Research Report of OTKA #**106374**

"***Finding Focus of Interest in freely configured sensor networks***"

during February 1 of 2013 – January 31 of 2016

by **Tamás Szirányi**

<u>**Összefoglaló**</u>:

A project célja annak vizsgálata volt, hogy a szenzorok (főként kamerák és infra/lézer eszközök) 3D terében milyen módon lehet a helyszín eseményeit detektálni, követni, felismerni, miközben a követett céltárgyak átmozognak más szenzorok terébe, és a mozgásról és alakzatokról nem rendelkezünk segítő információval. Ehhez meg kellett oldani olyan kérdéseket, mint a

1. Különböző modalitású érzékelők össze-regisztrálása 3D-ben;
2. Alakzatok (elsősorban járókelők) detektálása és követése;
3. A megfigyelt alakzatok kiemelése (Feature of Interest, FOI) és a különböző nézetekhez tartozó FOI-k összeépítése 3D-ben;
4. 3D követés nem egyenletes alapsík esetére;
5. Helyszín alaprajzának detektálása strukturális jellemzők kinyerésével;
6. Képi és videó adatbázisok rendszerezése és visszakeresése;
7. Az ágens hálózatban a hierarchiák keresése;
8. Mozgó objektumok körüli szenzorok figyelmének (FOI) szervezése;
9. Alakzatok felismerésének tanítása (mozgó és álló objektumok is);
10. Tanítóhalmazok (ground truth) ellenőrzött előállítása;
11. Nézetek összedolgozása (wide/random baseline esetekre).

A munka során valamennyi fontos kérdésben sikerült jelentős eredményeket elérni. Az eredmények alapján körvonalazódnak ennek a nagyon összetett témakörnek a lehetséges megoldási lehetőségei. A kidolgozás során számos érdekes egyéb tudományos kérdés is felmerült, ami a munka továbbvitelét inspirálja.

Summary:

The goal of the project was an investigation on sensor networks (containing optical and infra cameras, laser scanners), about how they can be used for detection, tracking and recognize in 3D, while we do not have additional information about the events and entities. For this purpose we should have solved the problems of

1. Registration of different sensor modalities in 3D;
2. Object (e.g. pedestrian) detection and tracking;
3. Defining feature of interest areas (FOI) of tracked objects, and registering the sensors by using their FOI-s;
4. Tracking in 3D in case of non-homograph grounds;
5. Detection of scene footprint based on new structural features;
6. Indexing and retrieval in image- and video-database;
7. Searching for hierarchies in agents' network;
8. FOI definition in large scale sensor networks is tracking objects;
9. Training for recognition of moving objects;
10. Generating Ground-truth and its validation process;
11. Fusing different camera scans of a chain of independent views.

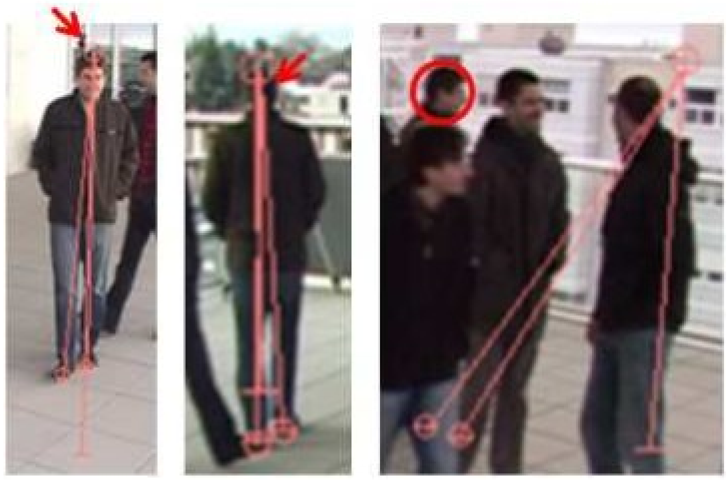
During the project all the important issues have been addressed and important results have been achieved. Based on the new results we can better outline the possible solutions for this complex topic. The project work initiated several new ideas and scientific problems, which should be considered in the future continuation of the efforts.

Research Report

First year

Report on the execution and results of planned actions:
Planned Task #1: *Building the testing laboratory*

Result: We have built up a multimodal/multisensor laboratory setup, implementing fixed network of high definition cameras, thermal cameras, Time-of-Light depth-sensors; the laboratory's LIDAR equipment is also available in the set-up. The research needs a huge amount of ground-truth measurement data (with human annotations. For each action we have generated these data/sets, and we also have investigated research on their usability for validation purposes. For the evaluation of the experiments we have created a multi-view GroundTruth database and qualification methodology: ”Evaluation of Manually Created Ground Truth for Multi-view People Localization” [4].



Using two views facing each other - the camera is marked with red arrows. Consequently, the lines are close to parallel in the reference space. As a result, small errors in marked position leads to incorrect position (matrix of problem will be poorly conditioned).
Ground-truth generating human subject didn't spot the error, however it is quite obvious.

Further devices (indoor UAV and miniaturized depth sensor cameras), and a system of tracking theatre are planned to install to continue the development of the laboratory.

Planned Task #2: *Data comparison of time-series or recognized patterns for correlating sensors' FOIs and finding joint objects*
Result: We have developed a method for fusion of multimodal camera sources in real-time. The proposed system performs automatic sensor registration of FOI areas and real-time feature level fusion by using visible light, thermal and depth imageries. Our approach utilizes the vertical plane homography and its transformation in depth. This method can fit the ROI of different sources in real-time, even in the case of missing/distorted imaging information from one of the sources: ”Calibrationless sensor fusion using linear optimization for depth

matching” [11]. The search for a FOI area is estimated by using our previous innovation of Co-motion analysis[1].

In these pictures one can see in the first column the outputs of a depth sensor, in the second column the outputs of a conventional camera and in the third column the image after merging the motion masks of these two sensors

- *First row: The object is placed behind the proper plane of the homography*
- *Second row: The object is placed in the proper plane of the homography*
- *Third row: The object is placed before the proper plane of the homography*

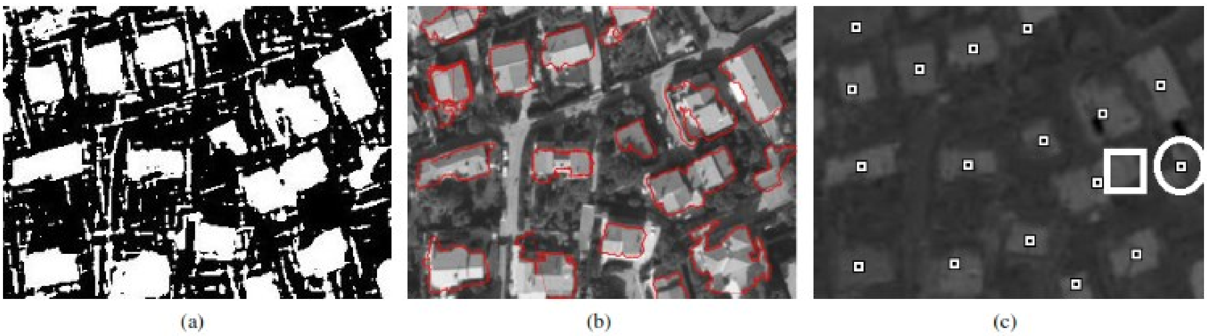
With using a new method based on ***geometric primitives we can localize people*** in the multiple views by detection of their feet.. The altitude of location is also retrieved, which eliminates the need of planar ground – which is a common restriction in the related literature: ”Localizing people in multi-view environment using height map reconstruction in real-time” [3] and ”Multi-view People Detection on Arbitrary Ground in Real-Time” [2].

[1] Z. Szlávik, T. Szirányi, L. Havasi, ”Stochastic view registration of overlapping cameras based on arbitrary motion”, **IEEE Tr. Image Processing**, Vol.16, No.3, pp.710 - 720, **2007**

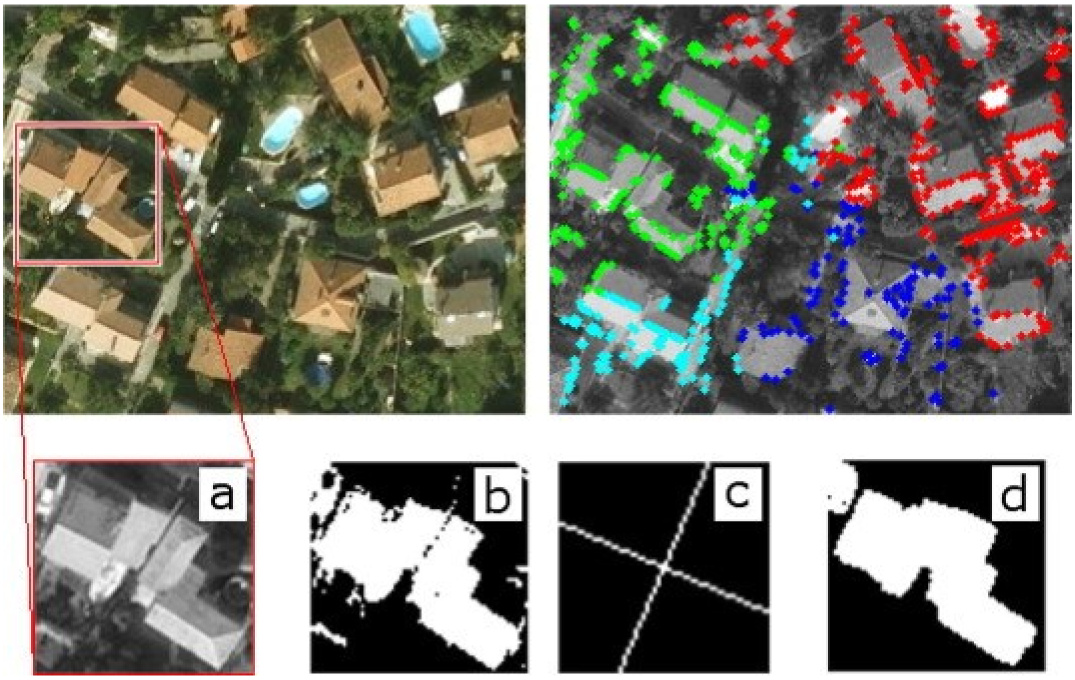
Sample results of localization and determining height map.

For remote positioned cameras a new building shape (footprint) detection algorithm has been developed; here the orientation information in the feature point neighborhoods is analyzed to define the most probable orientations.

Accurate building contours have been given without shape templates: “Multidirectional Building Detection in Aerial Images Without Shape Templates" [6][19]:



Steps of multidirectional building detection: (a) is the connectivity map; (b) shows the detected building contours in red; (c): marks the estimated location (center of the outlined area) of the detected buildings, the falsely detected object is marked with a white circle, missed object is marked with a white rectangle.



The building detection process: The first row shows the original image with the marked sample area; and on the right the result of the local orientation analysis with different directions marked with different colors. In the second row: (a) shows the sample building candidate area; (b) is the result of the Chan–Vese active contour algorithm; (c) shows the main directions [-24,66] of the area and (d) is the result of the orientation selective refinement process.

Planned Task #3: *Hierarchical Hidden Markov Models for the semantic description at a higher level meaning. At this stage, finding something similar in motion or structure among different modalities is the main purpose.*

Result: Semantic description at a higher level meaning. A multilayer fusion model for adaptive segmentation and change detection of optical remote sensing image series has been developed, where trajectory analysis or direct comparison is not applicable. Our method applies unsupervised or partly supervised clustering on a fused-image series by using cross-layer similarity measure, followed by multilayer Markov random field segmentation: ”Segmentation of remote sensing images using similarity measure based fusion-MRF” [7], ”Improved segmentation of a series of remote sensing images by using a fusion MRF model” [13].

Planned Task #4: *Indexing and retrieval methods based on the fused modalities The comparison of categories of detected objects/events from different sensors through cross-modal interaction is a main challenge here since the correlation should be estimated from different modalities and disjunctive range of parameter sets.*

Result: A flexible framework for research purposes has been built for testing features, metrics, distances and indexing structures. The core part of the content based retrieval system is the LHI-tree, a disk-based index scheme for fast retrieval of multimodal features. The demonstration focused on the content based retrieval of Wikipedia images (Hungarian version): ”Search in WikiImages using Mobile Phone” [1].

Second Year:

Planned Tass:

- Graph model of the sensor system: Sensors are represented by nodes, respecting geometrical constraints; graph topologies and connectivity analysis will be based on the FOI coherence. Finding the appropriate graph structure will result in a novel formalism describing the coherence of a sensor cluster, finding optimal dense sub-graphs for grouping, and defining geometry based graph-optimization of sensor networks.

- Beside the agents, it can be useful to model some details of the environment as graphs as well, for example monitoring the obstacles around the agents.

- Investigating the directional constraints of the connectedness of a subgraph around a FOI, which may lead to temporary structures of the graphs, where the transitions may be controlled by continuously checking the eigenvalues of the Laplacian graph.

Solutions and results:

Agent network analysis:

- The VW factory in Wolfsburg supported us with data for analysis in a hierarchy and connectedness searching task in a network of software agents. These data characterized well our problem of real robot/sensor agents. We developed an analysis methodology for *Unsupervised Error Estimation of Autonomous Agent Networks*.

We worked out a method on how to investigate large software networks if we have a very limited information available to help us in the work. We know the software modules (nodes) and the communication lines ”calls” between the nodes (arcs) and nothing else. The proposed method can be used if the industrial partner does not want to share too much information with us not even any dynamic information about the network. Then our task is to find anomalous activity in a network without any knowledge of the activities. We try to highlight some nodes or subgraphs that can be faults and this will help our partner to find the source of the problem more efficiently. First we give an estimation of what could be the smaller processing units of the system. These units might work independently. Then we try to establish a pattern that should be followed by the system. This pattern should be adaptive and should be able to distinguish true randomness from artificial one. Our method is worked out to filter the subgraphs of the system and find the faulty ones with high probability. Our aim is to find a smaller part of the industrial partner’s system on what he should run an error fixing algorithm in order to save time and energy by decreasing the number of nodes to inspect.

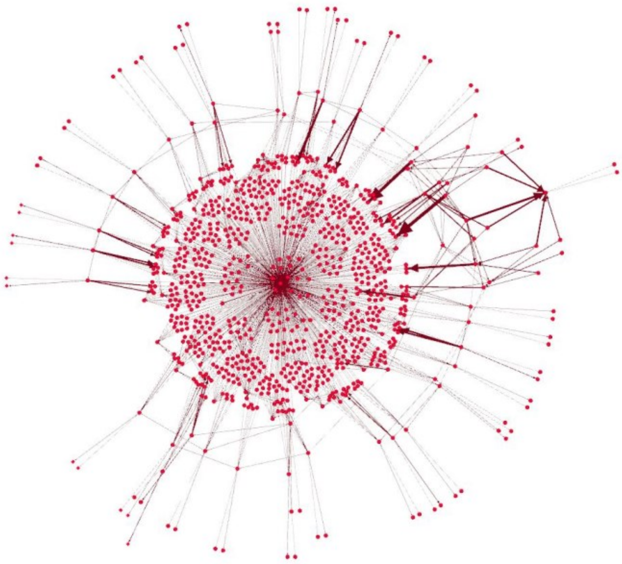
A preliminary result of our laboratory was when Keszler, Szirányi and Tuza[2] presented an algorithm to cluster graphs based on dense bipartite subgraph mining. Their method can be applied if we define abnormality with the help of density but we give some other heuristic definitions of abnormality at the end of this paper.

As main result, we proposed an algorithm, which can efficiently choose some subsystems of the whole network for further investigation. To do that, we defined four types of possible failures, and we gave a method to find small subsystems which can contain these failures. So

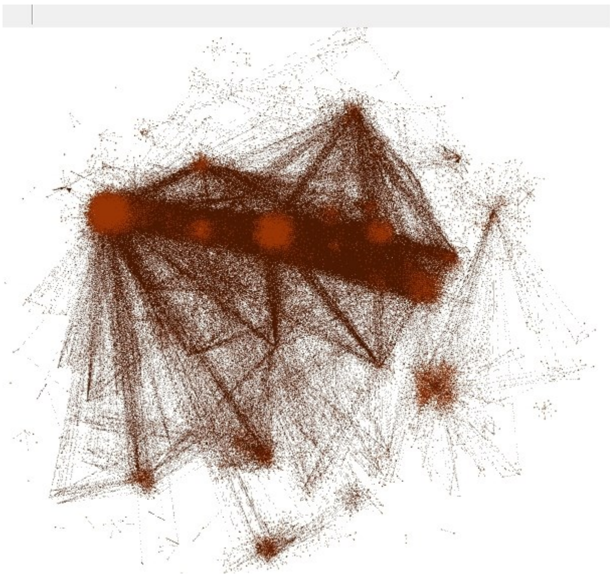
---

[2] A. Keszler, T. Szirányi and Z. Tuza: Dense Subgraph Mining With a Mixed Model, Pattern Recognition Letters, 2013, 34(11) 1252–1262.

we can give an error estimation of the system, using only structural information, and this will help to find the source of the problem more efficiently.



Picture of a centralized egonetwork

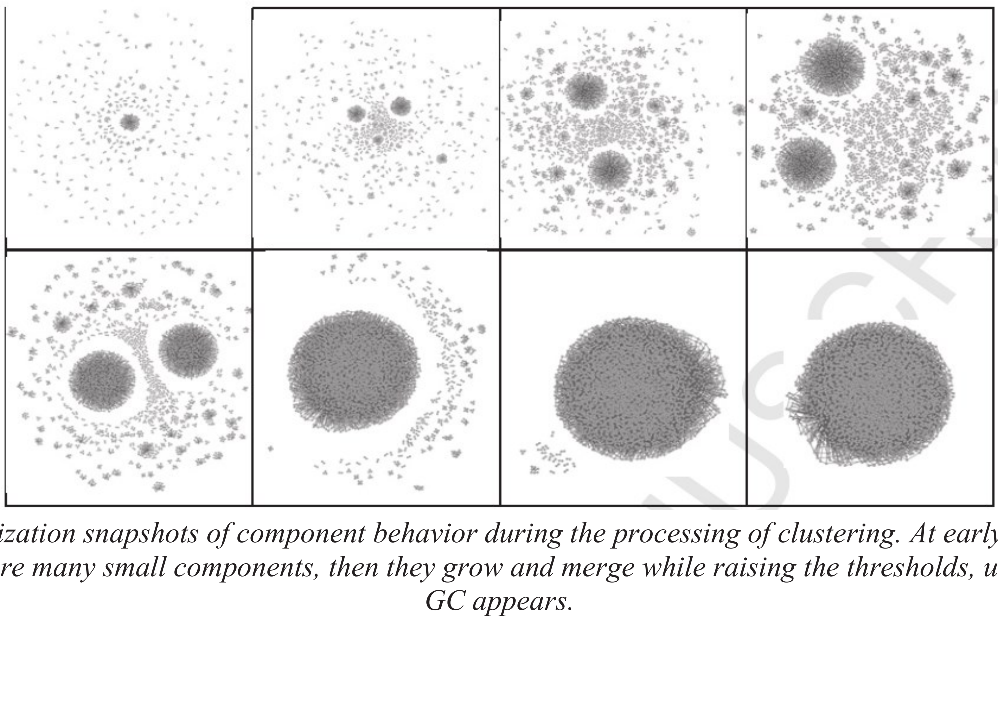


Our test database

Picture of a distributed egonetwork after analysing the connection system

Connectedness analysis:

- In the connectedness analysis of big agent-systems, where the connection is described by the geometrical distances, the giant component problem of clustering becomes a main issue. The mathematical model, following that of the Erdős-Rényi model of uniform random graphs, can characterize the geometrical connectedness of an agent system, but it also characterizes the descriptor graphs of multimedia database, including the retrieved patterns of high dimensionality. We analyze the phase transition and the evolution of components in such graphs, and based on their behavior, the corresponding descriptors are compared, ranked, and validated in retrieval tests.

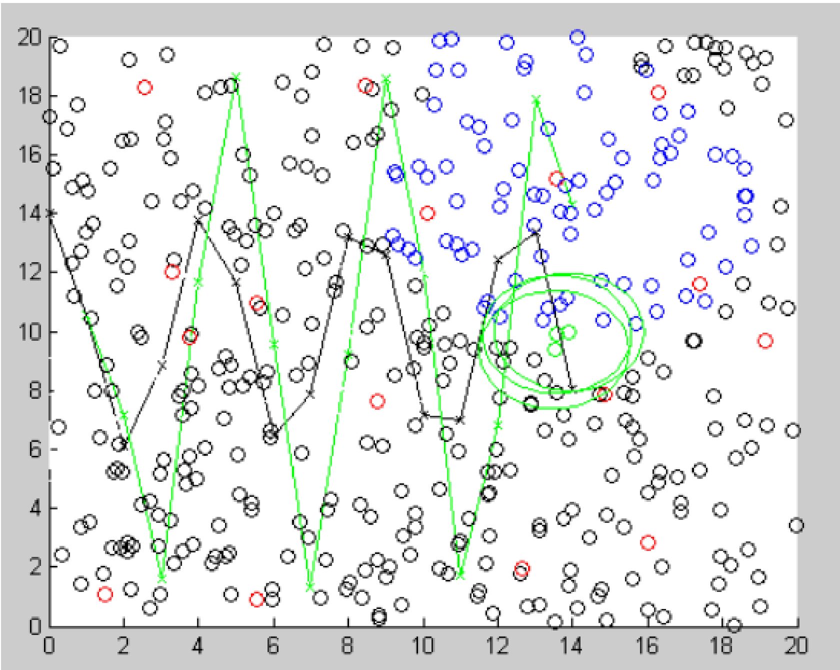


Visualization snapshots of component behavior during the processing of clustering. At early steps, there are many small components, then they grow and merge while raising the thresholds, until the GC appears.

Graph based modelling of the sensor-system:

- We have collected and compared several methods to show the possible graph-base solutions for arbitrarily positioned sensors, where the sensor-agents are characterized by local metric spaces, and the whole networks is to be organized in a continuously coherent system. The solution is given by methods in graph theory of realization and localization problems [17]. Solutions are given in case of anchor points and in others cases without them.

- One of the main applications of large scale sensor networks is tracking objects. In case of distributed networks we have to determine the sensors which will make the computational tasks, and considering the low energy consumption. Because of this we try to turn on as few sensors as possible. Another objective is to decrease the computational capacity and communication, too. We implemented the algorithm DPT (Distributed Predictive Tracking) found in the paper of Yang and Sikdar[3] [1]. This algorithm was developed for tracking the movement of a single object covered by random distributed sensors. The sensors are fixed and capable to measure their own distance from the object. Our aim was to modify the algorithm and to use as few sensors as possible at the same time (energy consumption). We have developed a model [21] for tracking an object with a changing cluster of possible sensors:



Black: turned-off sensors, Red: cell pointers, blue: ready-on sensors, green: turn-on sensors (and their effective radius), black line: track of the object, green line: estimated track of the object

Image fusion based on connectedness:

- In a related work, dealing with geometry and image registration problems[4], my colleagues (partly involved in the present project on the connected problem of sensor network optimization) developed a method for local processing of photos and associated sensor

[3] Yang, H., and Sikdar, B., A Protocol for Tracking Mobile Targets using Sensor Networks. IEEE Workshop on Sensor Network Protocols and Applications, 2003

[4] Tanacs A, Majdik A, Hajder L, Molnar J, Santa Zs, Kato Z: Collaborative Mobile 3D Reconstruction of Urban Scenes, ACCV Workshop on Intelligent Mobile and Egocentric Vision (ACCV-IMEV), Singapore, Springer, 2015

information on mobile devices. Its goal is to lay the image clustering and registration foundations of a collaborative multi-user framework where ad-hoc device groups can share their data around a geographical location to produce more complex composited views of the area, without the need of a centralized server-client - cloud-based - architecture. It focused on processing as much data locally on the devices as possible, and reducing the amount of data that needs to be shared. The main results are the proposal of a lightweight processing and feature extraction framework, based on the analysis of vision graphs, and presenting preliminary composite view generation based on these results. [5] This work is an important contribution to the approach of the present project, as it gives the methodology how to fuse images in the self-organized sensor network. In the present project we developed further this methodology for finding automatic navigation [18] and sensor organization in case of huge number of possible sensors [17] and multimodal sensors [16].

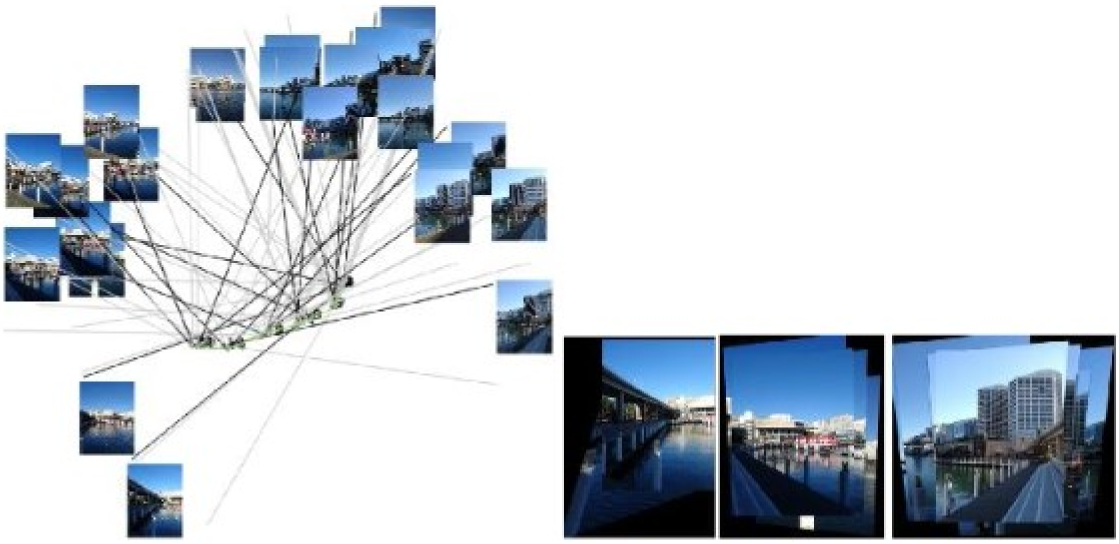


Image shots made by mobile phones, organized by the help of geo-data view information in a connection-graph, and fused in common view on the right.[6]

Change detection by using multiple sources:

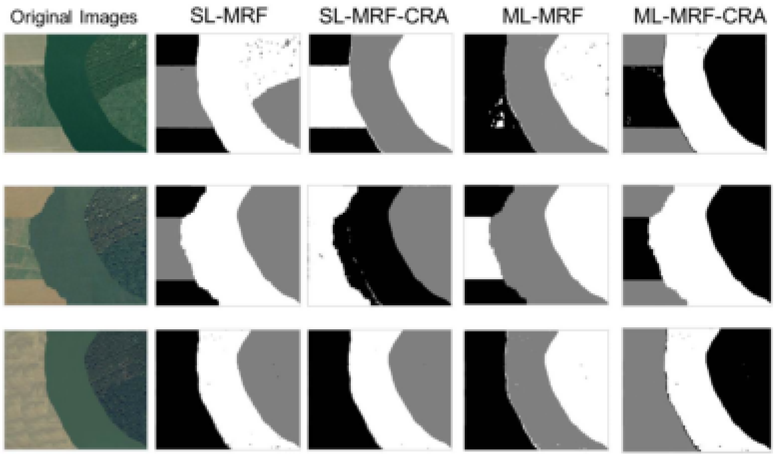
- A main problem of environment detention is to find changes through image information. Usually it is a non-trivial task because of the different shooting conditions and multiple image sources. Change detection itself is a basic task, when we segment the image into labelled areas, and the detection of changes between of labelled segments of images of different time instances is to be done. For processing multiple sources in different time instants, we have developed a multilayer fusion based Markov Random Field segmentation method; this method works in unsupervised or partly supervised training environment. In case of

---

[5] http://web.eee.sztaki.hu/~kla/cvcp13.html

[6] L. Kovács: *Processing Geotagged Image Sets for Collaborative Compositing and View Construction*, in Proceedings of 2013 IEEE International Conference on Computer Vision Workshops (IEEE Intl. Workshop on Computer Vision for Converging Perspectives in conjunction with ICCV 2013), pp. 460-467 (DOI 10.1109/ICCVW.2013.67, ISBN: 978-0-7695-5161-6).

unsupervised condition, a global super-mapping is done on the stack of images of different time instances, and it results in a training set for the individual layers [14]. The method has been tested on remote sensing images of controlled ground truth conditions.



Unsupervised segmentation results for the multi-layer image segmentation for meadow, river and three variants of forest/bush categories; the proposed similarity measure based fusion-MRF model gives the best result with automatically synchronized cluster definition.

Calibrationless sensor fusion:

- Recently the observation of surveillanced areas scanned by multi-camera systems is getting more and more popular. The newly developed sensors give new opportunities for exploiting novel features.

Using the information gained from a conventional camera we have data about the colours, the shape of objects and the micro-structures; and we have additional information while using thermal camera in the darkness. A camera with depth sensor can find the motion and the position of an object in space even in the case when conventional cameras are unusable.

How can we register the corresponding elements on different pictures? There are numerous approaches to the solution of this problem. One of the most used solutions is that the registration is based on the motion. In this method it is not necessary to look for the main features on the pictures to register the related objects, since the features would be different because of the different properties of the cameras. It is easier and faster if the registration is based on the motion. But other problems will arise in this case: shadows or shiny specular surfaces cause problems at the motion.

We have shown that registering the corresponding elements in a multi-camera system we can find a homography between the image planes in real time, so that we can register a moving object in the images of different cameras based on the depth information [11].

Third year:

Planned tasks: *Patching the scene of the HS ROI from inter-agent data: using state-of-the-art methods for processing structure-from-motion, stereo imaging, or mosaicing of the*

environment by registering/stitching the scanned views. Multiview/multitarget tracking will be used by implicit location detection.
We can build a model of the environment, based on a given set of geometries and a codebook of possible shapes and events.

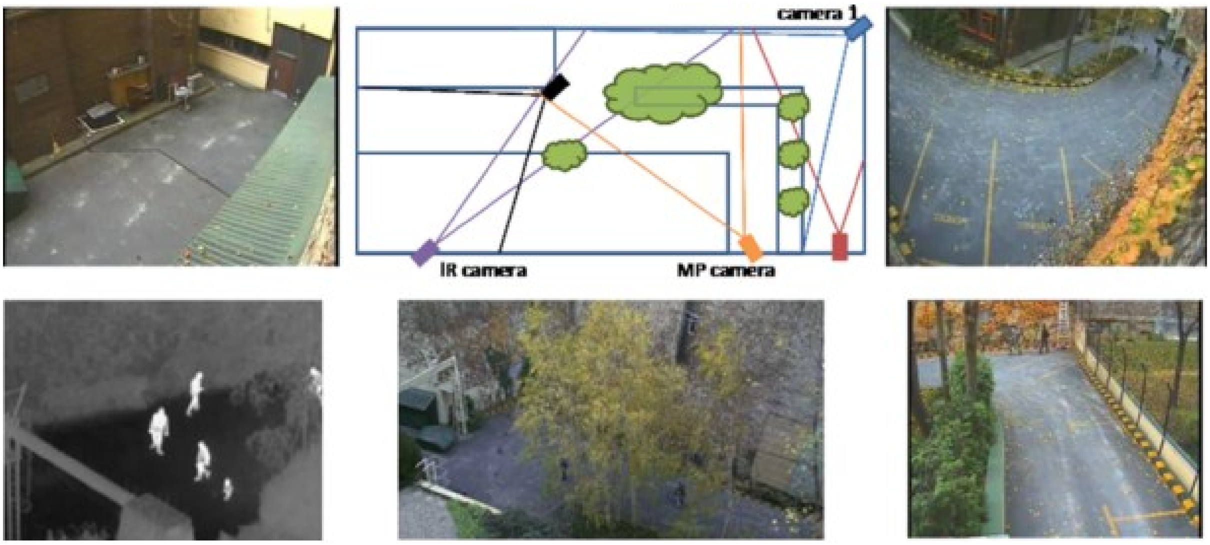
Solutions and results:

Structure from one view in motion: A special case of sensor positioning when obstacle detection is done in a semi-metric space. We presented a method [18] for low computational complexity single image based obstacle detection and avoidance, with applicability on low power devices and sensors. The method is built on a novel application of single image relative focus map estimation[7], using localized blind deconvolution, for classifying image regions. For evaluation we use the MSRA datasets and show the method's practical usability by implementation on smartphones.

Shape and event detection:
Combining multiple observation views has proven beneficial for pedestrian tracking. We developed [16] a methodology for tracking pedestrians in an uncalibrated multi-view camera network. Using a set of color and infrared cameras, we can accurately tracking pedestrians for a general scene configuration. We design an algorithmic framework that can be generalized to an arbitrary number of cameras. A novel pedestrian detection algorithm [20] based on Center-symmetric Local Binary Patterns is integrated into the proposed system. In our experiments the common field of view of two neighbouring cameras was about 30%. The system improves upon existing systems in the following ways: (1) The system registers partially overlapping camera-views automatically and does not require any manual input. (2) The system reaches the state-of-the-art performance when the common field of view of any two cameras is low and successfully integrates optical and infrared cameras. Our experiments also demonstrate that the proposed architecture is able to provide robust, real-time input to a video surveillance system. Our system was tested in a multi-view, outdoor environment with uncalibrated cameras.

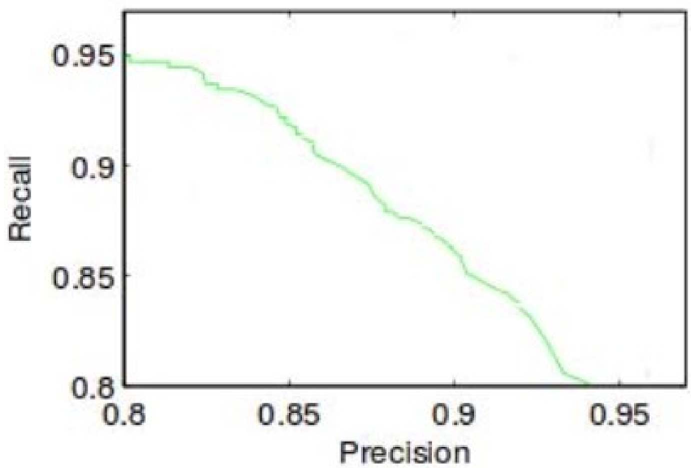
---

[7] Kovács, L., Szirányi, T.: Focus area extraction by blind deconvolution for defining regions of interest , (2007) *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 (6), pp. 1080-1085.



The test environment of our measurement.

Frames of three different cameras and the detection results
(the third camera was an infra one).



ROC curve of tracking measured in function of detection threshold.

Conclusion and epilogue

The project at the beginning addressed several key problems emerging in sensor networks and multimodal systems. Each key problem issue has been considered during the project. The project itself has been depending on the sensor network hardware and the extracted data for processing. Most of the data must be annotated for running learning systems (supervised and unsupervised). Before the project we considered other (independent) consortia projects with large infrastructure, from which we can gain data streams for evaluation and modelling. It was partly successful, since the FP7 ProActive project supported us with some data, but the level of annotation and the complexity of the network were limited. For this reason we should work on video data annotation problems first, what was not planned in the project earlier. Finally, for each separate problem we could generate good annotations, like those published in: [4, 9, 10, 11, 12, and 14]. It was a great fortune that the VW factory looked us with a problem about hierarchy searching in a practical agent network [9], which problem was identically same as our sensor clustering task in the mathematical sense; this investigation initiated an interesting research in graph theory and its application.

Most of our results are published during the project:

- Fusion of sensors in a network;
- Object tracking in 3D, FOI definitions;
- Object recognition in 2D and 3D;
- Scene detection with new features;
- Change detection of scenes on the half-semantic (label) level;
- Graph decomposition for agents' network;
- Graph interpretation in case of huge networks: Giant Component problem for geometric graphs;
- Navigating in 3D by a camera-system.

Other results are under investigation, evaluation and processing (e.g. [9, 21, 22]), planning publications in the near future. The support of the OTKA was fundamental for collecting excellent young colleagues in a good cooperation, and addressing really challenging topics.

Publication list of the project

[1] Havasi László, Szabó Mihály, Pataki Máté, Varga Domonkos, Szirányi Tamás, Kovács László: Search in WikiImages using Mobile Phone, In: Czúni L, Schöffmann K, Szirányi T (szerk.) 11th International Workshop on Content-Based Multimedia Indexing. CBMI 2013. Veszprém: IEEE, 2013. pp. 219-222., 2013

[2] Kiss Ákos, Szirányi Tamás: Multi-view People Detection on Arbitrary Ground in Real-Time, In: Sebastiano Battiato, José Braz (szerk.) (szerk.) VISAPP 2013 - Proceedings of the International Conference on Computer Vision Theory and Applications, Volume 1. Barcelona: SciTePress, 2013. pp. 675-680., 2013

[3] Kiss Ákos, Szirányi Tamás: Localizing people in multi-view environment using height map reconstruction in real-time, PATTERN RECOGN LETT 34: (16) 2135-2143, 2013

[4] Kiss Ákos, Szirányi Tamás: Evaluation of Manually Created Ground Truth for Multi-view People Localization, In: Concetto Spampinato (szerk.) VIGTA '13 Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications. New York: ACM Press, 2013. pp. 1-6., 2013

[5] Kovács Andrea, Szirányi Tamás: Improved Harris Feature Point Set for Orientation-Sensitive Urban-Area Detection in Aerial Images, IEEE GEOSCI REMOTE S 10: (4) 796-800, 2013

[6] Manno-Kovács A, Szirányi T: Multidirectional Building Detection in Aerial Images Without Shape Templates, In: (szerk.) International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W1, ISPRS Hannover Workshop . Hannover, Németország, 2013.

[7] Szirányi Tamás, Shadaydeh Maha: Improved segmentation of a series of remote sensing images by using a fusion MRF model, In: Czúni L, Schöffmann K, Szirányi T (szerk.) 11th International Workshop on Content-Based Multimedia Indexing. CBMI 2013. Veszprém: IEEE, 2013. pp. 137-142., 2013

[8] Attila Kiss, András Recski: Universal Spacings for the 3-Dimensional VLSI Routing in the Cube, Int. Colloquium on Graph Theory and Combinatorics, Grenoble, France, 2014.06.30-07.04., 2014

[9] Attila Kiss, Anna Mándli, Tamás Szirányi: Unsupervised Error Estimation of Autonomous Agent Networks, Seventh Cracow Conference On Graph Theory, Rytro, Poland, 2014, 2014

[10] Havasi L, Varga D, Szirányi T: LHI-tree: An efficient disk-based image search application, IWCIM 2014, IEEE, 2014

[11] Havasi László, Kiss Attila, Spórás László, Szirányi Tamás: Calibrationless sensor fusion using linear optimization for depth matching, In: International Workshop on Combinatorial Image Analysis: IWCIA 2014. Brno, Csehország, 2014.05.28-2014.05.30. #18. (Lecture Notes in Computer Science) LNCS, 2014

[12] Kovács Levente, Keszler Anita, Szirányi Tamás: Component evolution analysis in descriptor graphs for descriptor ranking, DIGITAL SIGNAL PROCESSING 31:(8) pp. 1-12. (2014), 2014

[13] Shadaydeh, M., Sziranyi, T.: An improved local similarity measure estimation for change detection in remote sensing images, IEEE Int.Conf. Aerospace Electronics and Remote Sensing Technology (ICARES), 2014

[14] Szirányi Tamás, Shadaydeh Maha: Segmentation of remote sensing images using similarity measure based fusion-MRF model, IEEE Geoscience Remote Sens. Letters, V.11, 2014

[15] Csaba Benedek, Maha Shadaydeh, Zoltan Kato, Tamas Sziranyi, Josiane Zerubia: Multilayer Markov Random Field Models for Change Detection in Optical Remote Sensing Images, ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING, 2015

[16] Domonkos Varga, Tamás Szirányi, László o Havasi, Attila Kiss, László Spórás: A multi-view pedestrian tracking method in an uncalibrated camera network, ICCV 2015 (IEEE Int. Conf. Computer Vision Workshops), 2015

[17] Kiss Attila, Szirányi Tamás: Szenzorhálózatok 3D összefűzése, KEPAF 2015 (Képfeldolgozók és Alakfelismerők Társaságának 10. országos konferenciája), 2015

[18] Kovács Levente: Single Image Visual Obstacle Avoidance for Low Power Mobile Sensing, LNCS 9386: Advanced Concepts for Intelligent Vision Systems. 16th International Conference, ACIVS , pp. 261-272, 2015

[19] Manno-Kovacs A, Sziranyi T: Orientation-selective building detection in aerial images, ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING 108: pp. 94-112. (2015)

[20] Varga D, Szirányi T: Robust real-time pedestrian detection in surveillance videos, Rev.version is under review: J Ambient Intelligence and Humanized Computing, 2016

[21] Mándli Anna (Kiss Attila, Szirányi Tamás supervisors): Intelligens Ágensek kapcsolata nagy méretű hálózatokon, project progress report, ELTE TTK, 2015.

[22] Varga D, Szirányi T: "Detecting pedestrians in surveillance videos based on Convolutional Neural Network and Motion", submitted to Eusipco, 2015