The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Motion Vision Based Structure Estimation in Forest Environment

Jakke Kulovesi

Abstract— **Motion vision can be used to determine world structure from a video sequence. In harvester machine automation, the potential is that trees could be measured from a distance. Based on the measurements, tree cutting could be optimized and harvester automation increased, resulting in higher resource utilization efficiency. However, a natural environment poses challenges to any computer vision task. This paper presents computer vision algorithms that are applied to a forest environment. The results show that dense optical flow can be computed from a real-world forest data accurately enough as to enable instantaneous dense structure estimates of the visible image scene.**

I. INTRODUCTION

According to a long-term vision, forest harvester machine operation could be both automatized and optimized using advanced measurement technology. Essential tree parameters like trunk dimensions, shape and branches could be measured while approaching a tree to be felled. The possibility of measuring the tree before it has been felled is a novel one. Presently, the mechanical measurements performed in the harvester head are obtained too late for true cut-location optimization.

The work presented in this paper is a part of a project that aims to evaluate and compare different sensory options for a harvester. This paper concentrates on the motion vision approach. Two main benefits of motion vision based structure estimation are: the requirement of only one camera (combined with some absolute distance measurement for fixing scale) and small discrepancy between consequent images. Small discrepancies in image data are advantageous for optical flow computation but on the other hand increase noise in depth reconstruction. Other sensory options such as stereo vision and laser scanners are part of the project but not discussed here. Instead, this paper proposes a structure from motion estimation procedure that is intended to work outdoors and produce high resolution dense depth estimates from image pair data.

Optical flow has been under extensive study for decades, resulting into several families of optical flow computation methods. In the following, focus in on dense optical flow estimates. Objective evaluation of the different methods has been attempted [1], [2] but has also proven to be difficult. For example in this work, the aim is to measure trees, in which case average error over the whole image is secondary to the error near critical object boundaries. Fleet and Jepson propose a phase correlation method in [3] that achieved good results in the benchmark [1] but for the price of relatively low estimate density. Promising enhancements have been suggested to phase correlation image registration [4], [5] but to the author's knowledge, these have not been integrated to the dense optical flow estimation task. Alvarez et al suggest in [6] a method claiming improved results against the benchmark competition but comparison with e.g. [3] is left out due to the restriction to 100% dense maps. A later work by Alvarez et al proposes a method utilizing symmetry in optical flow [7], attempting to address the occlusion problem but primarily achieving only detection of occlusion.

This article, however, pursues an application specific solution to the task by using a block matching i.e. correlation approach. The work is continuation to the works of Anandan in [8], and Singh, in [9]. Essential differences and task-specific improvements to the above have been done. Common features include hierarchical pyramid processing, window correlation and integer block search. However, in this work, no explicit smoothing is performed to obtain sharp object boundaries. No pyramid overlapping as in [8] is required either due to the use of adaptive filtering. The method in this paper differs from the above also in the addition of both sub-pixel refinement and filtering, implemented *inside* the hierarchy. This will allow the benefits of the both improvements to propagate within the hierarchical processing, answering to the challenges of the pursued application. In addition, block cropping is explicitly presented to allow for using the full image resolution in computation. Algorithm parameters have also been adjusted according to empirical in-site experiment data to suit the application needs. Main aim of this work was to evaluate if motion vision structure estimation could be applied to a real world measurement task in a challenging forest environment, where the main challenges supposedly relate to the optical flow calculation.

Structure from motion can be pursued in various different ways. A central problem is determining the motion, or change in pose for the camera. When both optical flow information and the pose change are known, a linear triangulation procedure can be utilized. Kalman filtering approach is popular in the literature, e.g. [10]. For images in the megapixel range, such approach would hardly be possible to implement in real-time. A more speed oriented integrative approach is given in [11] but the amount of high frequency noise in the camera motion makes continuity assuming filtering approaches infeasible. Instead, instantaneous independent motion estimation is required. Linearized formulations for the problem have been attempted in [12],

This work was supported by TEKES, the Finnish Funding Agency for Technology and Innovation, and participating companies. The author wishes to thank all the reviewers for their very helpful and detailed feedback.

J. Kulovesi is with the Department of Automation and Systems Technology, Helsinki University of Technology, PO Box 55000, 02015 TKK, Finland jakke.kulovesi@tkk.fi

978-1-4244-3804-4/09/$25.00 ©2009 IEEE

[13] but their precision was found insufficient. Later in this paper, a straightforward solution is presented to obtain a reasonable base estimate for motion without unnecessarily compromising quality.

II. Measurement Platform and Setup

Video data was collected using a Honda all terrain vehicle (ATV). The ATV carries batteries, a computer and the cameras attached to a camera stereo stand. While the data collection was performed with a stereo camera system, data from only one camera was used in this work as explained in the introduction. Fig. 1 shows the ATV sensor platform setup.

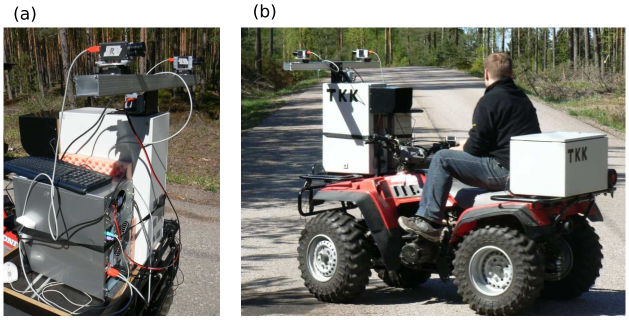

Fig. 1. Honda all terrain vehicle (ATV) sensor platform. (a) Cameras are attached to a PC computer on board the ATV. (b) Honda ATV operated by a researcher.

A. Video Hardware

Video hardware used in the tests consisted of an industrial monochrome firewire camera, product name Basler Scout, with a wide angle lens manufactured by Tamron to provide with a sufficiently large field of view (72° in vertical orientation) to allow measuring a tree from a distance. There were several criteria behind the choices. First, the ability to output uncompressed video data to avoid compression artifacts. Second, a relatively high resolution to facilitate measurement precision, i.e., 1.4 MP (megapixel) images. Third, a sufficient frame rate to ensure that relative changes between video frames are not too large, the Scout giving 17.4 fps at maximum.

B. Video Data Collection

The data used in this paper was collected by driving approaches toward trees from approximately 5 meters with a Honda ATV and recording the approach to a video sequence. Due to uneven ground and the ATV itself, the sequences contain a high amount of uncontrolled camera motion. Later experiments have verified that data obtained from a real harvester head behaves more smoothly.

C. Test Location and Conditions

The data collection was performed in a forest in Vantaa, Finland. The forest had a low density of trees, predominantly pines, with a low amount of undergrowth. The pines were of relatively the same age and had little branches on the lower parts of the trunks.

Conditions on the day of the data collection experiment were challenging due to half-cloudy conditions and moderate wind. Problematically high dynamic range of illumination results from the intensity variation between bright sunlight and cloud shadows. The wind on the other hand causes the forest to move (including shadow movement), violating the usual rigidity assumption.

III. Problem Framework

In this work, there are three main variable groups of interest relating to motion vision. Two of these are objective properties of the world: the *motion parameters* and the *scene structure parameters*. The third is an artificial, more ambiguous variable group – *optical flow* – that describes apparent visual motion in the image sequence.

The camera model used in this paper is the *perspective projection model*. Fig. 2 illustrates the camera coordinate system. Throughout this paper, in image pair data, the frame is fixed to the first (preceding) image corresponding camera frame.

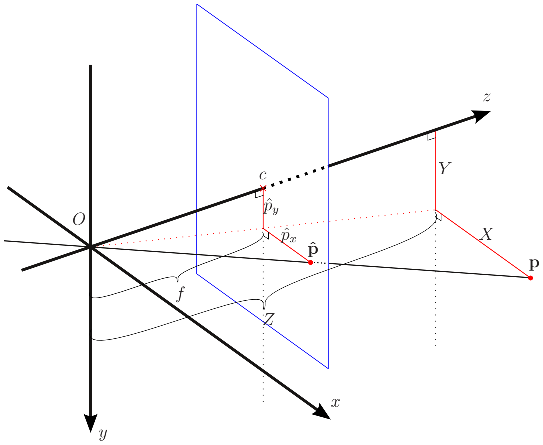

Fig. 2. The center of the virtual image plane is denoted by c. X, Y, and Z are the respective x-,y-,and z-coordinates of the scene point. $\hat{p}_x$ and $\hat{p}_y$ are the coordinates of the projected point $\hat{\mathbf{p}}$. The camera points toward the positive z-axis.

Because perspective projection model fails to accurately represent a true mapping taking place in the camera optics, the camera was additionally calibrated against typical radial and tangential distortions.

Camera motion is represented using an Euler angles rotation matrix formalism, expressing a precise discrete transformation in an equation, (assuming $\Delta t = 1$ for convenience),

$$\mathbf{p}^2 = \mathbf{R}_{\alpha_y,\alpha_z,\alpha_x}(\mathbf{p}^1 + \mathbf{t}), \tag{1}$$

where $\mathbf{p}^1$ is a scene point before motion and $\mathbf{p}^2$ is the same point location after the motion. The translation motion $\mathbf{t}$ and the rotational motion represented by the rotation matrix $\mathbf{R}_{\alpha_y,\alpha_z,\alpha_x}$ define the motion parameters. The sampling period Δt can be chosen arbitrarily because time dimension is irrelevant in this task - only the acquired change in pose matters, regardless of the time needed for the pose change.

IV. AN ADVANCED BLOCK MATCHING ALGORITHM FOR OPTICAL FLOW

Optical flow is a (dense) estimate of the true projected motion field. In this paper, a motion field (and consequently optical flow) represents discrete motion.

The overall algorithm work flow is presented below:

1) Preprocessing.
2) Hierarchical search loop.
    a) Integer search.
    b) Sub-pixel refinement.
    c) Active pixel shifting (adaptive filtering).
    d) Hierarchy propagation.

The algorithm takes an image pair as its input. The image pair consists of a *source image* that represents the present and a *target image* that represents the future. First, an *image pyramid* is constructed for both of the images. Then, the algorithm works hierarchically with integer search, sub-pixel refinement and filtering procedures at each pyramid level.

A. HIERARCHICAL APPROACH

In a hierarchical approach, an image pyramid is first constructed from each full-resolution image. The pyramid is a multi-resolution representation of the original image [14]. A factor of 2 is used in this work for decimation and resampling. Using factor 2 is both mathematically convenient and efficient and thus commonly used in the literature.

Hierarchical processing starts from the pyramid top, from the most coarse hierarchy level and proceeds downward, one layer at a time. Parameters involved in the search are the integer search range r and the block dimensions m and n. Based on empirical experimentation, block dimension values between 3 and 5 provide the best compromise between noise and sharpness. In forest where vertical edges are predominant, taller than wide blocks are preferred. At the highest pyramid layer, the search range has to be calculated so that it matches the maximum detectable motion desired for the task. For the subsequent pyramid levels, the range can be kept small, typically from 2 to 4. Any obtained results – and errors – in the hierarchical processing will propagate and spread during the pyramid processing but the severity of error propagation is greatly diminished with a filtering process introduced later in this paper.

B. INTEGER FULL SEARCH

Integer pixel search is a motion estimation search process that tries to find the best integer-precision candidate match between blocks from two images. An image block is a windowed rectangular-shaped area of its parent image.

The search process evaluates how well two blocks, a fixed source block from the source image and a varying-position block from the target image, match. The match "goodness" is evaluated with an error-metric that is minimized by the best match. The error metric used in this work is the normalized mean square error

$$E_{\text{MSE}} = \frac{1}{N}\sum_{i}^{N}(x_i^1 - x_i^2)^2, \tag{2}$$

where vectors $\mathbf{x}^1, \mathbf{x}^2$ are composed of image block data, and N is the length of these vectors. Normalization is used to ensure that match errors between different sized blocks are comparable.

A limited-range full-search is implemented to emphasize the quality of the search results. Alternatively, a heuristic search pattern could be utilized but doing so would sacrifice matching quality. The higher computational cost of running a full search can be compensated with parallel computing.

C. IMAGE BORDERS

When block matching happens in the "center" areas of an image, the image borders can be safely ignored. Close to the image borders, however, there are several issues. First, the block area should not overlap the image area. Second, the blocks need to share the same dimensions. Third, the correct motion may be directed outside of the image borders. Trivially restricting the computation to the image center is not feasible, especially with hierarchical processing as that would loose a significant part of the image data.

A solution to the first two problems, and a partial solution to the third consists of three phases. First, the source image block is cropped to image borders. Second, the shifted target block is additionally cropped against the image borders - this is the final target block. Third, the target block is reverse-shifted back and again cropped against the image borders - this is the final source block. The last step is required to ensure identical block sizes. Additionally, when the width or the height of a block is cropped to less than 2, matching is aborted. A more detailed visual representation of the cropping procedure is available at [15].

D. SUB-PIXEL REFINEMENT

In order to attain sub-pixel precision with block matching, image block interpolation is necessary. Linear interpolation is used to limit computational cost. For example, bicubic interpolation could also be used with the cost of more computation time. In this interpolation task, scale, orientation, and shape are maintained, only translation occurs. Thus, image block interpolation can be performed with only four weighted matrix additions [15].

The sub-pixel heuristic search refines the integer precision results. Different sub-pixel candidate translations are explored heuristically with the help of predictors to restrict computational cost. The used sub-pixel precision was $1/16$ – probably pushing the limit of the obtainable information in the image data.

An outline of the proposed sub-pixel precision refinement search is presented below.

1) Consider 6 predictors at the search space. Compute the error values for each and select the minimum error predictor. 5 of the predictors are fixed spatial predictors covering coarsely the search space and one is a continuity predictor that uses the location of the previously calculated (neighboring) location.

2) Perform an optimized simple gradient descent search from the selected predictor. Stop the search if a stop condition is fulfilled while performing the search.
3) Return the sub-pixel block translation at the stop location as the optimum sub-pixel refinement to the integer search result.

Predictors are heuristic guesses for search locations that are used to improve efficiency and quality of e.g. motion estimation algorithms [17]. Effectively, they restrict the amount of error evaluations required by the gradient search by giving a good starting location.

E. ADAPTIVE FILTERING

This far, block matching results have been associated with the block center pixel. However, shifting the associated pixel can provide benefits by avoiding overlapping object boundaries, i.e. occlusion areas. Ong and Spann use a more basic block shifting procedure in [18] but the filtering process used in this work is specifically implemented to benefit from the existing dense flow field.

It is assumed that both dense optical flow and corresponding match error measures exist. Consequently, all the off-center shifted correlations are present in the computed optical flow. This observation allows for implementing an optimal block shifting algorithm that works as a post-processing filter for the earlier acquired optical flow field, with little additional computational cost.

In the following, a three-criteria adaptive filter is proposed. A minimum (2) error yielding shift is searched from the candidate shifts that pass the criteria for the whole image area. The three parameters implementing the criteria are:

1) **Relative error threshold**, t_r
The relative error threshold defines the minimum relative improvement that has to be obtained by a shift – if the relative change in error values is below the threshold, no filtering will be done.
2) **Minimum error threshold**, t_m
The minimum error threshold defines the minimum allowed error for the initial flow – if the error value is below the threshold, no shifting will be done.
3) **Flow magnitude change threshold**, t_c
The flow magnitude change threshold defines the minimum required change in the flow vectors for the shift to be performed – if the norm of the change is less than the threshold, no shift will be performed.

The implemented optical flow computation uses this filtering at the end of each hierarchy level. Consequently, the benefits propagate with the hierarchical processing, greatly reducing the effects of false block matches. The use of the block match error measure (2) in the shift algorithm has some shortfalls. The match error metric biases low-texture areas, giving them smaller error values. To compensate for the weakness, simple thresholding is used in the main integer search algorithm to abort the search in blocks that are either too dark or bright, i.e., trivially containing low texture.

V. MOTION ANALYSIS AND STRUCTURE FROM MOTION

This section presents a straightforward approach to using the acquired dense optical flow field for constructing an instantaneous 3D structure estimate of the visible image area, i.e., a depth map. First, camera's relative pose change between images is estimated. Second, the relative pose information and the optical flow are used to compute a dense depth map.

A. CAMERA RELATIVE POSE ESTIMATION

Although the aim is to compute a dense structure estimate, sparse data set can be used for estimating motion. The benefits of using only partial data for the motion estimation are numerous, most importantly: computation is much lighter and quality of the data can be selectively controlled. In this paper, sparse sets of points have been selected manually at this stage of algorithm development, but several automatic methods exist and yet an additional automatic method is under development for the problem but has not yet been thoroughly tested. Shi and Tomasi propose a method for finding good features to track in [19] and Lowe implements an automatic method for extracting scale-invariant keypoints in [20]. However, these methods are not specifically designed to predict quality of a dense optical flow field that is used in this work because the methods implement their own feature matching and tracking algorithms. For this reason, more work would be required to integrate one of these methods to this work. Thus, a more simple solution specific to this problem might be easier to achieve and this work is still in progress.

Pose estimation in this work was done using numerical optimization for several reasons. Due to the problem nonlinearity, no closed form solution exists. Several approximated solutions from the literature such as [13] or iterative application of [21] with triangulation have been attempted but the results have proven unsatisfactory with respect to the precision requirements of this work. Numerical optimization, on the other hand, can be performed without approximations. However, proper care should be executed with respect to the typical problems such as local minima and unguaranteed convergence. Fortunately, advanced optimization algorithms are well-researched and readily available for use.

A flow fit error measure was used in the optimization as the function to be minimized. The flow error measure calculates the difference between expected motion field and the measured optical flow. First, the input data points $\hat{\mathbf{p}}^i$ are back-projected out from the image plane by using depth information Z^i and transformed according to (1)

$$\mathbf{p}^i = \mathbf{R}_{\omega_y,\omega_z,\omega_x}([x^iZ^i, y^iZ^i, Z^i]^T + \mathbf{t}).$$

The transformed points are then projected to the image plane $\hat{\mathbf{p}}^i = \mathbf{p}^i/p^i_z$, by dividing the vector with its z-axis coordinate value. The motion field $\Delta\hat{\mathbf{p}}^i$ is simply the difference $\Delta\hat{\mathbf{p}}^i = [\hat{p}^i_x, \hat{p}^i_y]^T - [x^i, y^i]^T$.

Because an optical flow field approximates the motion field, an error measure

$$E_{\text{f-fit}} = \sum_{i=1}^{N} \left(|u^i - \Delta\hat{\mathbf{p}}_x^i| + |v^i - \Delta\hat{\mathbf{p}}_y^i| \right) \quad (3)$$

can be defined, where u is the x-direction optical flow and v is the y-direction optical flow.

In addition, a scale regularization term was used to address the inherent scale invariance in a monocular vision system. Regularization can be performed by fixing the translation vector's norm to unity against a squared penalty

$$E_{\text{norm}} = (\sqrt{t_x^2 + t_y^2 + t_z^2} - 1)^2. \quad (4)$$

In the final optimization process, sum of $E_{\text{f-fit}}$ and E_{norm} is minimized.

B. STRUCTURE FROM MOTION

Once both the optical flow and the relative camera motion are known, the structure can be solved linearly. The following presents a "direct" solution to the problem.

In the direct approach, the coordinate system is fixed to camera coordinates of the source image. Two points, $\mathbf{p}^1$ and $\mathbf{p}^2$, represent the object location with respect to two viewpoints. The principle of the direct approach is that $\mathbf{p}^2$ can be represented in two ways using $\mathbf{p}^1$: first in the image plane, where $\hat{\mathbf{p}}^2$ is $\hat{\mathbf{p}}^1$ added with the motion field $\mathbf{m}$ estimate (optical flow)

$$\mathbf{p}^2 = Z_2\hat{\mathbf{p}}^2 = Z_2\hat{\mathbf{p}}^1 + [u, v, 0]^T.$$

On the other hand,

$$\mathbf{p}^2 = \mathbf{R}(\mathbf{p}^1 + \mathbf{t}) = \mathbf{R}(Z_1\hat{\mathbf{p}}^1 + \mathbf{t})$$

according to (1). Eliminating $\mathbf{p}^2$, denoting $\mathbf{z} = [Z_1, Z_2]^T$ and reshaping results in

$$\begin{bmatrix} -\mathbf{R}\hat{\mathbf{p}}^1 & \hat{\mathbf{p}}^1 + [u, v, 0]^T \end{bmatrix} \mathbf{z} = \mathbf{R}\mathbf{t}.$$

The above is an overdetermined linear system of equations. Denoting the left side matrix of the above equation as $\mathbf{A}$, and applying a pseudoinverse solution to minimize the squared error results in the depths, assuming rank 2 in matrix $\mathbf{A}$

$$\mathbf{z} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}\mathbf{t}. \quad (5)$$

VI. RESULTS

Fig. 3 shows an example visualization of an obtained optical flow field. The figure shows that several tree shapes are clearly distinct in both the component fields. This is a good result in itself as far as the difficult problem of object segmentation could be solved based on this data. In addition, the figure shows some typical errors found in optical flow results. The noisy areas in the top region of the image correspond to areas of uniform open sky where plain block-matching will fail. Some other noise in the results can be explained by the moving shadows in the data and other similar assumption violations. All in all, the dense optical flow field is smooth and the object borders are relatively sharp, even in the full resolution pixel detail level (which is not visible from the reduced size images). At this stage, the computation takes far too much time for real-time application but on the other hand no optimizations have been performed nor has the inherent parallel computation potential been utilized.

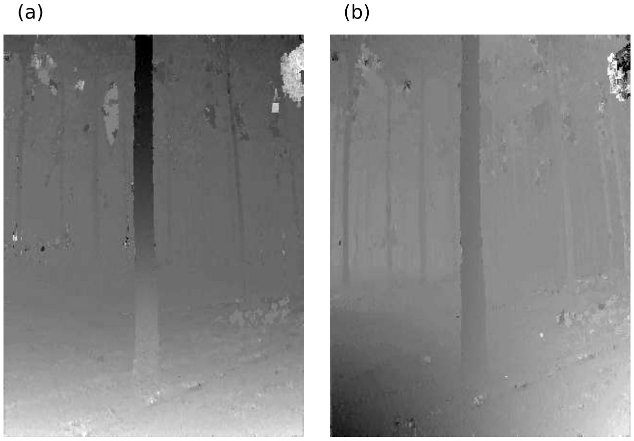

Fig. 3. Dense optical flow field. (a) y-axis direction component as an intensity plot. (b) x-axis direction component as an intensity plot. A middle gray color corresponds to zero optical flow, bright color to a high positive optical flow component, and dark black color to a large negative optical flow.

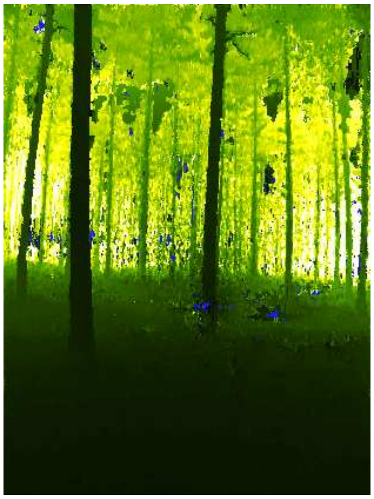

Fig. 4. Depth map corresponding to an image pair. The brighter the color, the more distant the image location is. Due to the scale invariance, the depth map information is relative.

Motion estimation with numerical optimization proved to be robust without data-specific parameter tuning. Error minimization converged in all test data cases and the results were sensible. The calculated motion estimates were used to compute dense depth maps for the image pair data. One example depth map that was obtained is presented in Fig. 4. Because the depth estimate relies strongly on the optical flow, the results reflect the quality of the optical flow estimates – assuming that the motion estimation succeeded. The relative depth map alone does not allow absolute measurements to be performed but it demonstrates that the general 3D forest structure can be reconstructed from an image pair using the methods presented in this paper. In general, the instantaneous structure reconstructions work well if the motion in the image pair provides with sufficient depth information through translational motion. In the cases where the translational motion was insufficient, the depth estimation naturally fails. In this sense, the instantaneous representations are only a starting point for building integrated 3D reconstructions that complement the information from different time instants.

Evaluation of results at this point of the work is unfortunately primarily qualitative. No ground truth optical flow is yet available from the forest scenes due to the lack of sensors capable of both measuring very small movements with high precision and the lack of full 3D scans of the forest. For motion estimation, affordable inertial measurement units are not sufficient for direct estimation of ground truth motion. Similarly, existing 3D laser scans from the approaches are under work and their resolution is very limited due to the low measurement times allowed. In the future, results from the parallel measurement techniques will be utilized to give a much more quantitative evaluation of the strengths and weaknesses in each approach. However, at this time the results do encourage further effort in the motion vision approach.

VII. CONCLUSIONS

This paper has presented a sequence of methods that produce a structure estimate of a forest scene based on two consequent video frames. The combined results from the tasks of determining optical flow, estimating discrete motion between image frames and constructing a full depth map of the visible scene have proved promising.

Lots of room for further work remains in all the submethods. Optical flow computation suffers in large areas of non-trivially low texture information. The computational cost of the calculation is high and parallel implementation should be pursued in the future. The motion estimation presented here is not sufficient for end application use. Automatic point selection will have to be incorporated to the method sequence and the motion estimation should preferably be changed to a RANSAC scheme based approach that should prove higher quality and robustness without the inconvenience of the numerical optimization.

On the whole, the results this paper have presented a solid basis for additional research. The approached tree trunk is clearly segmented from the background and other trees. In addition, more distant objects in the forest environment can be identified from the derived depth maps. From an image object segmentation point of view alone, the results are good, providing a possible solution to a generally very difficult computer vision problem. Many difficulties with using computer vision in natural conditions have been answered. In essence, the results have proved that it is reasonable to apply motion vision analysis to video data taken in uncontrolled and challenging forest conditions.

REFERENCES

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[2] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow," *Computer Vision and Image Understanding*, vol. 84, no. 1, pp. 126–143, 2001.

[3] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[4] J. Ren, T. Vlachos, and J. Jiang, "Subspace extension to phase correlation approach for fast image registration," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I –481–I –484, 2007.

[5] Y. Keller and A. Averbuch, "A projection-based extension to phase correlation image alignment," *Signal Processing*, vol. 87, no. 1, pp. 124–133, 2007.

[6] L. Alvarez, J. Weickert, and J. Sánchez, "Reliable estimation of dense optical flow fields with large displacements," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 41–56, 2000.

[7] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez, "Symmetrical dense optical flow estimation with occlusions detection," *International Journal of Computer Vision*, vol. 75, no. 3, pp. 371–385, 2007.

[8] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, no. 3, pp. 283–310, January 1989.

[9] A. Singh, "An estimation-theoretic framework for image-flow computation," *Computer Vision, 1990. Proceedings, Third International Conference on*, pp. 168–177, December 1990.

[10] Y. Xiong and S. A. Shafer, "Dense structure from a dense optical flow sequence," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 222–245, 1998.

[11] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, 2002.

[12] T. Mukai and N. Ohnishi, "Recovery of motion and structure from optical flow under perspective projection by solving linear simultaneous equations," in *ACCV '98: Proceedings of the Third Asian Conference on Computer Vision-Volume II*. London, UK: Springer-Verlag, 1997, pp. 583–590.

[13] D. J. Heeger and A. D. Jepson, "Subspace methods for recovering rigid motion i: Algorithm and implementation," *International Journal of Computer Vision*, vol. 7, no. 2, pp. 95–117, 1992.

[14] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, 1984.

[15] J. Kulovesi, "Structure from visual motion in forest environment," Master's thesis, Helsinki University of Technology, Department of Automation and Systems Technology, Espoo, Finland, 2008.

[16] X. Li, E. Li, and Y.-K. Chen, "Fast multi-frame motion estimation algorithm with adaptive search strategies in h.264," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 3, pp. iii–369–72, May 2004.

[17] A. Tourapis, O. Au, and M. Liou, "Highly efficient predictive zonal algorithms for fast block-matching motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 10, pp. 934–947, October 2002.

[18] E. P. Ong and M. Spann, "Robust optical flow computation based on least-median-of-squares regression," *International Journal of Computer Vision*, vol. 31, no. 1, pp. 51–82, 1999.

[19] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, June 1994. [Online]. Available: citeseer.ist.psu.edu/shi94good.html

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, vol. 20, 2003, pp. 91–110. [Online]. Available: citeseer.ist.psu.edu/lowe04distinctive.html

[21] X. Zhuang, R. Haralick, and Y. Zhao, "From depth and optical flow to rigid body motion," *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*, pp. 393–397, June 1988.