right images are shown as red points in Figures 7c and 7d, respectively. It is finally verified that the red points in Figures 7c and 7d are also corresponding points found by stereo matching.

A sequence of 11 image pairs was processed using the tracking-based method for motion estimation. A TIN model of the reconstructed trunk is shown in Figure 8. The model in Figure 8a has been produced using all the 3-D points found by stereo matching after postprocessing while the model in Figure 8b has been generated using only those 3-D points which have been successfully tracked from the previous frame. It can be realized that the first model contains many outliers while the second one is more reliable.
4.2 Harvester sequence

In the second test case, the Basler cameras were mounted on-board a harvester just above the grab. The optics of the cameras was improved by using high quality Schneider Cinegon lenses. An image pair with radial and tangential image distortions corrected is shown in Figure 9. Figures 10a-b illustrate the AOIs extracted semi-automatically using the second alternative approach for determination of trunk boundaries. The gray levels have been scaled so that levels between 85 and 255 have been set to zero and levels between 1 and 84 multiplied by three.

![Figure 9. An image pair of the sequence taken with cameras on-board a harvester. Radial and tangential image distortions have been corrected.](image)

![Figure 10. AOIs in the a) left and b) right image. The gray levels have been scaled. c) Interest points in the left image after postprocessing. d) Corresponding points in the right image after postprocessing.](image)
Figures 10c-d and 11a-b illustrate the corresponding points found by stereo matching after postprocessing at the last level of the hierarchy. All the correspondences appear on the trunk after successful postprocessing in Figures 10c-d, although the AOI in the left image is broader than optimal and interest points initially appear within the whole area. The red points constitute similar shapes in Figures 11a-b, which indicates that the corresponding points are most likely correct.

A TIN model of the trunk reconstructed from a sequence of 9 image pairs using the tracking-based method for motion estimation is shown in Figure 12. There exist outliers in the model triangulated from all the points obtained from stereo matching after postprocessing in Figure 12a while most of the outliers are removed when only those points are used that have been successfully tracked from the previous frame in Figure 12b.

The motion estimation based on surface matching is illustrated in Figure 13. Figure 13a shows two TIN models triangulated from the 3-D points of consecutive frames. Since the sensor has moved between the frames, the TIN models do not match properly when plotted in the same coordinate system. The result after surface matching is shown in Figure 13b. It can be realized that the surfaces are perfectly aligned and outliers removed when only points within the matching area are used in surface model generation. The matching area is given by points where the correspondences are compatible and thus have non-zero weights in surface matching.

A TIN model of the trunk triangulated from the points of 9 sequentially registered consecutive 3-D data sets is shown in Figure 14. The model in Figure 14a has been generated using all the points of the data sets while the model in Figure 14b contains only points located in the matching areas. The removal of unwanted points is clear. A qualitative comparison between the models in
Figures 12b and 14b shows that the surface matching based method for motion estimation provides a more detailed surface with a higher point density than the tracking-based method.

The stereo matching and motion tracking are computationally heavy. We have implemented the algorithms using matlab software and used fast convolution operations for computing the image correlations. The computations have been performed in a HP CP4000 BL ProLiant supercluster. However, it takes for example about 165 seconds to find the corresponding points in the right image for 13700 interest points in the left image at the first level of the hierarchy. At subsequent levels, the computing times are shorter due to reduced search space. For example, it takes about 109 seconds to find the corresponding points for 30446 interest points with 27 x 61 pixels search window. Moreover, it takes about 383 seconds to track 11443 corresponding points to the next frame. Consequently, the stereo matching and motion tracking methods are currently suitable only for off-line purposes. However, the motion estimation based on registration of the surface models is fast. It takes only 2-3 seconds to match the surfaces shown in Figure 13 excluding the surface normal estimation, which takes about 13 seconds. The surface models contain more than 14000 points and the number of corresponding points having a positive weight is 10855 after registration.

Figure 12. A TIN model of the trunk triangulated from a sequence of nine image pairs using a) all the points b) only those points which have been successfully tracked from the previous frame. The color is proportional to the z coordinate.

Figure 13. Registration of two consecutive surface models. a) The original surface models before registration. b) The surface models triangulated from points within matching areas after registration. The color is proportional to the z coordinate.
5. CONCLUSIONS

The paper applied stereo matching, motion tracking, and surface matching techniques for digitizing the 3-D shape of a trunk from a sequence of image pairs in a challenging forest environment. Stereo matching worked well but required postprocessing for removal of false matches. A novel similarity criterion was introduced to deal with false matches. Two methods for motion estimation were considered, namely, tracking of corresponding points in the left and right images and sequential registration of successive 3-D data sets using surface matching techniques. Both methods were successful although the method based on surface matching provided a more detailed model of the trunk. The surface matching also performed computationally much faster than the correlation-based tracking of corresponding points.

Our plans for future research include extraction of geometric information from the TIN model of the trunk. For example, the diameter of the trunk as a function of height might be estimated by dividing the point set into narrow horizontal slices perpendicular to the main direction of the trunk and fitting cylinders to the points within each slice.

6. REFERENCES


Gilbert, S., Laganiere, R., and Roth, G., 2005. Stereo motion from feature matching and tracking,

Jokinen, O., 1998. Area-based matching for simultaneous registration of multiple 3-D profile

Jokinen, O., 2000. Matching and Modeling of Multiple 3-D Disparity and Profile Maps, Doctoral

Machine Intelligence, Vol. 16, No. 5, pp. 543-549.

Kivinen, V.-P., 2004. A genetic algorithm approach to tree bucking optimization, Forest Science,
Vol. 50, No. 5, pp. 696-710.

scanning laser range finders in forest harvesters, The IASTED International Conference on
Robotics, Telematics and Applications (RTA 2009), Beijing, 12-14 October, 2009, 6 p.

Molton, N. and Brady, M., 2000. Practical structure and motion from stereo when motion is

Moyung, T.J. and Fieguth, P.W., 2000. Incremental shape reconstruction using stereo image
September, 2000, Vol. 2, pp. 752-755. IEEE.

Pan, B., Xie, H., Xu, B., and Dai, F., 2006. Performance of sub-pixel registration algorithms in

Pfeifer, N., Gorte, B., and Winterhalder, D., 2004. Automatic reconstruction of single trees from
terrestrial laser scanner data, International Archives of the Photogrammetry, Remote Sensing and

Salgado, A., Sánchez, J., 2005. 3D geometry reconstruction from a stereoscopic video sequence,
Second International Conference Image Analysis and Recognition, Kamel, M., Campilho, A.

Thies, M., Pfeifer, N., Spiecker, H., and Gorte, B., 2004. Three-dimensional reconstruction of
stems for assessment of taper, sweep and lean based on laser scanning of standing trees,

Tomasi, C. and Kanade, T., 1992. Shape and motion from image streams under orthography: a

scenes, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,
Madison, June 2003, pp. 367-374. IEEE.

ACKNOWLEDGMENTS

We would like to thank our colleagues in the Department of Automation and Systems Technology at Helsinki University of Technology for providing the test imagery. Tomi Rosnell and Heikki Kauhanen are acknowledged for calibration of the cameras. The computations were performed in HP CP4000 BL ProLiant supercluster of CSC, the Finnish IT center for science. The research was funded by Tekes, the Finnish Funding Agency for Technology and Innovation, and by Finnish companies including Metsäliitto Group, UPM-Kymmene Oyj, Stora-Enso Oyj, Ponsse Oyj, Vansco Electronics Oy, John Deere Forestry Oy, Komatsu Forest Oy, and Logset Oy.