Multiple View Motion Tracking of Gridded Surfaces using Topological Structure

M. H. Kinsner[1], T. S. Kenyon[2], D. W. Capson[3] and A. D. Spence[4]

[1]McMaster University, kinsnemh@mcmaster.ca
[2]McMaster University, kenyot@mcmaster.ca
[3]McMaster University, capson@mcmaster.ca
[4]McMaster University, adspence@mcmaster.ca

ABSTRACT

This paper describes a novel algorithm for interframe motion tracking of gridded surfaces across close range video sequences. The topological approach exploits the regular grid structure, simplifying feature extraction requirements, and reducing dimensionality of the projective tracking problem from eight to two. Multiple view grid asymmetry is implemented using loosely defined tracking fiducials that do not need to be explicitly detected or measured. Results from an experimental setup validate the accuracy and robustness of the algorithm. Timing results demonstrate throughput greater than the 15 frames per second camera data acquisition rate.

Keywords: computer vision, interframe motion, grid tracking, topological structure.
DOI: 10.3722/cadaps.2013.221-229

1 INTRODUCTION

To provide a reference structure not otherwise found in a natural scene, vision-based applications require a reference geometry such as a gridded surface [1]. Furthermore, when a camera frame encompasses only a small surface region, the corresponding grid features must be tracked across multiple views. In practical applications such as metal forming strain analysis [2], conventional solutions result in a parameter space that is unmanageably large.

Herein, we report a novel interframe motion tracking solution, designed to robustly determine the trajectory of line-based grids as a camera moves across a target surface. A topological approach is chosen to exploit the regular grid structure, reducing dimensionality of the interframe transform, and creating a parameter space small enough to be searched exhaustively. To overcome interframe motion ambiguity, grid pattern asymmetry is implemented using fiducial marks added between the grid lines. These fiducials need not be defined precisely, and in this case are hand drawn marks placed randomly

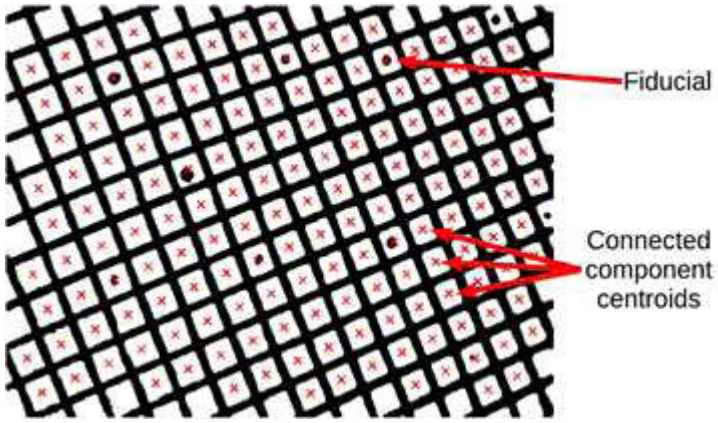


Fig. 1: Thresholded grid image with the centroids of connected components denoted by '×' marks.

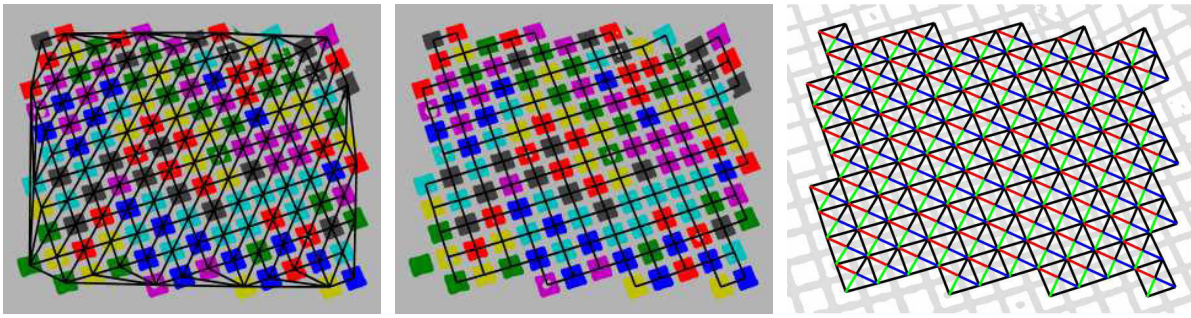
between the grid lines. The implemented algorithm achieves throughput greater than the 15 frames per second camera data acquisition rate.

The remainder of the paper is organized as follows. Section 2 reviews previous work. Section 3 describes topological construction, including image thresholding, connected region labeling, Delaunay triangulation, topological merging, and consistency verification. In Section 4, topological tracking is described. Section 5 presents experimental results, followed by conclusions in Section 6.

2 PREVIOUS WORK

Previous work on motion tracking of gridded surfaces is limited to applications where the target is smaller than the camera field of view, allowing boundaries to be detected and used for absolute grid positioning. Solutions typically involve either block-matching approximations [3] or explicit detection of features [4], followed by putative matching and Random Sample Consensus (RANSAC) based refinement [5]. From this correspondence set, the interframe transform (homography) is computed. The robustness of such techniques is challenged by difficult imaging conditions, such as varying illumination, focus and camera range. Moreover, the tracking search spaces in practical applications can become large, necessitating heuristic or other adjustable search space reductions. Additional challenges are imposed under close range imaging, as required by many applications of hand-eye and camera calibration [6][7].

The solution proposed herein combines grid topology with fiducial structure, and thereby removes the need to explicitly detect fiducial features. This relaxes the challenging requirement for a robust detector and interframe matching system. Fiducial features are measured during the topology construction process, and are represented as active pixel ratios within thresholded connected components. No explicit feature detector is required to locate the fiducials, and the detection is robust to changing lighting and scale conditions. By building a topological representation of the grid structure [8], dimensionality of the problem is reduced from a general homography (eight continuous degrees of freedom (DOF) [5]) to discrete shifts in the directions of the grid axes (two discrete DOF). This reduction makes an exhaustive search of the possible grid shifts feasible. Subsequent sections present the proposed algorithm details.



(a) (b) (c)

Fig. 2: (a) Delaunay triangulation of connected component centroids; (b) a set of complete quadrilaterals connecting grid units in the principal grid directions; (c) vertex orientation relative to principal grid directions.

3 TOPOLOGICAL CONSTRUCTION

Starting with a raw frame from a grayscale video sequence, the following algorithm sequence is applied. First, the image is thresholded to obtain a binary output that segments the black grid lines and fiducials from the (white) remainder of the image. Herein, a local adaptive thresholding technique based upon the Shafait method [9] was selected. This technique uses a varying threshold level to compensate for spatially changing illumination or reflections in different image regions. A morphological close operator is then applied to remove orphan black pixel noise and repair fragmented grid lines.

A connected component labeling algorithm, that attaches a unique identifier to each contiguous white area, is then executed (Fig. 1). The centroid of each white area is used as vertex coordinates (x, y) for the topological data structure construction.

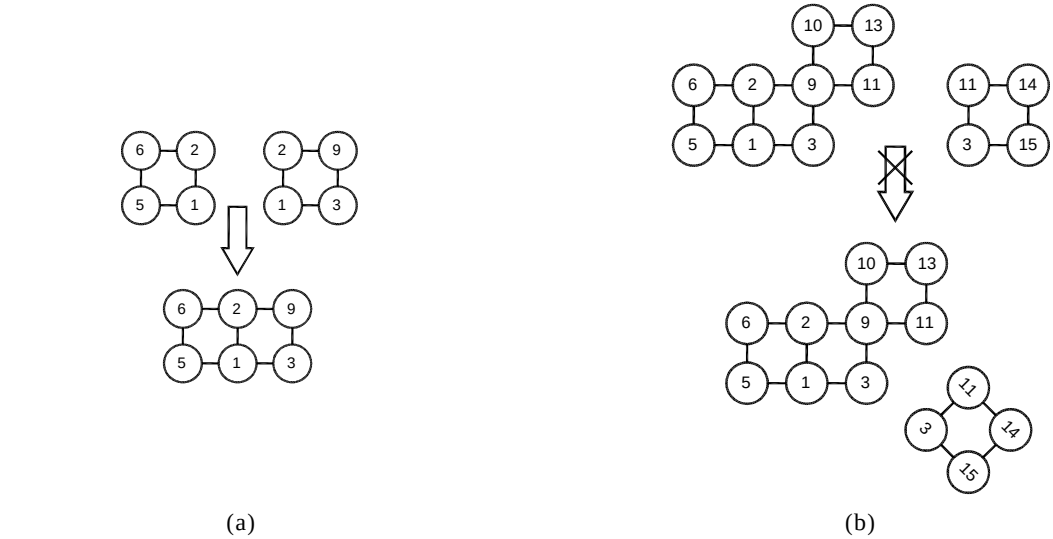
To extract the topology of the connected component regions, a Delaunay triangulation is computed using each connected component centroid. This triangulation is formed from the set of centroids such that the circumcircle of any three vertices, joined in a triangle, is empty [8]. Since the nearest neighbor graph is a subgraph of a Delaunay triangulation, this approach allows neighboring regions in the regular grid topology to be rapidly located.

Delaunay triangulation of the connected components nominally produces six edges connected to each vertex (Fig. 2(a)). Four of these edges correspond to principal grid directions, and the remaining (hypotenuse) edges must be pruned. Triangles that share a common hypotenuse are paired to form a quadrilateral (Fig. 2(b)). Principal grid directions of the quadrilateral representation are aligned between frames. In the first frame of a video sequence, two orthogonal edges of a completed quadrilateral are selected to define the principal grid directions of the grid structure. In all subsequent image frames, the edges of a completed quadrilateral that best match the previous principal grid directions are selected as the defining vectors in the current frame. Based upon the principal grid directions, all quadrilaterals in the structure are oriented to establish a common vertex ordering. As illustrated in Fig. 2(c), 1- black, 2- blue, 3- green, and 4- red vertices are consistently oriented about the quadrilateral centroid. At this stage, adjacent quadrilateral structures have coincident sides.

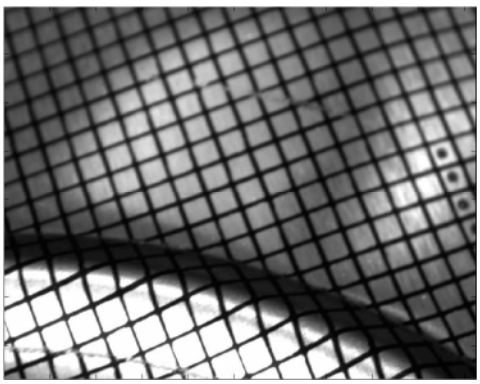
Using each vertex as a node, the adjacent quadrilateral structures are then merged to generate the overall topological structure of the grid. A vertex near the image center is chosen as the initial node,

and a search identifies all quadrilaterals containing this vertex. All new neighboring vertex relationship information from just previously identified quadrilaterals is merged into the topology. A First- In, First- Out (FIFO) vertex queue stores neighboring vertices and the merging process is repeated until all vertices have been traversed (Fig. 3(a)).

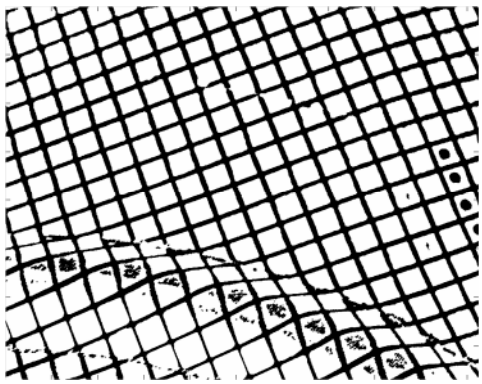
Real world noise, physical surface grid damage, and extreme surface geometry may lead to imperfections in the thresholded image, and relative inconsistencies between the topological vertices forming quadrilaterals. Spurious topologies include instances where a single topological node is entirely missing, erroneously detected by two or more differing neighbor sets, or has only a single (dangling) neighbor (found at the outside edge of the overall topology). Verification traversals through the topology identify and reject inconsistent cases so that only a regularly ordered, four- neighbor structure remains (Fig. 3(b)). An actual image triangular (rather than square) topology rejection is illustrated in Fig. 3(c- f) [10]. A final filtering phase then considers the geometry of the quadrilateral structures formed by topological vertices, and removes any structures that are keystone or non-rhomboid in shape.



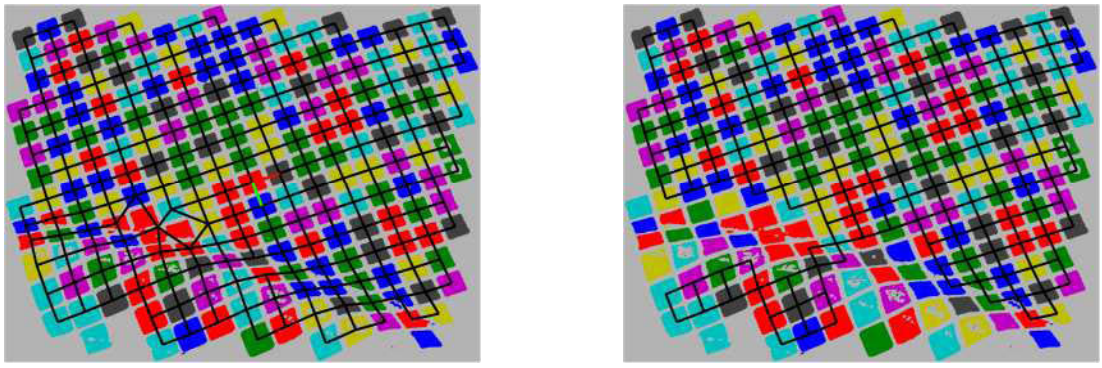
(a) (b)



(c)



(d)



(e) (f)

Fig. 3: Merging quadrilateral vertex neighbor information: (a) topological merging; (b) topological inconsistency detection and rejection; (c) flanged surface image; (d) thresholded flanged surface; (e) inconsistent topology; (f) inconsistencies rejected.

Fig. 4: Sample metric computation for connected component x,y with fiducial present. The black/white regions have been inverted for ease of illustration. Metric $\sigma_{x,y}^{k} = 2572/2922 \approx 0.88$. (a) Connected component x,y, $n_{\text{active}}^{x,y} = 2572$; (b) Filled region, $n_{\text{total}}^{x,y} = 2922$.

4 TOPOLOGICAL TRACKING

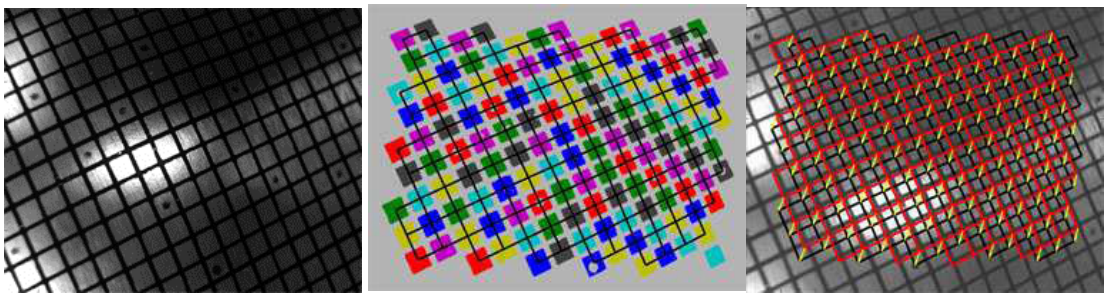
When a regular grid fills an entire camera field of view, the symmetry of the pattern can present ambiguous grid feature correspondences between frames. Asymmetry is therefore added by introducing black fiducials between grid lines. Fiducial information is used during the topology construction process to formulate a metric based upon active pixel ratios within each connected component region (Fig. 4). For connected component x,y in frame k, a fiducial metric value is computed from the ratio:

$$\sigma_{x,y}^{k} = \frac{n_{\text{active}}^{x,y}}{n_{\text{total}}^{x,y}} \tag{1}$$

where n_{active} is the number of pixels that are active (white) within a given connected component, and n_{total} is the total number of pixels in a filled instance of that region. ($n_{\text{active}} = n_{\text{total}} - n_{\text{fiducial}}$).

For components without fiducials, this ratio will be unity. Components with a fiducial will have a ratio less than one. This fiducial metric is then incorporated with the associated component node in

the topological representation. During tracking, this metric is later evaluated with an optimization objective function to determine the most appropriate shift of the grid.



(a) (b) (c)

Fig. 5: (a) Sample image of a deformed metal surface before processing; (b) Sample topological data from the deformed metal dome; (c) Sample interframe motion vectors between image frames.

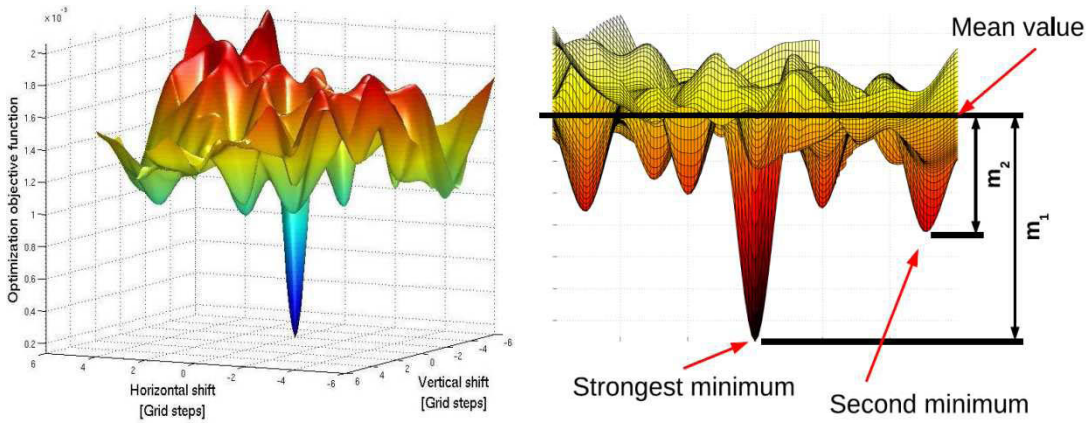
To determine the interframe motion between two views of a gridded surface, the topological grid representations of each image are integrally shifted across one another. An interframe objective function generates a scalar output for each tested motion vector, with the goal that across the search space, an extreme value (minimum) is associated with true registration. As noted earlier, during topology construction, grid directions were normalized to those extracted from the previous frame. This pre-processing alignment step eliminates the possibility of rotations between topological representations, so they need not be considered in the optimization. The optimization objective function is defined as:

$$F(i,j) = \sum_{\substack{(x,y)|(x,y)\in G^k \\ (x+i,y+j)\in G^{k-1}}} (\sigma_{x,y}^{k} - \sigma_{x+i,y+j}^{k-1})^2 + P(i,j) \tag{2}$$

where i,j define the interframe shift to be tested and variables x,y denote the location of the node in the kth topological representation of the grid G^k. The penalty term $P(i,j) = \sqrt{i^2 + j^2}$ represents the Euclidean distance of the integral grid shift.

For each possible motion vector, Eq. (2) logically overlays data defining the grid topologies between adjacent frames. At every overlaid node where both frames have a connected component metric $\sigma_{x,y}^{k}$ defined, the sum of squared differences is computed. This sum is generated across all the topological nodes, and then normalized. The connected component metrics take into account fiducial area and other geometric features, so the objective function provides a measure of the aggregate grid similarity for the current registration.

When testing large interframe motion vectors near the boundaries of the search space, only a small subset of the topological nodes may overlap between frames. This small subset is often not sufficient to provide reliable interframe comparisons, and can lead to erroneous results. The penalty term $P(i,j)$ reduces the strength of objective minima near the search space boundaries, and causes optimization to favor small motions. Consistently finding the correct motion vector near the edges of a search space indicates that camera motion is too rapid for the imaging conditions. To overcome this, either a larger field of view camera or a faster frame rate must be used to provide sufficient overlapping grid elements.



(a) (b)

Fig. 6: Optimization objective function for typical interframe motion: (a) A large downward spike indicating the minimum objective value represents the best estimate for interframe grid registration, (b) Function mean value and strongest minima are marked with black lines. Measurements m_1 and m_2 used to define r_m are labeled.

Using the output from the grid registration, each node in the topological representation can be cast back into image coordinates and used to establish the interframe correspondences between connected component centroids. These correspondences provide the required input, in image coordinates, for computation of the interframe transform. Although the topological grid shift is an integral number of grid units, once cast back into image coordinates, the transform is computed to pixel accuracy. The interframe transform can then be computed using conventional homography estimation techniques [5].

4.1 Tracking Robustness

Tests were performed on various experimental calibration data sequences to validate the algorithm and verify robustness in changing range, focus, and lighting conditions. Tab. 1 shows results from four independent video sequences, with statistics provided on the ratio $r_m = |m_2|/|m_1|$ between the absolute magnitudes of the largest and second largest minima in the optimization objective function. Magnitudes were measured relative to the objective function mean value, and the ratio r_m provides a measure of the contrast between the correct and second best (incorrect) grid registrations. All calculated interframe transforms were manually inspected to confirm correct algorithm output. Fig. 5(c) shows an example result with both the current and previous frame grid topologies overlaid, with the detected interframe motion indicated by arrows.

These results demonstrate robust performance by both the topological construction process and the optimization objective function. The rightmost two columns identify the number of frames for which the second strongest objective minimum was more than 90% and 80%, respectively, of the true minimum. These ratios indicate a lower objective contrast between the correct and an erroneous grid motion. Of the 4350 interframe transforms reported, less than 0.2% of the transforms had a similarity ratio above 80%, and only one frame generated a ratio about 90% (ratio of 0.91). The mean values of r_m report average similarities of less than 50%, with standard deviation smaller than 10%.

5 EXPERIMENTAL RESULTS

5.1 Computational Speed

The computational performance of the topological tracking algorithm was evaluated to determine its applicability to real-time vision systems. A line-based camera calibration grid was tracked across a video sequence, producing a series of transformation matrices that describe the relative interframe motion between each successive frame. The video sequences were captured at 15 frames per second and consisted of grayscale images with a resolution of 1024 by 768 pixels. The implementation was written in C++ and compiled using Microsoft Visual Studio 2010 with release (/O2) optimization. The experiments were performed on a 2.8 GHz Intel Core i7. A breakdown of the average execution time required to process a single frame can be found in Tab. 2.

The algorithm is dominated by the image pre-processing, especially the connected component labeling of the grid elements and fiducials. As can be seen, however, the reduced dimensionality of the topological structure has been leveraged to simplify the remaining computational effort. On average, the implementation approaches 20 frames per second throughput and never operates below the video frame rate of our camera system. These timing results suggest that this algorithm is suitable for a range of applications that require real-time tracking of gridded surfaces.

6 CONCLUSION

This paper has described a novel and practical algorithm for accurate interframe tracking of line-based surface grids larger than the camera field of view. The topological approach exploits the grid structure, and is used to reduce the dimensionality of the projective tracking problem. Fiducial markers used to provide asymmetry in the grid have few design requirements, affording flexibility and avoiding the need for specific feature detectors. As an important stage in camera calibration, this work extends the literature by exploiting structure unique to such applications.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Post Graduate Scholarship and Discovery Grant financial support from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Luhman, T.; Robson S.; Kyle, S.; Harley, I.: Close Range Photogrammetry: Principles, Techniques and Applications, Wiley, 2007.

[2] Spence, A. D.; Capson, D. W.; Sklad, M. P; Chan, H.-L.; Mitchell, J. P.: Simultaneous Large Scale Sheet Metal Geometry and Strain Measurement, Trans. ASME, J. Manufac. Sci. and Eng. 130(5), 2008. DOI:10.1115/1.2976121

[3] Irani, M.; Anandan, P.: About Direct Methods, Vision Algorithms: Theory and Practice (Lecture Notes in Computer Science), 1883, 2000, 267-277. DOI:10.1007/3-540-44480-7_18

[4] Torr, P.; Zisserman, A.: Feature Based Methods for Structure and Motion Estimation, Vision Algorithms: Theory and Practice (Lecture Notes in Computer Science), 1883, 2000, 278-294. DOI:10.1007/3-540-44480-7_19

[5] Hartley, R.; Zisserman, A.: Multiple View Geometry in Computer Vision (2 ed.), Cambridge University Press, New York, NY, USA, 2003.

Sequence Number	*Number of Frames*	*Mean of r_m*	*Standard Deviation*	*Maximum value of r_m*	*# frames $r_m > 0.9$*	*# frames $r_m > 0.8$*
1	1110	0.4740	0.0823	0.9100	1	2
2	550	0.4837	0.0790	0.8552	0	3
3	1800	0.4772	0.0794	0.8886	0	2
4	900	0.4865	0.0756	0.7997	0	0

Tab. 1: Experimental results. Statistics on the ratio r_m are shown from four experimental video sequences, and the number of frame pairs with an r_m value larger than a threshold are listed. Correct interframe motion tracking was obtained for all frames.

Processing Stage	*Execution Time (ms)*	*Percent of Total Time*
Adaptive Thresholding	4.1	7.84%
Morphological Operations	7.6	14.53%
Connected Component Labeling & Metric computation	37.8	72.28%
Delaunay Triangulation & Quadrilateral Formation	0.5	0.96%
Grid Axis Alignment	0.3	0.57%
Topological formation and filtering	0.3	0.57%
Interframe Registration	0.3	0.57%
Interframe Transform Computation	0.4	0.76%
Total	51.3	100%

Tab. 2: Average Execution time for each stage of the topological tracking algorithm (single frame).

[6] Tsai, R. Y.: An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, 1986, 364-374.

[7] Zhang, Z.: A Flexible New Technique for Camera Calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 2000, 1330-1334.

[8] Shu, C.; Brunton, A.; Fiala, M.: A Topological Approach to Finding Grids in Calibration Patterns, Machine Vision and Applications, 21(6), 2010, 949-957. DOI:10.1007/s00138-009-0202-2

[9] Shafait, F.; Keysers, D.; Breuel, T.: Efficient Implementation of Local Adaptive Thresholding using Integral Images, Proceedings of the 15th Document Recognition and Retrieval Conference (DDR-2008), Part of the IS&T SPIE International Symposium on Electronic Imaging, San Jose, CA, 2008.

[10] Kinsner, M. H.: Close-range Machine Vision for Gridded Surface Measurement, Ph.D. Thesis, McMaster University, Hamilton, ON, 2011, http://digitalcommons.mcmaster.ca/opendissertations/6072/.