1 Executive Summary of the Proposal

Mass gatherings at large events and demonstrations in public urban spaces are an expression of freedom and openness in European societies. German and French citizens enjoy sport events, like marathon runs, festivals like the Cannstatter Wasn in Stuttgart, the Munich Oktoberfest, or Francofolies de la Rochelle, Feria de Bayonne, Fetes des Lumières in Lyon and many more. Festivals and politically motivated demonstrations alike involve threats for the safety and security of citizens. Therefore, organizers and public authorities are facing new challenges and are in need of new solutions. In this project, partners from academia, civil security forces, one large security industry partner and several small enterprises join forces to investigate and evaluate solution strategies, methods and concepts.

S²UCRE stands for Safety & Security of Urban Crowded Environments. The project addresses mass gatherings in **complex scenarios** and **widespread urban environments**. In such situations, the crowd is often inhomogeneous in terms of density and motion dynamics. In addition, pedestrians are distributed over large areas. The surveillance of the course of the event is challenging for Law Enforcement Agencies (LEAs) even without disturbances. An up-to-date overview is often missing. Fast reaction in case of accidents, offensive behaviour, or crowd densities that trap people and endanger lives (often referred to as “panic situations”) is difficult, pre-emptive measures seem impossible.

To overcome these problems, the S²UCRE project aims to interconnect **crowd monitoring based on video analytics** with simulation-based methods for **short-term prediction of crowd behaviour**. Two concrete events are selected as scenarios: the Hamburg Harbour Festival (1.5 million visitors) and the May 1st demonstrations in Paris (12000 attendees). S²UCRE furthermore intends to bridge the gap between the **macro level of crowd monitoring** and **observation of individuals or groups** when there is reasonable ground for suspicion. Carrying **out research** on **five interdependent support technologies for LEAs, event organizers, security staff and rescue teams,** S²UCRE will **improve the collaboration of stakeholders responsible for security**. Studies on **human factors** form an integral part of the technical work packages and thus go beyond accompanying research. A fundamental work package is dedicated to **legal and ethical aspects**. The S²UCRE demonstrator will integrate the support technologies, thus providing a proof of concept for predictive analytics tools. In particular, ways to guide crowds based on predictive simulation will be explored. The five technology areas are:

- Video-based crowd monitoring of distributed urban environments: crowd densities and crowd dynamics.
- Short-term prediction of crowd behaviour for fast and efficient evacuation.
- Semi-automated suspicious behaviour analysis for security applications.
- Detection & (geo-) localisation of perpetrators.
- Self-localisation of security and rescue team members and a communication platform for geo-registered information exchange between security and rescue personnel.

In contrast to current projects, S²UCRE will not develop these technologies individually but will focus the research on how to integrate a continuous workflow into a single system. This will combine the current security and safety situation with short-term simulations of the crowd that are used to prevent critical conditions.

S²UCRE envisions an **integrated surveillance and prediction system** that addresses safety and security and that is tested on concrete scenarios.

2 Objectives and Context of the Project

2.1 Overall Objective

Crowded places are potential targets for terrorists: they are locations with limited protective security measures and there is a potential for mass casualties. Yet, even without terrorist attacks, crowds are vulnerable. We have learned, from the bitter experience at Duisburg Love Parade in 2010 that, mistakes when guiding and managing the crowd may lead to deaths.

S²UCRE aims to carry out research towards new safety and security technology for monitoring, situation analysis, recognition and prediction of risky and security-critical situations at urban mass events. The overall idea is to design a combined safety and security monitoring system on a common platform. The following objectives are defined to devise an integrated monitoring system:

1. Real-time **video-monitoring** techniques shall be investigated and designed that allow people counting, density and speed estimates, so that input data for both, on-line simulation and direct risk assessment, are obtained. Algorithms for direct risk assessment are devised, notably detection of unusual (e.g. aggressive) behaviour or phenomena that indicate danger for the crowd, such as extreme densities or crowd quakes.
2. Online information on crowd behaviour, such as density, speed and choice of targets (e.g., beverage carts) shall be transmitted to a **simulation module** that produces **real time predictions of possible futures and threats**. The simulation module will also take into account psychological aspects provided by **psychological crowd studies** conducted within this project. Concrete sub-goals are: (1) Accelerate existing crowd models so that faster than real time code execution becomes possible. (2) Establish a framework for systematic sensitivity studies so that the likelihood of a certain outcome can be quantified (3) Conduct psychological studies to fill information gaps on typical crowd reactions (such as change of target). (4) Build a predictive building block for a demonstrator tool and validate it against evidence gathered at Hamburg Harbour Festival and the project's psychological studies.
3. Techniques for **detection, tracking and identification of individuals** in a crowd, who may represent a threat to others shall be investigated. Unusual behaviour detected on the macro level of crowd monitoring can indicate to aggressive behaviour of individuals or groups. The goal is to monitor individuals or groups of people with whom the LEAs need to establish contact to deescalate the situation. Scientific challenges include robust person/group tracking in dense crowds and multi-camera calibration including cameras from UAVs and body-worn cams.
4. **Communication technologies for** robust and efficient **interaction and information exchange** between crowd managers and security/rescue personnel shall be investigated. In particular, aspects of geo-referenced self-localisation of team members, as well as technologies for information hand-over regarding individuals that need to be approached shall be considered.
5. Finally, legal and ethical aspects of the monitoring in 3. and 4. shall be addressed. The goal is to define boundaries of usage and technical means to enforce these boundaries.

NOTE: Point 1 and 2 describe the research focus of the German consortium, point 3 and 4 the research focus of the French consortium. Point 5 is essential for both consortia. While each consortium could stand alone, there is benefit in an integrated system that provides functionalities for both applications, security and safety, and improves the collaboration of the stakeholders responsible for safety and security.

Safety and Security Goals Addressed by S²UCRE: The German consortium defines the following **safety goals** that will be investigated through the Hamburg Harbour Festival: a. Enable fast and efficient video network planning and deployment. b. Estimate crowd densities in real time and detect unusual crowd behaviour to assess risks. c. Predict crowd motion (short term) to support decisions and plan actions.

The Préfecture de Police de Paris has outlined the following **security goals**: a. Identify and track offenders in the crowd using a network of fixed and mobile cameras. b. Evaluate the potential threats associated with a person behaving strangely in the crowd. c. Identify vandalism or other groups that constitute a threat to the public. The security use cases will be investigated through the annual May 1st demonstration in Paris.

Figure 1, shows the overall concept in S²UCRE. In particular, it visualises cross-links between the system components (blocks) and their assignment to work packages (green, numbered circles). German partners are in charge of red components, French partners of the blue ones. German and French partners jointly work on bi-coloured components. In addition, the diagram shows the synergies and logical connections between components needed for safety applications (red) and security applications (blue). In addition, information exchange between all work packages will be ensured through binational project workshops.

The information flow and processing chain for safety scenarios is as follows: Algorithms for video-based crowd monitoring and density estimation in the urban area (3) extract data and give them to short-term simulation for risk assessment (5). Here the geo-registered knowledge about escape paths (4) and human factors analysis (2) is used to simulate evacuation scenarios. The results of (3) and (5) are displayed in the surveillance and prediction system (7). Human operators assess them. In case of a critical situation, the human operator can send dedicated messages and action plans to security officers and rescue teams in the crowd (8), or the crowd itself (9).

For the security use cases, the information flow is as follows: Mobile cameras (6), and possibly drones (to be investigated) gather data that is displayed in the surveillance and prediction system (7) and observed by human operators. If the operators detect suspicious behaviour, they activate the monitoring application in the surveillance and prediction system (7). The potential offender is then detected and tracked by the system and tracked (3).

Information on offenders that are considered dangerous can be handed over to the mobile security forces so that the right person can be approached (8). In particular, this includes the geo-location of the offenders and their appearance as shown in high-resolution images.

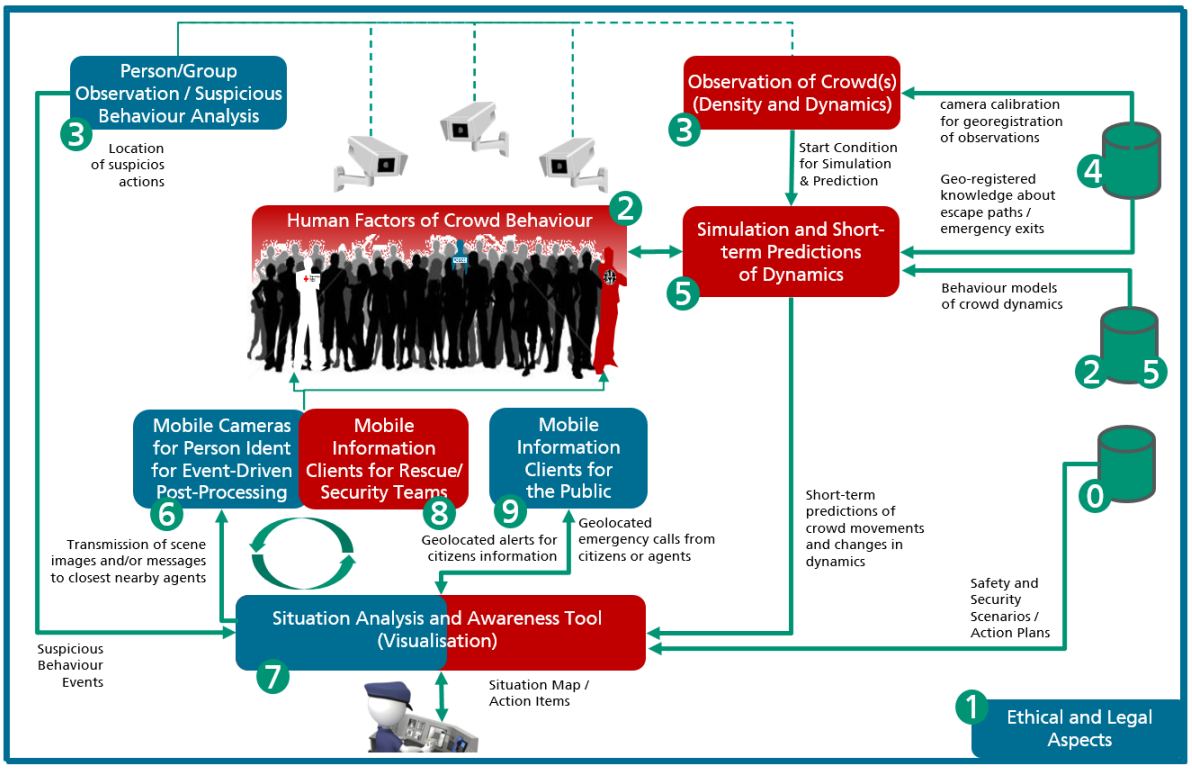

Figure 1: Overview of the Concept and Basic System Components

As shown in Figure 1, S²UCRE covers the complete process chain, from observation of crowded urban scenes, to risk assessment and situation awareness using geo-located observations, to short-term predictive simulations, and to visualisation for decision makers. In parallel, human operators who can trigger measures in case of reasonable suspicion use the sensor (camera) network for surveillance. Thus, S²UCRE unifies two sets of system requirements and concepts, safety and security, on one platform. Ethical and legal aspects (1) are evaluated in a separate work package that will focus at the security scenario.

2.2 Context and Relevance of the Proposal to the Call

S²UCRE addresses the topic of protection of citizens in publicly accessible urban spaces: Assistive technologies to evaluate and predict the security situation of large-scale public events are investigated. S²UCRE combines crowd monitoring based on video analytics with pedestrian simulation to provide short-term prediction of crowd densities and flows at public events, such as the Hamburg Harbour Festival and the annual May 1st demonstration in Paris. The goal is to predict danger and to match predictions with evacuation scenarios worked out in advance to better guide large crowds in emergencies. Unusual behaviour in a crowd, such as aggression by individuals or groups, can lead to incidents that are relevant for both, security and safety. S²UCRE therefore combines safety- and security-related technologies in an integrated surveillance and prediction system.

Despite some success stories such as traffic monitoring and intrusion detection, video analytics systems remain brittle and function only under restrictive conditions that are not compliant with most real-time operational conditions. Most unresolved issues regard detection, tracking and identification. S²UCRE will significantly extent VOIE [0], a demonstrator for processing large volumes of video data for criminal investigations that is currently developed under the guidance of partner Morpho in response to objectives set by COFIS, “Comité de Filière des Industries de Sécurité. COFIS was created by the French prime minister in 2013 to improve access of operational governmental stakeholders to key security technologies, including video protection systems, and to strengthen the French national security industry.

Guiding large crowds requires close cooperation between all stakeholders responsible for public safety and security: LEAs, crowd managers, organizers, security and emergency teams. S²UCRE will support this cooperation by contributing new communication technologies for information exchange between control rooms and personnel on-site. The S²UCRE demonstrator will also include means to directly instruct and guide crowd members.

Researchers from the area of data protection and ethics collaborate in the consortium in order to establish a development process that follows the principles of privacy by design. From the beginning, threats and risks for visitors, that may be induced by S²UCRE’s new technologies shall be identified, analysed and means to mitigate them shall be devised.

2.3 State of the Art and Patents Filed

Five core technologies form the backbone of the S²UCRE demonstrator:

Video analytics for crowd monitoring (safety applications), in particular estimation of densities, people counting and changes in motion behaviour are researched worldwide. An excellent overview of state of the art methods can be found in [37], [38]. However, many approaches have been developed as “single-camera” systems in which density, counting, as well as motion behaviour is estimated for each video stream (camera) independently. This is a major drawback for intuitive interpretation and assessment of situations, which can only be interpreted correctly when one takes into account information from multiple cameras, overlapping cameras or even cameras that are sparsely distributed over a large environment.

Within the past decade **simulation of crowds for evacuation planning and risk assessment** has been investigated intensively by projects such as SinoVE [25], REPKA [23], BASIGO [35], MulitkOSi [24] (German), PEDIGREE [33], CBDif-Fr [34] (French). Controlled

experiments have been conducted to allow comparison to empirical data. The previously dominant cellular automata and social force type models have been complemented by models that more closely mimic human stepping [2], [3] and, to a limited extent, include human decision processes and social behaviours [5]. In addition, a number of simulation frameworks have become available, among them HM's VADERE research platform, which will be accessible open source in 2016. The number of companies that offer planning reflects the progress in the area and consulting based on proprietary software. One of them is S²UCRE partner accu:rate whose unique selling proposition is the use of cutting-edge technology. On the other hand, there are very few results on short-term predictions. Several issues have hampered progress: Numerical problems in some models lead to instabilities [1]. High computational costs prohibit faster than real time simulation in more complex scenarios and cannot always be resolved by parallelization. Moreover, crucially, but little discussed, changes in each of the many model parameters may strongly effect simulation outcomes [5], [6]. Thus, predictive power may be lost.

Fast and efficient pose estimation/calibration of stationary and mobile cameras in particular for highly distributed camera networks with non-overlapping field of views, is an active field of research. Based on recent achievements as proposed by [7], [8], [9] good results can be obtained for specific applications (e.g. overlapping fields of vision). These approaches, however, fail for scenes typical for urban areas: large distances, difficult interest point matching due to homogenous areas or repetitive structures, significant amounts of dynamic "blobs" (groups of people) without a clear background. To overcome these issues ways to allow highly accurate camera pose estimation (geo-registration) must be found. Preliminary work on Visual SLAM based calibration has been carried out by Fraunhofer IOSB in the BMBF-project Muskat [26], with promising results on calibration of "mid-scale" camera networks with non-overlapping fields of vision. In S²UCRE these approaches must be extended to large scale distributed urban environments. Estimating the pose of a dynamic/moving camera is also imperative in order to relate elements to a larger context, which is available only in the wide field of vision of a static camera. This problem is notoriously difficult, a variant of it having been studied in robotics [10].

In addition to research on safety applications, S²UCRE also addresses security-related **video analytics for re-identification of individuals (offenders).** All monitoring technologies will be based on an integrated platform, consisting of stationary cameras (as used by the safety system) as well as mobile cameras (body-worn cameras, hand-held cameras, UAVs). The state of the art on the re-identification problem deals mainly with pedestrian re-identification or face re-identification. Pedestrian re-identification is usually studied on video-surveillance images with small definition, while face re-identification usually needs high resolutions images to ensure minimal accuracy. Liao et al. [17] compare the descriptors used in recent research. Some methods focus on the construction of accurate colour histograms [18], others use local descriptors [19] or filters [20]. Further studies propose robust metric learning methods to measure the similarities between the objects [21]. Recently, a new kind of methods based on deep convolutional networks exceeds previous results in several fields of computer vision. These methods are now used for re-identification and since 2015 some methods have begun to achieve state of the art results [14]. Interesting results were reported in the TRECVID Challenge [22] for the Instance Search Task. However, the algorithms cannot be directly applied to the video-surveillance problems, as the definition of the images is better in the TRECVID Challenge than in video-surveillance

The state of the art on **Communication platforms for real-time information exchange between command and control officers and mobile security and rescue teams** is as follows: Several mobile solutions, based on real-time location have been studied in European projects: Public Warning and Information Solution for crisis management (FP7: iSAR+), and emergency call solutions (FP7: SOTERIA and H2020: Nexes). These developments offer text, audio and picture transmission for population alert, as well as compliance to the future

emergency call standard. In addition, photos and video transmission from citizens to authorities and secure chatting with the Public Safety Access Points is offered.

2.4 Scientific and/or Technical Working Objectives of the Proposal, Including Envisioned Innovations and Project Feasibility

While camera-based systems have been used as distributed sensor networks in many projects showing, in principle, the applicability of video analytics to support security personnel who monitor crowds, most approaches share several main weaknesses: First, the systems were designed for monitoring of crowds in a “closed” area, with fixed known entry-exit points. Thus, they count people and measure densities under more “controlled” conditions. Second, real-time observations and, in particular, short-term predictions have not been investigated in sufficient detail. Third, some projects focused on security applications, while others focused on safety aspects of crowds. Synergies, cross-linked aspects, and integrated platforms have not yet been taken into account. S²UCRE aims at closing these technology gaps:

Video Analytics for Crowd Monitoring (Safety Applications). S²UCRE addresses the issue of multi-camera crowd monitoring in the following steps. Video analytics will first extract basic features on crowd density and motion behaviour for each video separately. However, in the next steps, all observations are geo-registered on a global map of the urban environment. Using sensor data fusion approaches, all observations are merged to generate a joint crowd density and motion behaviour map for further analysis and evaluation. Preliminary work towards these concepts has been presented by Monari et al. [36]. Still missing are approaches to get a highly accurate camera calibration in respect to a map of a given location.

Simulation of Crowds for Evacuation Planning and Risk Assessment. Predictive crowd simulations must be very fast: S²UCRE will build on HM’s well-tested research platform VADERE to investigate methods for code acceleration that go beyond parallelization and computation on GPUs. In particular, techniques from computational fluid dynamics, such as adaptive mesh generation, will be transferred to crowd simulations. Predictive crowd simulations must also be accurate: Studies of human behaviour will be considered to complement the input data available from video analytics, in particular with respect to likely decisions on where to walk next. Furthermore, methods to conduct sensitivity studies will be investigated and implemented so that the system can identity input parameters where little changes of measurement errors have great impact on the simulation outcome. Sensitivity studies will be conducted systematically on the main input parameters for predictive simulations: crowd densities, speeds, preferred targets. Thus, the user will be able to judge the reliability of the predictive simulations. Finally, a control cycle will be devised that takes data from video analytics as input for predictive simulations. Simulation outcomes and measurements will be compared to recalibrate the simulation system.

Fast and Efficient Pose Estimation of Stationary and Mobile Cameras for Geolocalisation. Camera pose estimation allows knowing the current position of fixed cameras and, additionally, the alignment of pan-tilt-zoom cameras. In S²UCRE, new Visual-SLAM based methods will be proposed to allow highly accurate camera pose estimation (geo-registration). In crowded scenes, this problem is more difficult than for augmented reality because of occlusions and since the perspective change of the static cameras is significant. Previous results indicate that exploiting additional sensors injects helpful information for the pose estimation [11], [12]. Based on previous work performed on GPS filtering in urban environments [14] and on inertial-vision data fusion for navigation [15], S²UCRE intends to use additional data sources in order to constrain and solve the pose estimation problem for wearable and UAV cameras with focus on robustness. In addition, previous results obtained on road detection from on-board car cameras [16] will be adapted and applied in order to perform ground plane detection from dynamic cameras.

Video Analytics for Re-identification of Individuals (Offenders). One of the most challenging problems in mobile vision applications and for mobile camera platforms remains, geo-registration of observations, tracks, position of individuals. S²UCRE will propose an accurate