Final Report of Project OTKA#101598 ``Comprehensive Remote Sensing Data Analysis''

Call: Hungarian Scientific Research Fund – Support of Basic Research
Duration: 01.01.2012-31.12.2014.
Track: Postdoctoral (PD)
Principal Investigator: Csaba Benedek, PhD,

Summary of the project achievements

The main goal of the project has been to develop widely applicable methods for detailed and multi-level content analysis of visual information provided by various remote sensing technologies. Starting from solutions of concrete tasks in different application areas, we explored the common problems from the different fields, which were formalized, so that we could propose general methodologies for various applications. Ensuring the theoretical basis of the proposed models, we have chosen the widely established Marked Point Process (MPP) method family as a starting point, however, it has shortly became obvious that the traditional MPP models need to be generalized in various ways to adapt them to several real life problems. In particular, three main issues should be often incorporated in the MPP framework: 1) object grouping 2) sub-object encapsulation 3) temporal sequence analysis. Apart from MPP models, further stochastic and machine learning approaches were investigated and proposed during the project work.

Following the workplan of the project, in the first year, we proposed solutions for various definite remote sensing (RS) applications. Meanwhile we collected and organized corresponding problems appearing in different RS tasks, so that we can handle in similar ways.

In the second year we adopted the previously developed model elements to other application areas, studying model interoperability within a given problem family and exploration of model interoperability for tasks related to significantly different data sources.

In the third year we implemented and documented a generalized MPP framework and procedure library for the addressed remote sensing topics. Experimental validation of the proposed concepts and studies of algorithmic prospects and limitations have also been conducted. In addition, we continued developing algorithms for vision of self driving cars and scene reconstruction using a terrestrial Lidar sensor. Finally, we participated in preparing a comprehensive state-of-the-art overview on different multilayer change detection techniques for multi-temporal remote sensing images.

As outcome of the project **6 journal papers**[1] (summa impact factor = 18.589), **1 book chapter** (in publication), **13 international conference** (including 5 IEEE Xplore, 5 LNCS, 2 ISPRS Annals), and **7 national conference papers** have been published.

In this report, we briefly introduce the main results of the project. The description is organized into 6 sections corresponding to different concerning research topics. Some sections are decomposed into multiple parts, where each subsection emphasizes a given contributions.

1. Multiframe Marked Point Process model for target sequence analysis in ISAR images

[1] The latest ISPRS J. P Rem. Sens. paper is under minor revision.

Conventional Marked Point Process (MPP) based vision models are designed to analyze static scenarios, however several applications request object level investigations on multitemporal measurements.

We have introduced a dynamic Multiframe Marked Point Process (FmMPP) model framework for moving object analysis in image sequences, and demonstrated its efficiency for the task of automatic target structure extraction and tracking in series of Inverse Synthetic Aperture Radar (ISAR) images.

Identification and motion analysis of ship targets in airborne Inverse Synthetic Aperture Radar (ISAR) image sequences are key problems of Automatic Target Recognition (ATR) systems which utilize ISAR data. Remotely sensed ISAR images are able to provide valuable information for target classification and recognition in several difficult situations, where more traditional SAR imaging techniques fail. However, robust feature extraction and feature tracking in ISAR image sequences are usually difficult tasks due to noise and the low level of available details about the structure of the imaged targets.

We have proposed a new Multiframe Marked Point Process (FmMPP) model of line segments and point groups for automatic ship and airplane structure extraction and target tracking in ISAR image sequences. A robust joint model has been developed for axis extraction, feature point detection and tracking. For the purpose of dealing with scatterer scintillations and high speckle noise in the ISAR frames, the resulting target sequence has been obtained by an iterative optimization process, which simultaneously considered the observed image data and various prior geometric interaction constraints between the target appearances in the consecutive frames.

Quantitative evaluation has been performed on 8 real ISAR image sequences of different carrier ship and airplane targets, using a test database containing 545 manually annotated frames. We have experimentally shown that in case of noisy sequences, the introduced FmMPP schema can significantly improve the results of the frame-by-frame detection steps.

The proposed approach was published in *IEEE Trans. Geoscience and Remote Sensing* in 2014 (the Principal Investigator (PI) of the project was the first author), and partially presented in a Radar-focused remote sensing conference and at Képaf, (*Benedek* & Martorella, TGRS'14[2]), (*Benedek* & Martorella, IET RADAR'12), (*Benedek* & Martorella, Képaf'13),.

2. Multi-level Object Population Analysis with an Embedded MPP model

The PI has introduced a three-layer Embedded Marked Point Process (EMPP) framework for extracting complex hierarchical object structures from various digital images used by machine vision applications. The proposed method has been demonstrated in three different application areas: optical circuit inspection, built in area analysis in remotely sensed images, and traffic monitoring on airborne Lidar data.

Classical MPP-based image analysis models focus purely on the object level of the scene, and they cannot be suited to hierarchical pattern recognition problems in a straightforward way.
For overcoming the above limitations, the PI has proposed the EMPP framework, which extends conventional Marked Point Process models by two key components: (i) admitting object-subobject ensembles in parent-child relationships and (ii) allowing corresponding objects to form coherent object groups, by a Bayesian segmentation of the population. These two contributions were initially motivated by definite practical requirements from optical circuit inspection, respectively remote sensing traffic monitoring applications, (see 2.1 and 2.2 contributions). After solving tasks specific problems - with OTKA team members and external collaborators -, we have defined and implemented

[2] Please see the bibiliographic details in the project's publication list. OTKA project members in the author field are typeset with *italic*, e.g (*Benedek* & Martorella, TGRS'14).

a general three layer model framework, which has been simultaneously tested and validated in three significantly different domains.

Task 2.1 An automated Bayesian visual inspection method has been introduced for Printed Circuit Board (PCB) assemblies, which is able to simultaneously deal with various shaped Circuit Elements (CE) on multiple scales, by including object-subobject ensembles in the Marked Point Process schema. We have demonstrated the efficiency of the approach on the task of solder paste scooping detection and scoop area estimation, which are important factors regarding PCB qualities.

Automatic optical inspection (AOI) technologies provide very high resolution (10μm) images, thus the analysis needs a hierarchical modeling approach of the PCB structure, focusing jointly on circuit regions, individual Circuit Elements (CEs), CE interactions and characteristic patterns within the CEs, like the geometric scooping artifacts.

We have proposed a new visual inspection method, with the following key properties:

- Hierarchy between objects and object parts is described as a parent-child relationship embedded into the MPP framework.
- To simultaneously deal with variously shaped circuit elements, different types of geometric objects are jointly sampled, by adopting the multi-marked point process schema to the hierarchical entity extraction problem.
- To efficiently sample the SP population space, a Bottom-Up (BU) stochastic object proposal strategy has been developed, by combining low level statistical image descriptors with prior information based structure estimation. This step has kept the computational complexity tractable, although due to properties (i) and (ii) the dimension and size of the solution space were significantly increased.

The proposed method has been evaluated on real PCB data sets containing 125 images with more than 10.000 circuit elements. Performance efficiency has been demonstrated versus a conventional morphology based fault detection technique.

The model was published in *IEEE Trans. Industrial Electronics in 2013* (PI as first author), where the technological background of the selected AOI problem was provided by experts of electronic technology at the Budapest University of Technology and Economics (*Benedek* et. al. IEEE TIE'13).

Task 2.2 A new Two-Level Marked Point Process (L2MPP) approach has been proposed for modeling of objects and object groups within an entity population based on visual features. We have constructed a hierarchical modification of the Multiple Birth and Death optimization algorithm and shown by experiments that its performance is stable and efficient for tasks requiring simultaneous object extraction and population segmentation. Adopting point cloud based descriptors the L2MPP model has been deeply validated in the application field of airborne Lidar based traffic monitoring.

Automatic traffic monitoring is a central goal of urban traffic control, environmental protection and aerial surveillance applications. Complex traffic analysis needs a hierarchical modeling approach: at low level individual vehicles should be detected and separated, meanwhile at a higher level we need to extract coherent traffic segments, by identifying groups of corresponding vehicles, such as cars in a parking lot, or a vehicle queue waiting in front of a traffic light.

By the motivation of the above real-world practical need, we have introduced a new Two-Level Marked Point Process (L2MPP) structure for joint probabilistic modeling of objects a semantically consistent object groups in large entity population. Using our proposed Lidar point cloud based features for scene classification and geometric vehicle description it has been shown that the L2MPP approach can be well suited to the problem of vehicle detection and traffic segmentation from airborne LIDAR data.

A two-level iterative stochastic optimization algorithm has also been proposed which extracted the output vehicle and traffic segment configurations simultaneously.

The new model has been tested with real data of a discrete return Lidar sensor providing up to four range measurement for each laser pulse. Using manually annotated Ground Truth information on a data set containing 1009 vehicles, quantitative object, pixel and group level evaluation results have been provided, showing that the L2MPP model surpasses two earlier grid-cell-based approaches, a 3-D point-cloud-analysis process and a single layer MPP solution.

The L2MPP approach focused on the aerial traffic monitoring application has been published in ***IEEE Trans. Geoscience and Remote Sensing*** in 2015 - both authors were project members - (*Börcs* & *Benedek*, TGRS'15), some model elements have been presented at the ICPR, ISPRS Congress and Képaf conferences, (*Börcs* & *Benedek*, ICPR'12), (*Börcs* & *Benedek*, ISPRS C'12), (*Börcs* & *Benedek*, Képaf'13).

Task 2.3 The PI has defined a general three-layer Embedded Marked Point Process (EMPP) model with a corresponding multi-layer layer energy optimization algorithm, which can simultaneously extract object groups, objects and object parts from high resolution digital images. With ensuring flexible designing options of the data based and prior constraints in the model, we have shown that the EMPP approach can be fit to various real world hierarchical pattern recognition problems. The performance of the new technique has been validated in three different application domains.

Previous approaches on MPP-based hierarchical scene modeling (including the ones introduced in tasks 2.1 and 2.2), have conducted highly tasks specific attempts to model the object encapsulation or the Bayesian object group management issues. Practical experiences showed however, that for such complex, application dependent models, the adaption to another application domain was rarely straightforward, and usually a significant amount of modeling work and code (re-)implementation was needed to transform or modify the framework for a different field. For this reason, in this work, we followed a reverse way by collecting similar tasks appearing in different application areas, and addressing them by a joint methodological approach.

By the proposal of the Embedded Marked Point Process (EMPP) framework, we defined and implemented the structure elements of the complex model and the energy optimization algorithm at the abstract level, while we kept focus on ensuring very simple and highly flexible interfaces to the possible applications.

The proposed method has been demonstrated in three different application areas (from the domains also addressed in previous chapters of this dissertation): built in area analysis in remotely sensed images, traffic monitoring on airborne Lidar data and optical circuit inspection. In addition, a detailed methodological validation process has been conducted.

The general model has been presented (PI as sole author) in the IEEE ICASSP 2014 and ICIAR (LNCS) 2013 conferences (*Benedek*, ICIAR'13), (*Benedek*, ICASSP'14) and in (*Benedek*, Képaf'15).

3. Multi camera people localization in 3D

We have developed a Bayesian approach on multiple people localization in multi-camera systems. First, pixel-level features have been extracted, which were based on physical properties of the 2-D image formation process, and provided information about the head and leg positions of the pedestrians, distinguishing standing and walking people, respectively. Then features from the multiple camera views have been fused to create evidence for the location and height of people in the ground plane. This evidence accurately estimated the leg position even if either the area of interest were only a part of the scene, or the overlap ratio of the silhouettes from irrelevant outside motions with the monitored area were significant. Using this information we have created a 3-D object configuration

model in the real world. We also utilized a prior geometrical constraint, which describes the possible interactions between two pedestrians. To approximate the position of the people, we used a population of 3-D cylinder objects, which was realized by a Marked Point Process. The final configuration results are obtained by an iterative stochastic energy optimization algorithm. The proposed approach has been evaluated on two publicly available datasets, and compared to a recent state-of-the-art technique.
To obtain relevant quantitative test results, a 3-D Ground Truth annotation of the real pedestrian locations has been prepared, while two different error metrics and various parameter settings were evaluated, showing the advantages of our proposed model. The proposed annotation tool allowed the user to accurately select the ground occupancy of people by aligning an oriented rectangle on the ground plane. In addition, the height of the people could also be adjusted. In order to achieve precise ground truth data the user was aided by the video frames of multiple synchronized and calibrated cameras. Finally, the 3D annotation data could be easily converted to 2D image positions using the available calibration matrices.
The proposed approach has been published in *IEEE Trans. Circuits and Systems for Video Technology* in 2013 - PI as second author - (Utasi & *Benedek*, IEEE TCSVT'13), and a conference paper has been prepared (Utasi and *Benedek* VIGTA'12).

4. Dynamic environment perception and 4D reconstruction using a mobile Rotating Multi-beam Lidar sensor

We presented various vision and reconstruction algorithms for the data streams of a Rotating Multi-beam Lidar sensor. An overview of our corresponding results will appear in an invited book chapter of the book "Visual Perception and Intelligent Control for Single and Multiple Robots", to be published by Springer in 2015. The Reader can find the details in the corresponding publications regarding each task, as follows.

Task 4.1 - *3D people surveillance using static Lidar platforms* We have proposed an approach on real-time 3D people surveillance, with probabilistic foreground modeling, multiple person tracking and on-line re-identification using terrestrial Lidar point cloud sequences. We have exploited here our background subtraction and multi target tracking modules (presented in the last year's report), and as novelty we focused on the re-assignment of the temporarily lost trajectories during the tracking process, which is usually inevitable for a scenario with 6-8 people. For this reason, we have derived so called weak biometric identifiers from the Lidar sequence, featuring the dressing, height and uniques gait prints of the people, and used these features for re-identification. Quantitative evaluation has been performed on seven outdoor Lidar sequences containing various multi pedestrian scenarios.
The 3D people surveillance approach was published in *Pattern Recognition Letters* (PI as sole author), (*Benedek*, PRL'14), some further improvement are presented in (*Nagy*, *Benedek* & Jankó, Képaf'15).

Task 4.2 *Large scale urban scene analysis and reconstruction* We have proposed a joint approach on virtual city reconstruction and dynamic scene analysis based on point cloud sequences of a single car-mounted Rotating Multi-Beam (RMB) Lidar sensor. The aim of the addressed work is to create 4D spatio-temporal models of large dynamic urban scenes containing various moving and static objects. Standalone RMB Lidar devices have been frequently applied in robot navigation tasks and proved to be efficient in moving object detection and recognition. However, they have not been widely exploited yet for geometric approximation of ground surfaces and building facades due to the sparseness and inhomogeneous density of the individual point cloud scans. We have developed here an automatic registration method of the consecutive scans without any additional sensor information such as IMU, and introduced a process for simultaneously extracting reconstructed surfaces, motion information and objects from the registered dense point cloud completed with point time stamp information. We also perform real-time localization and identification of typical urban objects, such as traffic signs, vehicles or crosswalks. In contrast to most existing works, the proposed algorithm does not use hand-labeled

training datasets to perform object classification. Experimental results are carried out on real LIDAR measurements in the streets of Budapest.
The results were published at ISPRS and IEEE conferences (*Józsa*, *Börcs* and *Benedek*, VCM'13) , (*Börcs*, *Józsa* and *Benedek*, IEEE CBMI'13).

Task 4.3 *On board real time vehicle detection for autonomous cars* Detection of vehicles in crowded 3-D urban scenes is a challenging problem in many computer vision related research fields, such as robot perception, autonomous driving, self-localization, and mapping. In this work we presented a model-based approach to solve the recognition problem from 3-D range data. In particular, we aim to detect and recognize vehicles from continuously streamed LIDAR point cloud sequences of a rotating multi-beam laser scanner. The end-to-end pipeline of our framework working on the raw streams of 3D urban laser data consists of three steps:

1) producing distinct groups of points which represent different urban objects In the proposed framework we present a simple, yet efficient hierarchical grid data structure and corresponding algorithms that significantly improve the processing speed of the object detection task. Furthermore, we show that this approach confidently handles streaming data, and provides a speedup of two orders of magnitude, with increased detection accuracy compared to a baseline connected component analysis algorithm
2) extracting reliable 3-D shape descriptors specifically designed for vehicles, considering the need for fast processing speed
3) executing binary classification on the extracted descriptors in order to perform vehicle detection. The extraction of our efficient shape descriptors provides a significant speedup with and increased detection accuracy compared to a PCA based 3-D bounding box fitting method used as baseline.

The results were published at workshops of ECCV 2014 and ACCV 2014 conferences (*Börcs*, *Nagy* & *Benedek*, RCUAD'14) , (*Börcs*, *Nagy*, Baticz & *Benedek*, SUAS'14) . From an overview of the above contributions and invited book chapter has been submitted (*Börcs*, *Nagy* & *Benedek*, Book Springer'15). Further related publications are (*Börcs*, *Nagy* & *Benedek*, COGINFOCOM'13) and (*Börcs*, *Nagy* & *Benedek*, Képaf'15).

Task 4.4: *An Integrated 4D Vision and Visualization System:* Efforts on real time reconstruction of 3D dynamic scenes receive also great interest in video communication and augmented reality systems. Obtaining realistic 4D video flows of real world scenarios may result in a significantly improved visual experience for the observer compared to watching conventional video streams, since a reconstructed 4D scene can be viewed and analyzed from an arbitrary viewpoint, and virtually modified by the user. However, building an interactive 4D video system is highly challenging, as it needs in parallel automatic perception, interpretation, and real time visualization of the environment
We have developed a pilot system, called "integrated 4D" (i4D) for reconstruction and visualization of complex spatio-temporal scenes by integrating two different types of data: outdoor 4D data measured by a rotating multi-beam LIDAR sensor, and 4D models of moving actors obtained in a 4D studio. A typical scenario is an outdoor scene with multiple walking pedestrians.
The LIDAR monitors the scene from a fixed position and provides a dynamic point cloud. This information is processed to build a 3D model of the environment and detect and track the pedestrians. Each of them is represented by a point cluster and a trajectory. A moving cluster is then substituted by a detailed 4D model created in the studio. The output is a geometrically reconstructed and textured scene with avatars that follow in real time the trajectories of the pedestrians.
The scientific and technical novelties of the i4D system were described in an international patent, and published at the ICVS 2013 (LNCS) conference (*Benedek* et. al, ICVS'13) and at (*Horváth* et. al, Képaf"13).

5. Survey on Multilayer Markov Random Field Models for Change Detection in Optical Remote Sensing Images

We prepared a comparative study on three Multilayer Markov Random Field (MRF) based solutions proposed for change detection in optical remote sensing images. The Multicue MRF model integrates

two different features in a three-layer framework, based on various data-dependent and label fusion constraints. The Conditional Mixed Markov model has a four-layer structure, where the fourth layer creates configurable links between the remaining model components. The Fusion MRF considers color, texture and cross layer similarity information by proposing a joint multilevel segmentation and cluster selection technique. In the beginning of the paper, we highlight the significance of the focused model family by giving an overview on various existing approaches on change detection. Then, we introduce the three models following the same presentation scheme so that the reader can follow the main similarities and differences in the model structures, used features and the working constraints.
Finally, we provided qualitative and quantitative comparison results using a publicly available change detection database which contains aerial image pairs and Ground Truth (GT) change masks, extended with some additional images. We concluded that the discussed models are competitive against alternative state-of-the-art solutions proposed for the same task. However, we also demonstrate that the relevant GT information and the appropriate evaluation metrics may strongly depend on the needs of the concrete applications, therefore these factors should be carefully chosen during the validation process.
The overview paper received a *Minor Revision* decision in October 2014, upon acceptance it will be published in *ISPRS Journal of Photogrammetry and Remote Sensing* (*Benedek* et. al, ISPRS J'15), Special issue " Multitemporal remote sensing change detection", in 2015 (PI as first author).

6. Virtual city model reconstruction from airborne Lidar point clouds

We have developed automatic and robust algorithms producing detailed 3D virtual city models from aerial Lidar measurements. The proposed approach is able to construct realistic models for rooftops following different architectural trends, and to process large Lidar measurements covering areas of several square kilometers without user interaction.
As for our methodological contributions, first we have designed a procedure for surface normal based roof segmentation, which takes into account the boundaries of each roof segment, so that the adjacent segments connect without gaps. Second we have developed an algorithm to detect 3D edge lines of the rooftops, and a further polygon generator on the basis of these boundary lines. Third, for overcoming the limitations of triangulation algorithms working on convex hulls of the point set, we have proposed Markov Random Field based method for filtering out the incorrect triangles lying on the concave parts.
We have tested our approach on large datasets, containing clouds of over 8 million points covering a 240m x 2.3km large region and on other datasets of smaller territories having area up to 200m2.
This work was conducted with a BSc student, results have been published in the national GRAFGEO conference (*Polcz* & *Benedek*, GRAFGEO'14).