OTKA Final Report

Beyond Point-based Geometric Alignment, Fusion, 3D Reconstruction and Recognition of Visual Objects

Principal Investigator:	Zoltan Kato (SZTE) and Laszlo Czuni (PE)
Project ID:	K 120366 (main) and 120367 (assoc.)
Institutions:	Consortium of University of Szeged and University of Pannonia
Duration:	60 (2016-12-01 - 2021-11-30)
Written by:	Zoltan Kato, Laszlo Czuni, Levente Tamas

1 Introduction

Today, different sensors and approaches are often combined to build a detailed, geometrically correct and properly textured 3D or 4D (spatio-temporal) model of an object, a scene, or to achieve reliable object detection and recognition. Visual and non-visual sensor data are fused to cope with varying illumination, surface properties, motion, and occlusion. This research project aimed to produce generic ways of working with image patches (without point correspondences) and to directly provide the corresponding 3D information (surface patches, pose, or recognition). From a practical point of view, these methods open new possibilities for the application of reconstruction, recognition, and fusion of 3D depth data with 2D imagery. Important examples occur in security (surveillance), industry (visual inspection), and intelligent transportation (autonomous driving). The algorithms proposed in this project are computationally efficient and require little user interaction, paving the way for embedded vision systems, where limited computing power does not allow the use of many current techniques. Thus, our algorithms can contribute to exciting technologies such as mobile computing, autonomous vehicles, or drones. The contents of this report are grouped into 5 parts, which directly correspond to the main chapters of the original project plan:

- Patch-based camera independent reconstruction using 3D priors (Section 2.1): in addition to patch-based reconstruction we included Markov-based curvilinear reconstruction methods (Section 2.1.1)
- Beyond point correspondences (Section 2.2)
- Registration across different dimensions (Section 2.3)
- 3D-3D nonrigid registration for heterogeneous data (Section 2.4)
- Video object recognition (Section 2.5)

In Section 2.6 we included such results which were not categorized in the above classes but are closely related to this research.

2 Results

2.1 Patch-based camera independent reconstruction using 3D priors

The mainstream approach to passive stereo reconstruction is based on projective geometry that provides a full reconstruction framework. These methods typically rely on point correspondences and the epipolar geometry of points. The reconstruction theory is well developed and tested for central perspective cameras [1]. Region-based methods proved to be highly accurate and robust. We proposed a novel solution for multi-view reconstruction, relative pose and homography estimation using planar regions. The proposed method doesn‘t require point matches, it directly uses a pair of planar image regions and simultaneously reconstructs the normal and distance of the corresponding 3D planar surface patch, the relative pose of the cameras as well as the aligning homography between the image regions. When more than two cameras are available, then a special region-based bundle adjustment is proposed, which provides robust estimates in a multi-view camera system by constructing and solving a non-linear system of equations. The method is quantitatively evaluated on a large synthetic dataset as well as on the KITTI vision benchmark dataset [2].

Nowadays, depth (RGBD) cameras are gaining more and more attention. Having partial depth data of a scene, one can use it as a prior for efficient image-based (stereo or multiview) 3D reconstruction to produce

an accurate 3D scene representation. We proposed a novel graph-cut based 3D reconstruction method which is able to take into account partially available depth data as a prior. We formulate the energy as a function written as a sum of terms that can be minimized in two different representations: (1) assignment-based, which yields a standard binary energy, in this approach occlusion and uniqueness is handled naturally; as well as (2) a multi-label representation which yields to a non-binary energy where occlusion is only handled by assigning a special label to occluded pixels, while uniqueness is not handled explicitly. Both representations have their advantages and disadvantages, which are analyzed and discussed into details through various experimental results on the publicly available Middlebury stereo datasets and on real stereo images. Results show, that the use of depth prior information from different sources produces better 3D reconstructions [3].

2.1.1 2D and 3D models for the reconstruction of curvilinear objects

Besides the recognition methods of 3D objects from multiple directions, we turned to other problems such as segmentation and structural reconstruction of objects in cluttered environments [4]. We focus on objects that have line-like (piecewise-linear) or curvilinear structures such as trees (with branches), veins in organs or road structures in aerial images [5], [6], [7]. We tried different approaches for the representation of such patterns and we first found parts-based methods to be the most efficient. We created a Marked Point Process (MPP) with Reversible Jump Monte Carlo Markov Chain Optimization for the optimal reconstruction of such structures from parts and used it for the delineation and segmentation of tree images. For the initial detection of part of the object, we used different neural networks such as SegNet and DeepLab, and the MPP was building its model over the initial estimates [8], [9]. Finally, we found that the simultaneous detection of segments, segment borders, and parts' center lines with CNNs is very accurate and computationally efficient, and even the layering of occluding parts can be determined in many cases [10]. As the final approach we are planning to include stereo images to generate not only 2D or 2.5D but realistic 3D models of trees.

2.2 Beyond point correspondences

Recently, there has been much interest in object detection and matching using local invariant features (*e.g.* SIFT [11], SURF [12]) or regions (*e.g.* MSER [13]). Since these features are invariant to translation, rotation, and scaling, points or regions can be matched between images accurately. Moreover, if the matches have already been obtained, the transformations between the patches around the feature centers or regions can be obtained as well. Omnidirectional cameras are particularly interesting here, because state of the art point descriptors implicitly assume a perspective geometric relation between images. We show that novel camera calibration methods can be constructed to estimate camera parameters for non-conventional optics too: a novel method is proposed [14] for the absolute pose estimation of a central 2D camera with respect to 3D depth data without the use of any dedicated calibration pattern or explicit point correspondences. The proposed method has no specific assumption about the data source: plain depth information is expected from the 3D sensing device, and a central camera is used to capture the 2D images. Both the perspective and omnidirectional central cameras are handled within a single generic camera model. Pose estimation is formulated as a 2D-3D nonlinear shape registration task which is solved without point correspondences or complex similarity metrics. It relies on a set of corresponding planar regions, and the pose parameters are obtained by solving an overdetermined system of nonlinear equations. The efficiency and robustness of the proposed method were confirmed on both large-scale synthetic data and on real data acquired from various types of sensors.

While many handcrafted features have been proposed for keypoints, only a few methods exist for line segments. It is well known, however, that line segments are commonly found in man-made environments, in particular urban scenes, thus they are important for applications like pose estimation, visual odometry, or 3D reconstruction.

First, we addressed the problem of estimating the absolute pose of a multiview calibrated perspective camera system from 3D - 2D line correspondences [15]. We assume, that the vertical direction is known, which is often the case when the camera system is coupled with an IMU sensor, but it can also be obtained from vanishing points constructed in the images. We proposed two solutions, both can be used as a minimal solver as well as a least squares solver without reformulation. The first solution consists of a single linear system of equations, while the second solution yields a polynomial equation of degree three in one variable and one systems of linear equations which can be efficiently solved in closed-form. The proposed algorithms have been evaluated on various synthetic datasets as well as on real data. Experimental results confirm state of the art performance both in terms of quality and computing time. Then the method has been extended to generalized camera systems [16]: The only assumption about the imaging model is that 3D straight lines are projected via projection planes determined by the line and camera projection directions, i.e. correspondences are given as a 3D world line and its projection plane. Since modern cameras are frequently equipped with various location

and orientation sensors, we assume that the vertical direction (e.g. a gravity vector) is available. Therefore we formulate the problem in terms of 4 unknowns using 3D line - projection plane correspondences which yields a closed form solution. The solution can be used as a minimal solver as well as a least squares solver without reformulation. The proposed algorithm have been evaluated on various synthetic datasets as well as on real data. Experimental results confirm state of the art performance both in terms of quality and computing time.

Subsequently, the problem of multicamera absolute and relative pose estimation has been considered in [17]: The algorithm relies on two solvers: a direct solver using a minimal set of 6 line pairs and a least squares solver which uses all inlier 2D-3D line pairs. The algorithm have been validated on a large synthetic dataset, experimental results confirm the stable and real-time performance under realistic noise on the line parameters as well as on the vertical direction. Furthermore, the algorithm performs well on real data with less then half degree rotation error and less than 25 cm translation error on a 10m range outdoor scene.

Finally, pose estimation without the knowledge of the vertical direction becomes a more complex problem yielding various formulations and solutions: For estimating the absolute and relative pose of a camera system composed of general central projection cameras [18] such as perspective and omni-directional cameras, we derive a minimal solver for the minimal case of 3 line pairs per camera, which is used within a RANSAC algorithm for outlier filtering. Then, we also formulate a direct least squares solver which finds an optimal solution in case of noisy (but inlier) 2D-3D line pairs. Both solver relies on Grobner basis, hence they provide an accurate solution within a few milliseconds in Matlab. The algorithm has been validated on a large synthetic dataset as well as real data. Experimental results confirm the stable and real-time performance under realistic outlier ratio and noise on the line parameters. Comparative tests show that our method compares favorably to the latest state of the art algorithms. Then a new algorithm for estimating the absolute and relative pose of a multi-view camera system is proposed in [19]. We derive a direct least squares solver using Grobner basis which works both for the minimal case (set of 3 line pairs for each camera) and the general case using all inlier 2D-3D line pairs for a multi-view camera system. The algorithm has been validated on a large synthetic dataset as well as real data. Experimental results confirm the stable and real-time performance under realistic outlier ratio and noise on the line parameters. Comparative tests show that our method compares favorably to the latest state of the art algorithms.

2.3 Registration across different dimensions

This problem arises in calibration of nonconventional optics/sensors (*e.g.* omnidirectional optics, Lidar, etc.). While standard projective camera calibration is extensively studied and has many working solutions, the calibration of camera systems consisting of different sensors (*e.g.* Lidar, traditional color camera, or infra camera) is less studied. The problem of extrinsic calibration for 3D Lidar and color camera was first addressed in [20] which generalized the algorithm proposed by Zhang in [21]. For high precision aerial image registration the work presented in [22] is based on the information of the Lidar scan intensity. For low precision and high frame rate systems used *e.g.* for navigation purposes, the registration challenges are addressed in different ways. Usually there are several Lidar-camera scan pairs acquired and the registration is performed on these image pairs [23]. Other works are related to Lidar-omnidirectional camera registration [24]. In such mixed environments, correspondence-free and target-less calibration is particularly important since, due to unusual optical distortions and different sensory information, correspondences are difficult to establish. Furthermore, target-less calibration is important when images taken at different time (*e.g.* a Lidar scan and an infra image) need to be fused. A strongly related area is image-based navigation, which is becoming more and more important with the widespread use of smart mobile phones and UAVs.

Our preliminary results on calibrating a pair of perspective and Lidar-cameras was presented in [25]. Based on these results, we formulated pose estimation as a region based registration for central and non-perspective optics, such as omnidirectional cameras and various depth sensors.

As a real-life application, a workflow was proposed [26] for Cultural Heritage applications in which the fusion of 3D and 2D visual data is required. Using data acquired by cheap, standard devices, like a 3D scanner having a low quality 2D camera in it, and a high resolution DSLR camera, one can produce high quality color calibrated 3D model for documenting purpose. The proposed processing workflow combines a novel region based calibration method with an ICP alignment used for refining the results. It works on 3D data, that do not necessarily contain intensity information in them, and 2D images of a calibrated camera. These can be acquired with commercial 3D scanners and color cameras without any special constraint. In contrast with the typical solutions, the proposed method is not using any calibration patterns or markers. The efficiency and robustness of the proposed calibration method has been confirmed on both synthetic and real data.

Another workflow is proposed for cultural heritage applications [27] where the fusion of 3D and 2D visual data is required. Using a metric 3D point cloud acquired by a Lidar scanner and 2D images of a commercial high-resolution DSLR camera, we show how to produce a high-quality, metric 3D model for documenting or architectural planning purpose. The proposed processing workflow describes the data acquisition tasks, the

steps of the data processing, and the proposed method used for colorizing the point cloud from multiple cameras by choosing the camera with the best view based on different conditions. We show results on two Reformed Churches: Kolozsnema and Somorja.

2.3.1 New pose estimation and sensor calibration techniques

During our research we were focusing on different facets of the pose estimation problems including applications in the cultural heritage [28] and industrial robotics as well. We analyzed the relative pose estimation between augmented reality (AR) systems and a fixed robot, based on cross calibration between 2D cameras, using SVD techniques [29], [30]. With this approach one can bring into the same coordinate system an AR visualizing tool and an arbitrary other device with a 2D camera. Special focus was dedicated to the change detection problem in urban scenes using 2D and 3D data as well [31].
We proposed an approach for the calibration of the depth camera mounted on the arm and the gripper frame, which is based on a relative calibration to a fixed frame in space [32]. The main idea is to use a calibration pattern as an external reference with an internally calibrated depth camera [33]. The position of this checkboard and fixed frame with respect to the depth camera can be determined with SVD based methods taking into account the physical size of the pattern[34].
For the 3D data based pose estimation we developed an algorithm with for the efficient normal estimation (base operation for 3D point cloud data processing) from ToF camera specific depth image processing based on Feature Pyramid Networks. The results were tested on both embedded and server grade devices and the related results were published in [35], [36], and [37].

2.4 3D-3D non-rigid registration for heterogeneous data

The need for heterogeneous 3D data registration is becoming a must in several applications including autonomous navigation, mapping or even cultural heritage use cases. The main challenge in this type of data fusion is the relaxation of the rigid-transformation constraints. Due to the different physical measurement principles of the different sensors the resulting data undergoes a non-linear distortion, for which some un-distortion can help, but the overall consistency of the resulting 3D models may suffer from problems like occlusion, reflectance, shadow effects of surface unevenness. In order to deal with such situations a local, patch level 3D fusion approach is proposed which integrated in a global alignment procedure ensures the consistent 3D data fusion for the slightly non-rigid deformation cases too. Typical use cases for this approach would include the merging of the SfM 3D models with the geometrically correct (LiDAR, CAD model, etc) data. This can compensate the shortcomings of the input data (e.g holes, surfaces with scattered points, uneven sampling, lack of color information). Our approach [38] deals with the problem of fusing different (potentially partial) 3D meshes to fill in missing parts (holes) of an accurate reference 3D model using a less accurate but more complete moving 3D model. Typically, accurate 3D models can be produced by range devices (Lidar) which is often limited in setting viewpoints, while traditional Structure from Motion methods are using 2D images which are less restricted in viewpoints, but overall produce a less accurate 3D mesh. Combining the advantages of both modalities is an appealing solution to many real world problems. Herein we proposed a novel method which detects holes in the accurate reference mesh and then each hole is filled from the less accurate 3D mesh by gradually estimating local affine transformations around the hole's boundary and propagating it into the inner part. Experimental validation was done on a large real dataset, which confirms the accuracy and reliability of the proposed algorithm.

In [39], a region-based approach is proposed to find a thin plate spline map between a pair of deformable 3D objects represented by triangular surface meshes. The proposed method works without landmark extraction and feature correspondences. The aligning transformation is simply found by solving a system of integral equations. Each equation is generated by integrating a non-linear function over the object domains. We derive recursive formulas for the efficient computation of these integrals for open and closed surface meshes. Based on a series of comparative tests on a large synthetic dataset, our triangular mesh-based algorithm outperforms state of the art methods both in terms of computing time and accuracy. The applicability of the proposed approach was demonstrated on the registration of 3D lung CT volumes, brain surfaces and 3D human faces.

2.5 Video object recognition

It is obvious that video gives much more information about 3D objects than simply 2D projections. Not only the different views of the objects can be recorded but the 3D structure can be reconstructed by structure from motion techniques. However, these later approaches require good quality images and camera calibration with relatively large computational power still far from most of the mobile computing platforms and intelligent sensor

motes. Luckily mobile computing devices often contain inertial measurements units (IMUs) and the calibration of cameras can be combined with IMUs [40]. However, it is still an open question how to exploit the IMUs in video recognition without going through the structure from motion processing methodology. Our research is focused on a viewer centered recognition model where the relative position of the target object and the camera is utilized. Our preliminary experiments already showed [41] that IMUs can help in the recognition process with low computational demands. However, fast object tracking and/or segmentation still can be a problem in this framework being also a subject for research. Most object recognition are *passive* from the model side. We propose to build up model-driven interactive retrieval methods where the search engine gives hint how to move the camera around the object to get the fastest and most reliable recognition result.

2.5.1 Multiview recognition with information fusion of optical and directional information of IMUs and probabilistic models for object recognition with 3D sensors

One main aspect of the proposed methods was the low computational demand regarding complexity and memory usage. As planned we made new models for the retrieval and recognition of 3D objects with lightweight devices with information fusion of optical and inertial information. Also we created techniques with 3D sensors using Markov processes. The main ideas which were successfully exploited:

- Hough paradigm based approach: An object retrieval method was developed where compact visual descriptors of different views of the object and the camera's orientation was fused. We have shown that the utilization of the very informative and lightweight orientation results in the increase of the hit-rate [42], [43].
- Hidden Markov Model (HMM) based retrieval: In this approach, we modeled the views of objects from different directions as hidden states and the compact visual descriptors were considered as noisy observations [44], [45]. This approach is also applicable to pose estimation [46].
- Fusing information from CNNs (convolutional neural network) confidence values [47].
- Applying active vision to improve performance of the above Hough transformation and HMM models [48], [49].
- Active perception approach for object detection using 3D cameras based on Partially Observable Markov Decision Processes [50], [51] while in the work [52] simple and robust deep network architectures were trained for object recognition in indoor environment.

2.6 Other results related to object classification, recognition, and pose estimation

In the first months of the project we finished the research stretching over from the 2016 to develop lightweight methods for the binary classification of time domain signals [53].
We also continued our research related to the analysis of remote sensing images. The analysis and classification of image regions are targeted to estimate the urbanization used in many areas of biological research. There are two main improvements on this field: we investigated the scalability of our method and we made experiments with eye-tracking tools to discover the relations between the human's recognition and the automatic mechanisms [54], [55]. Our implementation of the urbanization score estimation method reached over 300 downloads from 19 countries (used in a total of 30 countries).
An application motivated research topic was the usage of eye-tracker for the measurement and evaluation of users' behavior during operators' training for manufacturing. With the help of the developed technique we can design more efficient visual tutorials to enhance the process of training in industrial environments [56]. Another computer vision technique being developed is to apply a surveillance method for the analysis and assessment of human posture and motion activities at workplaces. The developed technique can be applied to estimate human body parameters related to workplace ergonomic recommendations [57], [58].
Furthermore, advanced data analysis [59] and network latency mitigation were proposed in industrial robotics perception scenarios in the works [60],[61]. Both research direction proved to be relevant for the next generation Industry X.0 specific sensing and perception research domain.
For the outdoor robot 2D perception and navigation problem a grid based projection variant was proposed in the work [62]. An active perception approach was studied based on Partially Observable Markov Decision Processes (POMDP) used in an optimistic control setup [51]. A Recurrent Neural Network (RNN) based control law for 2D image based navigation was studied in the experimental reporting [63].
In the last months of the project we investigated siamese neural networks for defect detection, first implementations were tested in traffic signs [64] but in future we are planning to apply the methods to other industrial

products as well.

In the future, we plan to extend the methods developed within this project both on fundamental levels focusing on closed form solutions with optimistic optimization procedures for faster convergence suitable for embedded GPU platforms. On the technological side we intent to validate our prior results in applied knowledge transfer projects towards the industrial environment. These efforts hopefully will converge in a larger scale H2020, EIDN, or COST project with our existing international network from the three institutions involved in the current project.

3 Summary of Activities

The research project ran for 60 months after an extension due to Covid-19. Beside senior researchers BSc/MSc/PhD students were also involved in the research. We organized 4 mini workshops for the participants of the project to present and to discuss the research results and possible research directions. It was held twice in Veszprém, once in Szeged and once in Transylvania.
We sincerely appreciate the support of OTKA to reach the above detailed results and the constructive and anonymous reviews of the project proposal and midterm reports.
Quantitative summary of the scientific outputs of the project:

- 11 International journal papers[1]: 7 Q1 and 4 Q2 articles, cumulative IF: 59.426
  [14]: Scimago (Computer Vision and Pattern Recognition): Q1, IF: 16.389,
  [39]: Scimago (Computer Vision and Pattern Recognition): Q1, IF: 7.917,
  [3]: Scimago (Operations Research): Q2, IF: 2.345,
  [42]: Scimago (Computer Vision and Pattern Recognition): Q2, IF: 1.836,
  [43]: Scimago (Electrical and Electronic Engineering): Q2, IF: 3.021,
  [48]: Scimago (Signal Processing 2020): Q2, IF: 2.157,
  [9]: Scimago (Computer Science): Q1, IF: 3.367,
  [53]: Scimago (Electrical and Electronic Engineering): Q1, IF: 2.617,
  [30]: Scimago (Computer Science Applications): Q1, IF:5.666,
  [59]: Scimago (Computer Science): Q1, IF:4.098,
  [61]: Scimago (Industrial and Manufacturing Engineering): Q1, IF: 2.861
- 23 International conference papers [15, 16, 17, 19, 18, 27, 2, 26, 38, 65, 44, 46, 47, 50, 52, 4, 5, 10, 29, 36, 57, 60, 64]
- 11 Other presentations (with abstracts only): [45, 49, 6, 7, 8, 32, 33, 34, 54, 55, 56]
- 6 PhD Thesis: Abdellali Hichem (2022), Robert Frohlich (2019), Zsolt Santa (2018, KEPAF Best PhD Prize 2019), Metwally Rashad (2018), Amr Nagy (expected in 2022), Daniel Mezei (2021)
- 1 Technical report: [31]

Awards:

- KEPAF Best PhD Prize: Zsolt Santa (2019)
- KEPAF Best Paper Award: Hichem Abdellali (2021, based on [18])
- TDK 1. Prize: Nora Horanyi (1. Prize & Morgan-Stanley Prize, 2017)
- ETDK 1. Prize: Szilard Molnar (1. Prize, 2020)

Publicly available demo implementations of the methods developed within the project:

1. *Absolute Pose Estimation of Central Cameras Using Planar Regions* [14]: http://www.inf.u-szeged.hu/~kato/software/AbsolutePoseCentralCam.html
2. *Absolute and Relative Pose Estimation of a Multi-View Camera System using 2D-3D Line Pairs and Vertical Direction* [17]: http://www.inf.u-szeged.hu/rgvc/demos.php?did=pose_dicta
3. *Multiview Absolute Pose Using 3D – 2D Perspective Line Correspondences and Vertical Direction* [15]: http://www.inf.u-szeged.hu/rgvc/demos.php?did=pose

[1]Please note: in some cases only previous year data is available and given.

4. *Generalized Pose Estimation from Line Correspondences with Known Vertical Direction* [16]: http://www.inf.u-szeged.hu/rgvc/demos.php?did=pose_gPnLup
5. *2D Change detection based on 3D priot information* [31]: http:ftp://users.utcluj.ro/public_html/download/private/otka-tl.zip

International collaborations:

1. Naver Labs Europe, Meylan, France
2. Image science & computer vision group, University Jean Monnet, France
3. Janos Selye University, Slovakia
4. Technical University of Cluj-Napoca, Romania

Publications [57] and [58] do not contain project identifiers.

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge, UK: Cambridge University Press, 2003.

[2] R. Frohlich and Z. Kato, "Simultaneous multi-view relative pose estimation and 3D reconstruction from planar regions," in *Proceedings of ACCV Workshop on Advanced Machine Vision for Real-life and Industrially Relevant Applications*, ser. Lecture Notes in Computer Science, G. Carneiro and S. You, Eds., vol. 11367. Perth, Australia: Springer, Dec. 2018, pp. 467–483.

[3] H. Abdellali and Z. Kato, "3d reconstruction with depth prior using graph-cut," *Central Eur. J. Oper. Res.*, vol. 29, no. 2, pp. 387–402, 2021. [Online]. Available: https://doi.org/10.1007/s10100-020-00694-6

[4] L. Czúni, A. Kürtösi, and K. B. Alaya, "Color based clustering for trunk segmentation," in *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP).* IEEE, 2018, pp. 1–4.

[5] L. Czúni and K. B. Alaya, "Low-and high-level methods for tree segmentation," in *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1. IEEE, 2019, pp. 189–192.

[6] K. B. Alaya and L. Czúni, "Using graphs for high-level 2d trunk modeling," Pannonian Conference on Advances in Information Technology (PCIT), Abstract, 2019.

[7] L. Czúni and K. B. Alaya, "Segmentation of complex structures with parts based RJMCMC," VOCAL Optimization Conference: Advanced Algorithms, Abstract, 2018.

[8] K. B. Alaya and L. Czúni, "Vectorizing curvilinear objects with a variational method," Pannonian Conference on Advances in Information Technology (PCIT), Abstract, 2020.

[9] ——, "Stochastic modeling of trees in forest environments," *IEEE Access*, vol. 9, pp. 69 143–69 156, 2021.

[10] ——, "Cnn-based tree model extraction." in *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2021.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of the British Machine Vision Conference 2002, BMVC 2002, Cardiff, UK, 2-5 September 2002*, 2002.

[14] R. Frohlich, L. Tamas, and Z. Kato, "Absolute pose estimation of central cameras using planar regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 377–391, Feb. 2021.

[15] N. Horanyi and Z. Kato, "Multiview absolute pose using 3D - 2D perspective line correspondences and vertical direction," in *Proceedings of ICCV Workshop on Multiview Relationships in 3D Data.* Venice, Italy: IEEE, Oct. 2017, pp. 1–9.

[16] ——, "Generalized pose estimation from line correspondences with known vertical direction," in *Proceedings of International Conference on 3D Vision.* Qingdao, China: IEEE, Oct. 2017, pp. 1–10.

[17] H. Abdellali and Z. Kato, "Absolute and relative pose estimation of a multi-view camera system using 2d-3d line pairs and vertical direction," in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications.* Canberra, Australia: IEEE, Dec. 2018, pp. 1–8.

[18] H. Abdellali, R. Frohlich, and Z. Kato, "Robust absolute and relative pose estimation of a central camera system from 2d-3d line correspondences," in *Proceedings of ICCV Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving.* Seoul, Korea: IEEE, Oct. 2019.

[19] ——, "A direct least-squares solution to multi-view absolute and relative pose from 2d-3d perspective line pairs," in *Proceedings of ICCV Workshop on 3D Reconstruction in the Wild.* Seoul, Korea: IEEE, Oct. 2019.

[20] R. Unnikrishnan and M. Hebert, "Fast extrinsic calibration of a laser rangefinder to a camera," Carnegie Mellon University, Tech. Rep., 2005.

[21] Q. Zhang, "Extrinsic calibration of a camera and laser range finder," in *International Conference on Intelligent Robots and Systems*, IEEE. Sendai, Japan: IEEE, September 2004, pp. 2301 – 2306.

[22] A. Mastin, J. Kepner, and J. W. F. III, "Automatic registration of lidar and optical images of urban scenes," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Miami, Florida, USA: IEEE, June 2009, pp. 2639–2646.

[23] S. Bileschi, "Fully automatic calibration of lidar and video streams from a vehicle," in *12th International Conference on Computer Vision Workshops*, IEEE. Kyoto, Japan: IEEE, September 2009, pp. 1457–1464.

[24] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes," in *IEEE International Conference on Intelligent Robots and Systems*, IEEE/RSJ. San Diego, USA: IEEE, October 2007, pp. 4164–4169.

[25] L. Tamas and Z. Kato, "Targetless calibration of a lidar - perspective camera pair," in *Proceedings of ICCV Workshop on Big Data in 3D Computer Vision*, IEEE. Sydney, Australia: IEEE, Dec. 2013, pp. 668–675.

[26] R. Frohlich, Z. Kato, A. Tremeau, L. Tamás, S. Shabo, and Y. Waksman, "Region based fusion of 3D and 2D visual data for cultural heritage objects," in *Proceedings of International Conference on Pattern Recognition*, IEEE. Cancun, Mexico: IEEE, Dec. 2016, pp. 2404–2409.

[27] R. Frohlich, S. Gubo, A. Lévai, and Z. Kato, *3D–2D Data Fusion in Cultural Heritage Applications*. Singapore: Springer Singapore, 2018, pp. 111–130. [Online]. Available: https://doi.org/10.1007/978-981-10-7221-5_6

[28] A. Colson and L. Tamas, "Bremen cog: Three recording techniques for one object," in *Digital Techniques for Documenting and Preserving Cultural Heritage*. ARC, Amsterdam University Press, 2018, pp. 121–140.

[29] A. Blaga and L. Tamas, "Augmented reality for digital manufacturing," in *2018 26th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2018, pp. 173–178.

[30] A. Blaga, C. Militaru, A.-D. Mezei, and L. Tamas, "Augmented reality integration into MES for connected workers," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102057, 2021.

[31] L. Tamas, "Change detection for urban scenes," Technical report, TUCN, 2018. [Online]. Available: http://rocon.utcluj.ro/levente

[32] L. Tamas and L. Baboly, "Industry 4.0–MES vertical integration use-case with a cobot," in *International Conference on Robotics and Automation (ICRA), Dynamics Workshop Poster session*, 2017.

[33] L. Tamas and L. Busoniu, "Active perception for object detection on a conveyor belt," in *International Conference on Computer Vision (ICCV-WS) - 6DoF Pose Estimation Workshop Poster session*, 2017.

[34] C. Militaru, A.-D. Mezei, and L. Tamas, "Lessons learned from a cobot integration into MES," in *International Conference on Robotics and Automation(ICRA) - I3C workshop Poster session*, 2017.

[35] S. Molnár, B. Kelényi, and L. Tamas, "Feature pyramid network based efficient normal estimation and filtering for time-of-flight depth cameras," *Sensors 2021, Vol. 21, Page 6257*, vol. 21, no. 18, p. 6257, sep 2021.

[36] S. Molnár, B. Kelényi, and L. Tamás, "ToFNest: Efficient normal estimation for time-of-flight depth cameras," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1791–1798.

[37] L. Tamas and A. Cozma, "Embedded real-time people detection and tracking with time-of-flight camera," in *Real-Time Image Processing and Deep Learning 2021*, vol. 11736. International Society for Optics and Photonics, 2021, p. 117360B.

[38] L. Kormoczi and Z. Kato, "Filling missing parts of a 3D mesh by fusion of incomplete 3D data," in *Proceedings of Advanced Concepts for Intelligent Vision Systems*, ser. Lecture Notes in Computer Science, B.-T. J., P. R., P. W., P. D., and S. P, Eds., vol. 10617. Antwerp, Belgium: Springer, Sep. 2017, pp. 711–722.

[39] Z. Santa and Z. Kato, "Elastic alignment of triangular surface meshes," *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1220–1244, Nov. 2018.

[40] J. D. Hol, T. B. Schön, and F. Gustafsson, "A new algorithm for calibrating a combined camera and imu sensor unit," in *Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on.* IEEE, 2008, pp. 1857–1862.

[41] L. Czuni and M. Rashad, "Lightweight video object recognition based on sensor fusion," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on*, Oct 2015, pp. 1–5.

[42] L. Czúni and M. Rashad, "The use of IMUs for video object retrieval in lightweight devices," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 30–42, 2017.

[43] ——, "Lightweight active object retrieval with weak classifiers," *Sensors*, vol. 18, no. 3, p. 801, 2018.

[44] ——, "The fusion of optical and orientation information in a Markovian framework for 3D object retrieval," in *International Conference on Image Analysis and Processing.* Springer, 2017, pp. 26–36.

[45] A. M. Nagy and L. Czúni, "Temporal models for 3d object recognition," WCAIML, Abstract, 2019.

[46] L. Czúni and A. M. Nagy, "Hidden Markov models for pose estimation." in *VISIGRAPP (5: VISAPP)*, 2020, pp. 598–603.

[47] ——, "Improving object recognition of CNNs with multiple queries and HMMs," in *Twelfth International Conference on Machine Vision (ICMV 2019)*, vol. 11433. International Society for Optics and Photonics, 2020, p. 1143310.

[48] A. M. Nagy, M. Rashad, and L. Czúni, "Active multiview recognition with hidden Markov temporal support," *Signal, Image and Video Processing*, vol. 15, no. 2, pp. 315–322, 2021.

[49] ——, "About the temporal support of active object recognition," Pannonian Conference on Advances in Information Technology (PCIT), Abstract, 2020.

[50] A.-D. Mezei, L. Tamás, and L. Buşoniu, "Sorting objects from a conveyor belt using active perception with a POMDP model," in *2019 18th European Control Conference (ECC).* IEEE, 2019, pp. 2466–2471.

[51] ——, "Sorting objects from a conveyor belt using POMDPs with multiple-object observations and information-gain rewards," *Sensors*, vol. 20, no. 9, p. 2481, 2020.

[52] A.-O. Fulop and L. Tamas, "Lessons learned from lightweight CNN based object recognition for mobile robots," in *2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR).* IEEE, 2018, pp. 1–5.

[53] L. Czúni and P. Z. Varga, "Time domain audio features for chainsaw noise detection using WSNs," *IEEE Sensors Journal*, vol. 17, no. 9, pp. 2917–2924, 2017.

[54] Á. Lipovits, G. Seres, and L. Czúni, "Szoftvertámogatás területek urbanizációs indexének számításához," Urbanizációs Ökológia Konferencia , Abstract, 2018.

[55] A. Lipovits, "Determining the uncertainty of the urbanization index method based on image information and eye movement data," Urban Transitions, Abstract, 2018.

[56] Á. Lipovits, K. Tömördi, Z. Vörösházi, and R. Jinda, "Investigating the visual forms of dynamic electronic work instructions to improve learning efficiency and productivity in assembly processes," in *Pannonian Conference on Advances in Information Technology (PCIT 2019)*, 2019, p. 84.

[57] B. Szakonyi, T. Lőrincz, Á. Lipovits, and I. Vassányi, "An expert system framework for lifestyle counselling," *Proceedings of the eTELEMED 2018*, 2018.

[58] ——, "Rule-base formulation for clips-based work ergonomic assessment," *Hungarian Journal of Industry and Chemistry*, pp. 79–83, 2019.

[59] D. Mitrea and L. Tamas, "Manufacturing execution system specific data analysis-use case with a cobot," *IEEE Access*, vol. 6, pp. 50 245–50 259, 2018.

[60] L. Márton and L. Tamás, "Wireless data rate controller design for networked control applications," in *2018 26th Mediterranean Conference on Control and Automation (MED).* IEEE, 2018, pp. 1–6.

[61] L. Tamas and M. Murar, "Smart CPS: vertical integration overview and user story with a cobot," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 4-5, pp. 504–521, 2019.

[62] C. Marcu and L. Tamas, "Navigation of outdoor mobile robots with extended grid algorithms," in *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*. IEEE, 2020, pp. 1–6.

[63] C. Sándor, S. Pável, E. Wieser, A. Blaga, P. Boda, A.-O. Fülöp, A. Ursache, A. Zöld, A. Kopacz, B. Lázár *et al.*, "The ClujUAV student competition: A corridor navigation challenge with autonomous drones," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 17 511–17 517, 2020.

[64] A. M. Nagy and L. Czúni, "Detecting object defects with fusioning convolutional siamese neural networks." in *VISIGRAPP (5: VISAPP)*, 2021, pp. 157–163.

[65] G. Csurka, Z. Kato, A. Juhasz, and M. Humenberger, "Estimating low-rank region likelihood maps," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*. Seattle, Washington, USA: IEEE, Jun. 2020, pp. 1–10.