FRONT PAGE

PROJECT FINAL REPORT SUMMARY

Grant Agreement number: 315191

Project acronym: D-SENS

Project title: Depth Sensing Systems for People Safety

Funding Scheme: Research for SMEs FP7-SME-2012

Period covered: from 1.9.2012 to 31.8.2014

Name, title and organisation of the scientific representative of the project's coordinator1:
Esa Reilio, Managing Director RDnet Oy

Meri Helmi, Business Solutions Manager RDnet Oy

Tel: +358-40-502-6762

Tel: +358-40-485-7244

Fax:

E-mail: esa.reilio@rdnet.fi

E-mail: meri.helmi@rdnet.fi

Project website[2] address: http://www.d-sens.eu/

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

[2] The home page of the website should contain the generic European flag and the FP7 logo which are available in electronic format at the Europa website (logo of the European flag: http://europa.eu/abc/symbols/emblem/index_en.htm logo of the 7th FP: http://ec.europa.eu/research/fp7/index_en.cfm?pg=logos). The area of activity of the project should also be mentioned.

Executive summary

Due to recent technological developments in the visual depth sensing domain and driven by the available computational performance new perspectives are emerging for creating vision-based solutions which enhance human safety aspects. Human safety plays a central role in many fields of our daily life. In the *D-SenS* project (*D*epth *Sen*sing *S*ystems for People Safety) we brought state-of-the-art technology know-how and thorough understanding of end-user and stakeholder requirements together by engaging scientists and small-medium enterprises in joint technology development.

The main motivation of the *D-SenS* project was given by the fact that vision-based scene and object analysis in many application domains is still far from being reliable, thus remaining inapplicable for practical use cases. The main complexity of the underlying visual analysis task stems from the fact that image content is ambiguous by objective measures and it comprises a daunting variability. This prevailing ambiguity is one of the main motivating factors why 2D image information (RGB) and 3D depth data (D) are combined. These highly de-correlated information channels remove numerous geometric (e.g. scale) and photometric (e.g. shadows and illumination variations) ambiguities and therefore they are essential to meet the *D-SenS*'s research goals.

At the technology level the *D-SenS* project aimed at accomplishing practically relevant depth sensing vision systems around selected, previously unsolved problems which significantly impact human safety aspects. Given the heterogeneity of partner profiles and pursued goals across the project partners, one primary objective was to partition and link potentially marketable task-oriented ideas and available know-how in order to establish multiple actively networking teams within the project. These teams were essential to drive development from initial concepts towards prototypical solutions along a shortest path.

In the *D-SenS* project following research results have been accomplished: We addressed a set of novel visual monitoring applications for people safety, ranging from detecting and tracking humans in retail and public infrastructures, ambient intelligent and road environments to scenarios of left item and intrusion detection. Our solutions elaborated within the *D-SenS project* combine the "best of two worlds" (high-quality depth and intensity data, and modern vision algorithms) and enable the development of robust vision-based safety systems. Robustness in this context stems from the fact that the vision system operates on highly reliable data representing true scene geometry and remaining invariant with respect to photometric variations. In the project this increased robustness has been shown to yield meaningful high-level concepts (humans, left-items, intruders, scene model, motion patterns and activity) at a quality which seems to be adequate for practical use as demonstrated by the five use case prototypes tested and demoed in the application domains of smart buildings, assisted living and security.

An important vehicle towards efficiently reaching our task-oriented goals was to create a technology framework, the so-called *Common Framework* which shares common functionalities among vision system prototypes and enhances further development of system components and applications. The functionality of the *Common Framework* has been validated by the developed pilot applications in the selected applied domains and also joint demo sessions of a tutorial character were organised where consortium partners learned about the algorithmic mechanism, advantages and limitations of the created prototypes. We are confident that the project results and the *Common Framework* form a sustainable and easy-to-extend technology basis for commercial exploitation, long-term co-operations and additional project concepts even beyond the *D-SenS* project.

Project context and objectives

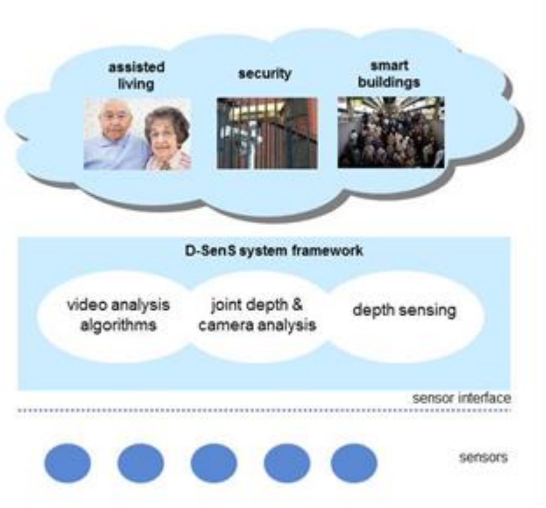

Figure 1: Illustration showing the depth sensing technology development concept of the *D-SenS* project. A visual analysis layer (*Common Framework*) is formed to hold all functional units in a structured manner (center). This layer is capable of processing arbitrary data (within a predefined set of sensor types) (bottom). At the same time it shares functionalities among the task-specific vision applications in the people safety domain (top).

All applied tasks targeted in the project involve new, emerging application scenarios; therefore prior to algorithmic development a thorough analysis has been carried out. The analysis involved aspects formulated in a use-case and marketing context: commercial potential, requirements, ease-of-use of deployment and operation, and probable limitations in a practical setting. At the same time the scientific and technology aspects were investigated: is the posed vision task feasible; are there probable limitations when applying the technology. The targeted pilot applications were decomposed into shareable functional units mapped to various levels of visual data processing within the *Common Framework*. The obtained three levels of abstraction (practical tasks, functional units and sensory data) are illustrated in Figure 1, created such that a maximum synergy between the pursued people safety applications can be accomplished.

The *Common Framework* library is an important result of the project which gathers and partitions camera and depth/intensity data processing algorithms into a unified software library. The jointly performed *Common Framework* design has identified various layers of information processing such as calibration, analysis and interfacing, and many recurring functional units are shared among the individual use case prototypes. Beside the established processing modules, the *Common Framework* specification includes also common data formats, interface definitions and software implementation practice guidelines. The entire code base was jointly managed and version controlled throughout the project, along with small sample data sets, exemplary processing chains illustrating a minimal solution of a problem for easy understanding and corresponding documentation.

Following focus areas were selected for use case definition, scientific investigation and implementation:

- People tracking and detection in smart surroundings,
- Intrusion detection within a spatial volume,
- People behaviour analysis,
- Object tracking and monitoring,
- 3D scene reconstruction and analysis from a moving platform,

- Traffic flow monitoring and
- Crowd management.

In light of the defined focus areas and use cases a review on the available technology (concepts, methodologies, and sensors) was performed and documented. The results of the review provided additional information for the conceptual design of the *Common Framework*, and lead to a set of detailed focus-area-specific requirements in terms of functionality, detection accuracy, system and hardware, context and operation environment, and interfaces as well as business-related requirements.

Gathering visual data in an early phase of the project was essential to assess the complexity of the given practical task. Data collection with depth sensors and stereo cameras was performed for guiding the development towards realistic scenarios. Depth sensing sensor setups (passive and active stereo) and acquisition software were distributed among the project partners, and a substantial amount (several hours) of data has been collected for each use case scenario. Data (time-varying data of depth, intensity and scene-specific priors) acquired in applied settings at the company partners beneficially complemented the laboratory data used at the scientific partners for elaborating solutions working under a large set of diverse conditions.

Five use case implementations were created to meet the commercial requirements of the participating SME's and demonstrating the *Common Framework's* flexibility and reusability. Next, we provide a concise overview on the current status of accomplished demonstrators:

People detection and tracking in smart surroundings: By employing depth data from two sensor types (MS Kinect/Kinect2 and a passive stereo camera setup) a real-time operational prototype has been elaborated performing human detection and tracking in indoor environments. Accurately quantifying indoor human demography in terms of number, location, motion patterns and dynamics provides information of great value in the retail, crowd management and safe infrastructures domains. Figure 2 shows some results of the human detection and tracking application.

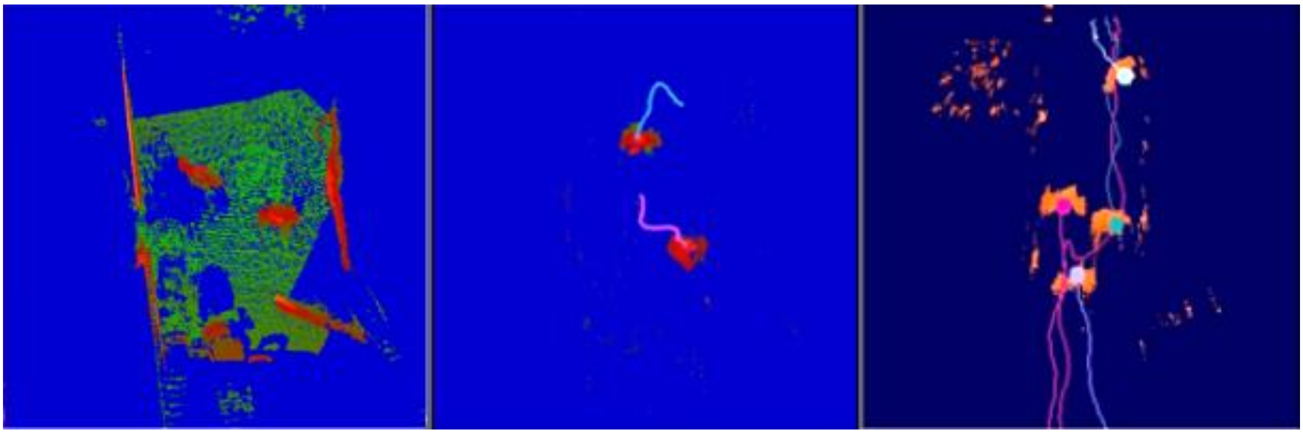

Figure 2: **People detection and tracking in smart surroundings**. Left: a top-view of the scene geometry represented by a point cloud (colour coding implies object height). Center and right: examples for tracking multiple interacting humans within this environment.

Detecting intrusion within predefined spatial volumes: Data representing time-varying spatial geometry is well suited for an analysis with respect to the presence or absence of arbitrary objects of sufficient size within predefined volume elements. Such a visual analytic functionality significantly impacts applied fields in safety (safe zones around machines) and security (intrusion detection for preventing theft). Figure 3 shows some result of the real-time intrusion detection prototype, capable of using two sensors observing the same intrusion volume, thus rendering the detection process substantially more robust with respect to occlusions or false alarms.

Figure 3: **Depth-based detection of intrusion within a predefined spatial volume element**. Left: Detection of an intrusion. Inset: point cloud representation of the scene along with a spherical volume element defining a monitored zone (box on the table). Center, Right: Sample detection results for our demonstrators running for several months.

Robust fall detection in indoor environments: Analysis of dynamic events and surrounding indoor scenes generates probabilistic alarms in the event of certain spatio-temporal patterns (fall, body pose) in this prototype. This application has a great practical impact in ambient assisted living scenarios (monitoring the elderly, children or chronically ill patients). Figure 4 displays a sample result for the awareness of human presence in the fall detection prototype.

Figure 4: **Fall detection prototype** segmenting spatial structures of a depth image detecting human shape pose variations (left and center) and its dynamics (height and velocity patterns - right) embedded into a statistical learning framework.

Reliable left-item detection in public spaces: Accurate detection of changes in specific crowded environments (subways, airports and other public spaces) is anticipated to enable the highly relevant security task: detecting left-items and their status (owner nearby or abandoned) in presence of clutter and occlusions. Figure 5 illustrates some results of the left-item detection prototype.

Figure 5: **Left-item detection** in scene based on combining depth and intensity information. Left: the demo scenario at UrbanMill used for demo session over many months. Center and right: Detected scene objects (blue – persons, green – stationary left item candidates, circle – vicinity for owner detection) in an indoor and outdoor environment.

Depth-aware road surveillance on a mobile platform: Future generation of vehicles are increasingly able to assess traffic and road conditions in front or around them by automated means. A specific high-resolution wide-baseline camera setup (Figure 6))has been developed and integrated into a road monitoring prototype targeting the detection of traffic participants up to large distances (<100 m), and assessing the road context from RGBD data, such as delineating rail tracks and thus assessing collision hazard prior to any incident.

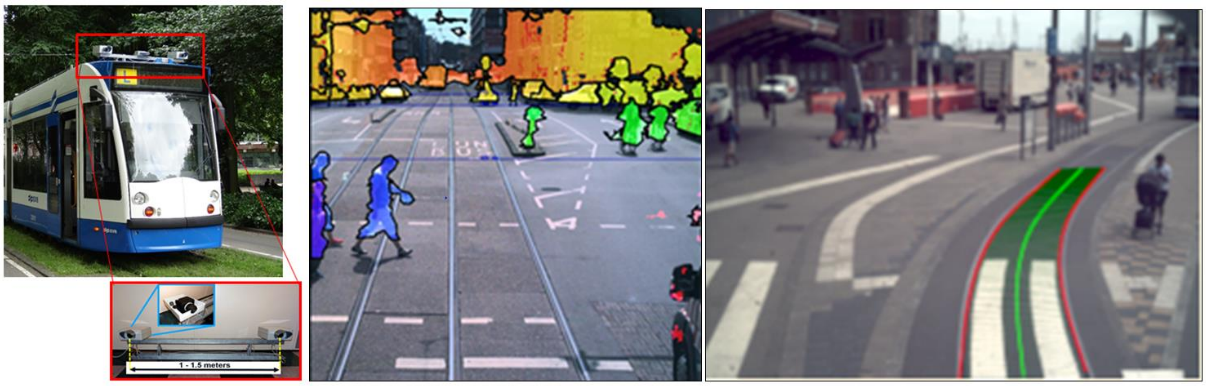

Figure 6: Left: Variable baseline stereo camera setup used on a **mobile tram platform for road surveillance.** Center: Detected objects of relevance such as pedestrians, vehicles and obstacles. Right: Sample result for detecting a curved rail track, representing the zone to be monitored for avoiding collisions.

Nearly all of these prototypes rely on privacy-preserving internal representations (point cloud, shape segments, geometry), therefore providing added support for a deployment in privacy-sensitive public (e.g. subway) or private (elderly home) scenarios.

Facilitating the take-up of results: Due to operating at large scale (geographic location, diverse scopes and background) within the project, the scientific and enterprise teams were faced with novel challenges: (i) how to integrate different research results into individual solutions which result in commercial benefits and (ii) how to establish a common understanding across viewpoints of technical and task-oriented nature. Partitioning on task-level (use cases) and on functionality-level (*Common Framework*) have proved to be a viable solution for creating actively engaged research-enterprise teams which share know-how with others. Organizationally, research was done *in situ* at the scientific partners, but as soon as a first framework became available, the development turned into an iterative process, bringing practical insights (verification, demonstration, end-user feedback) from the SME's back into the research process.

To achieve a collaborative integration of research and enterprise teams within *D-SenS* there was an important focus on Knowledge transfer and Knowledge management including the protection of IPR's. The consortium employed following means and channels to accomplish these goals:

- . Knowledge transfer and Training for SME's: regular telephone conferences between all consortium partners, scientific discussions between research institutes and research counselling and visits at the SME's have facilitated the joint design, analysis and modification of algorithms and frameworks. Additional 3-4 annual face-to-face workshops were held to present technical progress, to revise development plans and demonstrate achieved results.
- . Knowledge management: secure document and software configuration management frameworks (*Twiki*, *Git*) were employed to share knowledge and provide organizational flexibility so that we could support multiple use-case projects with shared functionalities and active engagement from all partners.
- . IPR protection: the consortium has agreed on legal principles governing the usage and right on previously existing and jointly developed know-how. The consortium has managed to achieve openness and flexibility posing no restrictions on the flow of ideas to support technical innovations. There were also emerging co-operations between SME's targeting joint business cases.
- . Dissemination and Exploitation: based on the accomplished use case results a *D-SenS*-Video has been prepared highlighting the technical innovations achieved in the developed technology prototypes and supporting commercialization efforts. The project website at www.d-sens.eu hosts demonstrations for accomplishment results. Dissemination at a scientific level (publications, tutorials, shared datasets) were also considered to be essential in order to generate valuable feedback from the scientific community, to produce further seeds of collaboration and to educate future staff. Accordingly, a number of scientific papers and presentation have been elaborated and presented at high-quality forums.