Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the “Confirmation of Authorship” file.

As corresponding author I, Federico Becattini, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature____ Date 29th of June, 2022

List any pre-prints:

Relevant Conference publication(s) (submitted, accepted, or published):

Justification for re-publication:

Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- We present eSERVANT, an IT infrastructure for smart management of event venues
- eSERVANT provides vision-based dynamic routing preventing crowds indoor and outdoor
- Through social media data we recommend similar users attending the event
- We evaluated user satisfaction of the eSERVANT app based on a 6 months sperimentation

ELSEVIER

Pattern Recognition Letters
journal homepage: www.elsevier.com

Events in Crowded Places: a Smart Service Management

Federico Becattini[a,**], Andrea Ferracani[a], Giuseppe Becchi[a], Alberto Del Bimbo[a]

[a] *University of Florence*

ABSTRACT

In this work we present eSERVANT (eventS in crowdEd places: a smaRt serVice mANagemenT), an ICT hardware and software infrastructure that allows the management of services tied to big events such as concerts, sport matches or fairs that take place in dedicated facilities. Beyond such events, there is a network of shared interests and needs, for which a smart environment can offer new opportunities as regard to location-based, mobility, security, profiling and social networking services.

1. Introduction

The IT-based functional control of infrastructures in the context of events regarding sport, entertainment, music or leisure is nowadays one of the main opportunities offered by new technologies for the management of medium and big facilities. People access and ticketing, administrative, accounting and commercial management, maintenance and connectivity are a big part of the services that can benefit from them. Nevertheless, commonly, these services are essentially addressed to facilities' operators, and only marginally to end users (i.e. the public). In this regard, they are almost completely ”disconnected” from the events themselves, the place and the surrounding area these are hosted in, and they are unable to get an adequate customization of services. The goal of eSERVANT has been that to develop a highly modular, flexible and configurable ICT infrastructure that would allow the exploitation of some of the latest technological innovations in user localisation, tracking and profiling in order to improve the quality of contextual and targeted services with respect to the facility, its urban context and the events' participants. eSERVANT provides an adaptable and configurable hardware and software infrastructure based on three core modules: 1) a video analysis module for the detection of flows of people, crowding and anomalies; 2) a user localization and dynamic routing module which enhances the fruition of services in indoor and outdoor environments; 3) a user profiling module which promotes contextual social networking through recommendation. The three modules can be orchestrated in order to ingest, analyse and manage data coming from several sources. This data and the results of the analyses are exploited to provide services and to give appropriate feedback to the end users, that, at this end, can exploit a mobile application also part of the eSERVANT infrastructure.

2. Related work

Video surveillance for large events. Safety is understandably an issue of primary importance when an event takes place. Threats and dangers can stem from many independent factors. The most critical situations take place when the threat is intentional, such as acts of terrorism. Nonetheless, more common scenarios can turn into a hazard when large crowds are formed in closed spaces without appropriate exit routes. These phenomena are often neglected, but can rapidly turn into a stampede when the crowd starts to perceive a sense of risk. Anomalous crowds as much as congested hallways and exit routes can therefore pose a threat to safety for all the people in the event site. At the same time a system capable of routing people across the building hosting the event could lower the load it has to sustain and prevent harmful situations. Such hazards have been largely amplified by the diffusion of the COVID pandemic, for which it is essential to avoid large crowdings, especially indoors [1]. This, in addition to the always increasing frequency of events gathering thousands of participants has led to a growing interest in the study of crowds through video analysis. The perks of relying on surveillance cameras lay in the broad coverage of the site that can be reached without being invasive for people attending the event. Whereas nowadays

[**] Corresponding author
e-mail: federico.becattini@unifi.it (Federico Becattini), andrea.ferracani@unifi.it (Andrea Ferracani), giuseppe.becchi@unifi.it (Giuseppe Becchi), alberto.delbimbo@unifi.it (Alberto Del Bimbo)

most buildings are equipped with surveillance systems, they are mostly manned by an operator which has to constantly monitor a large amount of video streams simultaneously. A first step to aid this process has been done with software frameworks used in combination with expert operators [2]. In the presence of big sport events as well as cultural gatherings, the need for a video surveillance system increases. During the soccer European tournament in 2008 Cisco provided a vast distributed system for surveillance [3] to monitor the large-screen monitors installed across the city of Zürich. For EXPO 2015 in Milan, a surveillance infrastructure has been developed, including an integrated architecture in a single operative center. The structure had to aggregate information streams from all the devices installed in the site and its surroundings such as cameras, smoke detectors, perimeter control and restricted area access gates. All the data was accessible in real time to the workers of the operative center. Similarly, for the Olympic games in Rio 2016, Dahua Technologies provided 1823 high resolution PTZ cameras, installed in strategical points and aggregated in a control center. Of particular interest were the artificial intelligence capabilities of the cameras which automatically signaled lost or abandoned luggage and unauthorized accesses. Apart from the hardware and software infrastructure, a behavioral study based on surveillance streams finds a great interest for safety but also for the organizational aspect of the events and the sites themselves since services offered to the visitors can be optimized for a better organization of future events [4]. Safety-wise, tragic episodes [5] can be recently found in the Love Parade disaster (Germany, 2010 [6]) and in the Phnom Penh stampede (Cambodia, 2010 [7]) where uncontrollable crowds caused up to hundreds of casualties. At the same time, in a controlled environment, people density estimation and analysis can help to establish a certain degree of comfort. In literature, observed areas are classified into five different categories: *free*, *restricted*, *dense*, *very dense* and *jammed flow* [8]. Methods that deal with analyzing crowds from video flows can be divided into methods that track single individuals [9, 10] and methods that directly estimate density [11]. The former focus on studying trajectories, while the latter are usually more scalable and are based on texture analysis and motion estimation. An important aspect in analyzing crowds is to take into account also the surrounding environment. The presence of intersections for instance, has an impact on people dynamics when groups split or merge across different paths [12, 13, 14]. An additional issue to take into account for ensuring safety is the ability to recognize anomalies such as fire outbreaks, abandoned luggage, unusual crowds or accidents [15]. Many techniques have been proposed in literature, based for instance on optical flow [16, 17], sparse representations [18] or deep learning [19, 20]. On the other hand, video surveillance streams can be used as a mean to offer services to visitors. People can in fact be engaged both from a smart infrastructure by being informed on the state of the ongoing event and by asking to the visitor itself to cooperate by handing in pieces of information [21]. User's participation can even be uncooperative and information about the event can be integrated from secondary sources such as social media [22].

User profiling. User profiling is the process of collecting, inferring and organising user profile information such as interests, preferences and behaviours. Profile creation is a relevant topic in Social Networks because it helps in providing personalised services, to filter relevant information and to perform targeted recommendation [23, 24, 25, 26]. A user profile model can be defined as a set of information that describes a user and consists of demographic information such as the user's name, age, country, level of education but also user's preferences and interests [27]. User profiling can be an effective technique to understand user's needs and to predict and condition his future behaviours. Alaoui *et al.* [28] have noted that this effectiveness depends on three major factors, i.e. digital traces' collection and management, computation of similarities and prediction through machine learning. In fact it is demonstrated that similarities in user's profile modules result in a similar behavioural model that can be exploited to elicit specific actions [29]. A survey on user profiling exploiting data from Social Networks has been conducted by Mezghani *et al.* [30]. The authors show how a tag-based modelling approach, based on the annotation of user profiles with 'basic demographic information', 'knowledge', 'interests', 'history', 'preferences', can be fruitfully used for recommendation purposes. Another important point in user profiling that has to be highlighted is the distinction between 'static profiles' and 'dynamic profiles' [31]. In fact, collecting user attributes that do not change over time is significantly different from analysing the temporal behaviour and the changes in interests and preferences of the user. Liang *et al.* [32] propose a Dynamic User and Word Embedding model (DUWE), and a Streaming Keyword Diversification Model (SKDM) to dynamically track the semantic representations of users and words over time, modelling their embeddings in the same space in order to measure their similarities. Information used to build user profiles can be gathered explicitly or implicitly [25]. Explicit information is declared by the user, e.g. on the personal data fields of the public profile page of a social network or submitted directly by filling out a form. On the contrary, implicit information is the result of the processing of undeclared but inferred data, e.g. the categorisation of the articles read or of the resources the user liked on a social network. Obtaining this data is part of the process called data collection, a preparatory step to data pre-processing, feature extraction, analysis, prediction and recommendation. Nowadays, despite the more and more stringent privacy concern, a very common way to collect user data is through the exploitation of Social Networks APIs. Chen *et al.* [33] uses a Social Tie Factor Graph (STFG) model to estimate a Twitter user's reference city area based on the user's followers network and personal demographic data. Relationships between users and locations are modelled as nodes; attributes and correlations are modelled as factors. Recently, proposed methods for feature extraction on Social Networks include behaviour, network and content analysis. Common approaches comprise profiling through interest, location and network analysis, establishing a strength correlation between users [33], keyword with their embeddings [32] and multimedia content-based feature extraction. Fernadi *et al.* [34] proposes a deep learning approach that extracts and fuses information across different modalities (status updates, page 'likes', images, relations) for inferring age, gender and personality traits of social media users. Users' annotations and tagging from Wikipedia have been used [35] to model user preferences and suggest videos, users and other resources on a Social Network. Content-based profiling, combined with collaborative filtering, is exploited by the same authors for improving recommendations systems through an hybrid approach in [23, 36, 37].

3. Monitoring Motion Flows for Dynamic Indoor Routing

eSERVANT features a video analysis module exploiting cameras for safety related functions, including dynamic indoor routing. The module allows the monitoring and management of flows of people within the facility where the event takes place and in the surrounding areas. The algorithm used to detect people was developed on SSD: Single Shot MultiBox Detector [38], a method for object detection in images based on a Convolutional Neural Network (CNN). The model has been readapted to be capable of efficiently detecting the single class of 'person' and to be able to work on multiple video streams rather than on an individual one. To efficiently process multiple streams together, a processing server queues frames from multiple cameras in parallel and feeds batches of frames to the model as soon as a sufficient number of images is available. The use of a GPU for evaluating the model makes it possible to obtain real time results, which are compliant with the needs of a surveillance application. A backend graphical interface, fully integrated with the system for the management and the orchestration of all the modules, allows the configuration of the cameras in the indoor/outdoor environment. For the purpose of detecting crowds, a ”walkable” zone is defined for each camera. Based on this area, the system provides an estimate on how filled the area under observation is and whether more people could occupy it safely. Each camera is connected to the processing server through a private Local Area Network and the video stream is always processed by the model and immediately discarded, for privacy reasons. Vision services are configured through a web backend exposing an administrative panel to add and configure cameras. Each camera is geo-tagged and linked to an area of observation, typically a room or a courtyard in the facility. When the camera is installed, a zone of interest can be defined through GPS coordinates and a radius. This area is used by the indoor routing module to define which areas of the building are crowded and should be avoided when moving in it.

Through the admin panel, operators can specify the following parameters tied to each camera: the camera identifier; a description to easily locate the camera; the local URL of the video stream read by the processing server; the floor of the building in which the camera is installed; GPS coordinates of the area of interest observed by the camera; radius in meters of influence of the observed area; a flag that can be toggled to activate or deactivate the camera. After adding a camera, an initial configuration step is required concerning the area of observation and the possible exits. When configuring the area of observation, a frame captured by the camera is shown and the operator can draw a polygon on it defining the walkable area. The walkable area is used to establish the occupancy of the room based on the amount of people detected in the scene. Similarly, exits are clicked on the frame and are treated as basins of attraction towards where people can be directed.

eSERVANT provides routing towards Points of Interest (PoIs) through its mobile app module. Whereas outdoor routing is nowadays a pervasive and consolidated technology, indoor routing is still rarely adopted for events in large venues. The eSERVANT smart routing algorithm adapts its routes in real time based on data coming from the sensors installed throughout the building, offering the best route avoiding crowds and queues and hence reducing potentially dangerous situations. This is of particular importance after the outbreak of COVID, since it can reduce unwanted gatherings and crowds indoor. At the same time, in presence of danger (extremely crowded areas, fire outbreaks, etc.) the algorithm can guide the visitor outside the building through the shortest safe path. The indoor routing module can also be configured by the user in order to find paths that fit with their needs, such as barrier free routes for disabled people. We built our indoor routing module by representing buildings as 3D graphs where nodes are rooms and edges are connections between rooms, such as doors and stairs. Each node has a set of 3D coordinates given by longitude, latitude and floor of the room inside the building. The proposed routing solution is integrated in OpenStreetMap[1] and built on INDRZ[2], a Django open source project for indoor navigation.

3.1. Administrative area

The administrative area of the routing module allows an operator to configure the algorithm and map venues and buildings. Buildings can be grouped in aggregated entities, referred to as campuses, in order to map venues for distributed events. Going in depth, buildings are hierarchically represented as a collection of floors, spaces (such as rooms or hallways) and Points of Interest (PoIs). This organization allows a full customization of the venue, obtaining dynamic maps which can include a whole campus or just a few rooms based on specific events. The creation of any of the aforementioned entities is performed through the admin panel. Every entity is geo-localized and stored into a PostgreSQL/PostGIS database, which offers specific models for geo-spatial applications. To facilitate the configuration of a building, an OpenStreetMap **.osm* file can be uploaded to be processed and stored automatically into the database. The creation of these files is made with JOSM[3], an external tool provided by OpenStreetMap which allows the user to superimpose building blueprints on maps to facilitate annotations. The annotation process consists in two phases: (i) manually identifying rooms through a polygon drawn on the map and labeling them with a description and a floor level (Fig. 1); (ii) outlining routes between rooms with JOSM's drawing tools. The result is a representation of the building as a graph where rooms are nodes and routes edges. Properties can be added to edges to specify its category (such as corridor, stairs or lift) and whether the route is barrier-free or not.

[1] https://www.openstreetmap.org/

[2] https://github.com/indrz/indrz

[3] https://josm.openstreetmap.de

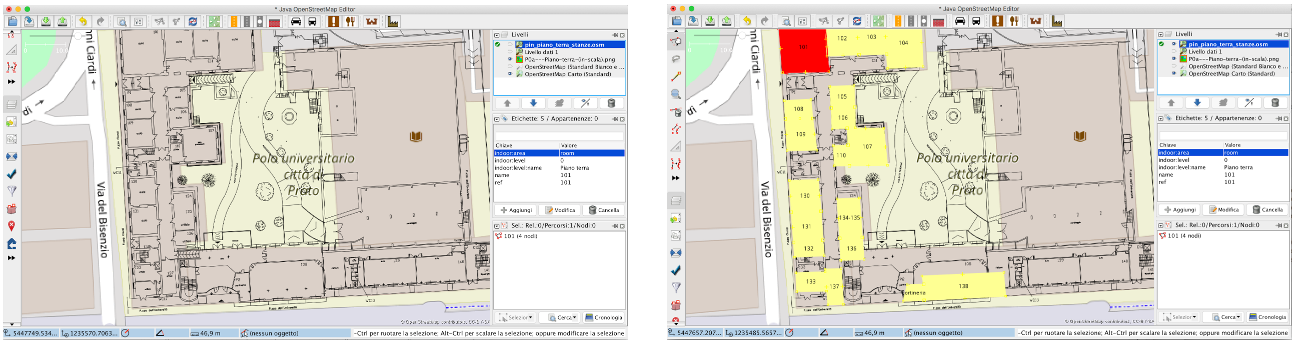

Fig. 1. Building annotation in JOSM. *Left*: blueprints can be superimposed on OpenStreetMap maps. *Right*: rooms can be labeled through a polygon.

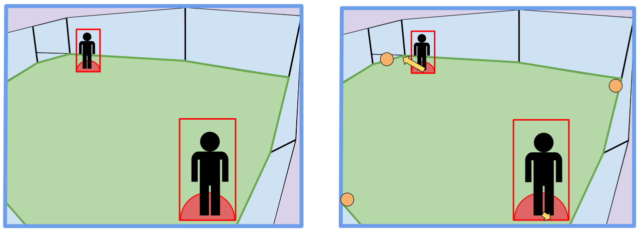

Fig. 2. *Left*: occupancy of detected people is estimated as a semi-circle on the walkable area (green). *Right*: motion-based association with exits.

3.2. Vision based indoor routing

The indoor routing algorithm developed for eSERVANT is tightly coupled with the vision module. Thanks to the cameras installed across the building, information about people flows and crowds are gathered, weighing the routes that users may undertake. Similarly to what happens in outdoor routing algorithms, in the presence of traffic jams, a penalty is given to a route that crosses the interested area when a crowd or an unusual flow is detected. To this end, the vision module registers two types of data: inbound/outbound flows and room occupancies. For each camera, the walkable area and the possible exits are defined as described in Sec. 3. Since cameras are calibrated, the overall room surface can be computed as the area of the rectified walkable polygon. For each detected person in the scene with its bounding box lying in the walkable area, an occupancy is estimated as the area of a semi-circle on the ground, having as diameter the lower segment of the bounding box. The overall room occupancy is then defined as the fraction of walkable area occupied by detected people.

To estimate inbound and outbound flows, we compute the moving direction of each observed person by tracking its position in subsequent frames with a tracking by detection algorithm [9]. We then identify the closest exit to its heading direction to understand which exit will be occupied. This allows us to estimate the flow of each exit (Fig. 2). Room occupancies and exit flows give us two indicators to understand if certain parts of the building should be avoided when routing people. The dynamic routing module in fact removes edges and nodes in the building graph when the vision module detects congestions, thus proposing different routes (Fig. 3).

4. User Profiling and Recommendation

The eSERVANT platform provides the end users with a mobile application which features several targeted services for receiving content updates, routing, notifications on the status of available facilities, recommendations. Standard multimedia information sharing, as well as innovative social networking strategies, are made available to users on this mobile app. The app services are based on profiling on social networks and contextual location understanding through digital traces analysis. The profiling module is mainly exploited for user recommendation and the creation of micro-communities. User data exploited by the profiling module is obtained from three different sources: social network login (Facebook, Twitter, Instagram); collaborative filtering on the application itself; location sensors. The users profiling module estimates a user to user affinity correlation between all the users of the application. The affinity estimate is exploited to increase the use of the social networking mechanisms proposed by the application through targeted recommendations of users to other users. The main social networking features provided by the app are: (i) visualization of friends and people present to events, with the detail of the estimated similarity to those users; (ii) creation and management of groups for *pre* or *post* event activities (e.g. celebrations, gatherings and other contextual events); (iii) creation and management of groups for car sharing.

4.1. Data collection

User data is collected from three main social networks: Facebook, Instagram and Twitter. The extracted data is exploited for the creation of the digital profiles of the users of the eSERVANT platform. Profiling is carried out with the purpose of implementing user recommendation algorithms used for the detection of micro-communities, the creation of groups for car sharing and activities contextual to events. These data include both static information (such as personal data, gender, age, address of residence) and dynamic data: status updates, user activities ('like', 'following' and 'followers', geo-locations, multimedia materials), which represent user's interests over time. These streams are combined with data deriving from physical analytics (i.e. data from location sensors and wi-fi networks) and from collaborative filtering on the mobile application. Twitter data is collected through Application-only authenticated requests. Facebook and Instagram use the OAuth open protocol for access delegation.

4.2. User profiling for social networking

The user-to-user similarity exploited by the recommendation systems is computed on four main macro-areas. These correspond to profiling domains that cover different aspects of ”sociality”: 1) “history”: general demographic information such as age, range of age, sex, friends; 2) “preferences”: favorite movies, books, TV series, categories of liked Facebook pages; 3) “activities” such as events or categories of events promoted by the users or in which they participated in, but also inferred activities derived from status messages or posted multimedia materials; 4) “localizations”: geo-referenced coordinates attributable to a user, useful to estimate a geographic area of interest. Two different approaches are exploited to recommending users in the scenarios implemented on the mobile app: 1) recommendation of users for pre-post event activities: a similarity based on user preferences is used; 2) user recommendation for the creation of car sharing groups: a similarity based on geolocations to estimate an overlapping area of interest is used.

4.2.1. Cold-start problem and segmentation

Recommendation is provided on different levels on the eSERVANT platform. In fact, in order to solve the so-called ‘cold start problem’ we use two strategies: 1) we compute user similarities on the fly, exploiting basic explicit data from social networks such as demographic information only between the users that are going to participate in the same events; 2) segmentation, i.e. we distribute users in groups. This is done because the process of computing user-to-user similarity, exploited for recommendation, can be ‘heavy’ and it can grow exponentially with the number of app users. In this way suggestions of similar users range from generic recommendations (e.g. recommendation of popular and most active users), to semi-personalized (in the case of users belonging to the same segments), to personalized (based on profiling and personal data modelling).

4.2.2. Analysing preferences

The algorithm for estimating the similarity between users on the preferences area is based on a user-to-user distance computed with respect to the level of interest of each user in a taxonomy of categories that actually represents most of the knowledge domains. This taxonomy is composed of 1542 categories obtained from the Facebook public API and used by Facebook to categorise all the resources of its knowledge-base (e.g. users, businesses, pages, films, TV series). A recommendation algorithm usually works through a series of votes or user ratings on particular items of interest (e.g. the product ratings on Amazon). The predictive model is built based on the basis of the distribution of votes expressed by each user. On the eSERVANT platform an implicit approach is used which exploits social profiling for inferring the user’s level of interest on all the categories through which the profile is represented, that is, a vote is estimated where explicit data is missing. Segments are used in order to pre-filter the set of recommendable users.

4.2.3. Recommendation using preferences

eSERVANT uses a user-item matrix (based on the 1542 Facebook categories) to define a user-to-user similarity. There are three main issues: 1) data sparsity: there might not be sufficient categories in common among users to estimate a similarity; 2) the need to compute a vote on each of the categories for each user; 3) defining a similarity that takes into account the weight of each category when calculating the affinity. Categories in the taxonomy are considered as user preferences. A categorisation system has been implemented that acts on the various types of multimedia data that can be extracted from social profiles. These are assigned to a category on an explicit or implicit basis dependently from the fact that a category is publicly assigned by a social network or not, e.g. a like on a page on Facebook is categorised as the category of the page. Different and configurable weights have been defined that impact the category score computed by the system for each user. A temporal decay function is also used, essential in assessing the importance of the data since user preferences can change over time. Next steps are: 1) normalization of category scores for comparing users; 2) sparsity reduction of the set of categories to have comparable data between users; 3) calculation of the probability of a category in the user segment; 4) calculation of the overall category probability; 5) calculation of user-to-user similarity based on the aforementioned probabilities and in-app recommendation of more similar users. The *weighted Pearson similarity* algorithm has been exploit to compute a linear and weighted correlation in the distribution pattern of the category values in the user-to-user pair. An *ad-hoc* algorithm was instead designed for 1) the definition of the score of each user action with respect to the type of feed (taking into account the weight of the feed and a time decay function); 2) sparsity reduction: carried out through an algorithm that propagates and re-evaluates the user categories according to the structure of the category graph. In the context of eSERVANT, the similarity/recommendation algorithm is exploited within the individual events on the platform. Users can view similarity scores with other participants to these events. The user-to-user similarity computed score is both global and divided into segments which are used on the application side to further specify the type of similarity. Four main segments have been identified concerning different types of data: ‘preferences’, ‘activities’, ‘history’, ‘locations’. Related users for a certain activity, as regards the use case of the pre and post event activity recommendation, are suggested considering the computed similarity. Recommendation of activities instead is performed mainly on a geographical basis with respect to the geo-location of the event itself, to temporal information about opening and closing times of facilities in the area and by the degree of affinity of the PoI with the characteristics of the group of recommended users. This depends on the analysis of the metadata associated with the PoI for its categorisation. Collaborative filtering on the eSERVANT mobile application is also exploited to define a segment of similar users in order to narrow down the number of users for which to compute recommendations. Interactions taken into account and logged in the system are: 1) user profile views by other users; 2) user’s invitations to a group of activities by other users.

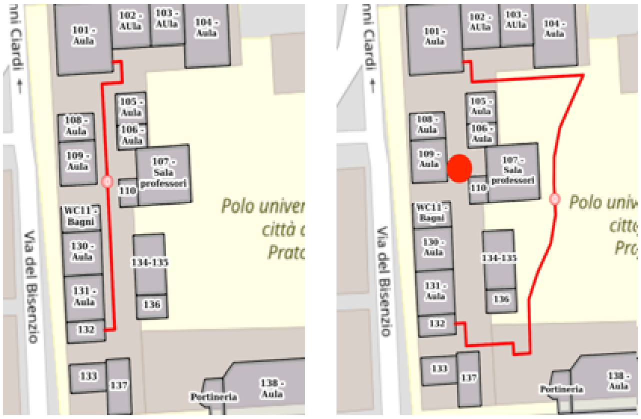

Fig. 3. Dynamic routing. When no crowd is detected the routing leads straight to the destination (left); when an anomalous flow of people is detected (red dot), the algorithm proposes an alternative route (right).

4.2.4. Estimating geographical similarity

For the car-sharing scenario it would have not made sense to recommend only similar people for general interests who, nevertheless, live and/or gravitate in very distant areas. In this sense, the recommendation based on user localizations acts as a filter to the standard recommender based on category affinity. Therefore, to measure this geographical similarity, we exploit all the geo-locations extracted from users' social accounts. Localizations of users inside the facilities, indoor and outdoor, collected through sensors, are also used. These data are expressed as latitude and longitude and are processed in order to establish this geographic user-to-user affinity. The eSERVANT backend platform allows the parametrization of the process of computation of user-to-user geographical similarity. Given that the points to be taken into account may not all have the same importance, the weight of some of them (such as the address of residence) can be manually increased, others instead are dynamically updated on the basis of their frequency in certain areas. Furthermore a coefficient can be configured with the purpose to give more or less importance to the average distances greater than a certain threshold for the final calculation of the affinity. This threshold is estimated through a calculation of the average distances that a user is willing to travel to participate in a specific event, data that is obtained through collaborative filtering on the app. User-to-user geographical affinity is calculated by creating a matrix D that contains all the possible distances between the points of the two sets of user A and user B localizations. For each point of a set, we calculate the distance from the closest point belonging to the other set, thus taking into account the points' density. The similarity based on the distance is normalized from 0 to 1. A value close to zero indicates an high similarity and vice versa

5. Test Case

To demonstrate the usage of the eSERVANT framework, it has been deployed in a university campus in Prato, Italy [4]. During an experimentation period of six months, the campus has hosted numerous events such as courses, workshops, screenings, lectures and sport gatherings. The campus has been equipped with six indoor cameras, covering the two main wings of the ground floor and the first floor, plus an outdoor wide-view camera in the courtyard connecting the wings of the facility. The eSERVANT application has been freely released to students and other participants of the events. The app was used by 976 users to obtain information about the events, recommendations for social gatherings and get direction to and within the facility. We asked 200 users to quantify their satisfaction in order to assess the benefits of eSERVANT for both event participants and facility management. In the following we report an analysis of the experimentation with respect to recommendations and dynamic routing.

Table 1. Recommender evaluation varying the number of suggestions J.

nDCG	J=5	J=10
standard recommendation	0.877	0.761
geo-based recommendation	0.903	0.789

Recommender evaluation. The recommendation systems exploits a custom ranking of user similarities. If we consider a user as a query term and the recommended users as the results in terms of affinity this can be seen as an information retrieval system problem. Consequently, relevance of recommendations have been evaluated exploiting the normalised Discounted Cumulative Gain measure (nDCG). Relevance scores have been computed comparing the list of recommended users with the ideal list given by a group of mobile app users who gave their consent to do the evaluation. To collect the ground truth we asked 200 users of the application to express a relevance score (on a 0 to 5 scale) for the first J people suggested by the system. In Tab. 5 we report the results for J=5 and J=10.

Dynamic Routing evaluation. To assess the quality of the vision-based dynamic routing, we have asked the 200 volunteers to select the best route between two options: the shortest past within the facility and the route proposed by our algorithm (Fig. 3). Each volunteer has answered multiple times (between 5 and 15) for different routes. On average, the users chose 57% of the time the route proposed by our method when no crowds were detected. That is if the detected room occupancy for all rooms along the shortest path is lower than 30%. Interestingly, when the room occupancy was higher than 30% in at least one room along the route, users preferred the vision-based route 93% of the times, demonstrating the usefulness of the algorithm. To further establish the effect of the routing algorithm, we measured the average room occupancy with and without enabling eSERVANT's routing service. Without providing routing information to users in the facility, the average room occupancy was of 56.49% with a standard deviation of 33.02. This means that several rooms were highly crowded while others instead were almost empty. On the contrary, when using the dynamic routing algorithm we registered an average occupancy of 28.64% and a standard deviation of 19.69. This confirms that

[4] https://www.pin.unifi.it/en/

the algorithm was able to effectively balance the occupancy of the rooms, distributing people along different routes.

6. Conclusions

We presented eSERVANT, a platform for smart buildings hosting events with many participants. We rely on computer vision methods to detect people and estimate room occupancies and provide dynamic indoor routing, which avoids congestions. The eSERVANT platform also offers services to users by analyzing social media data in order to facilitate and personalize the event experience. A mobile app is provided which performs recommendation of users for the organization of pre and post event activities and of groups for car sharing.

References

[1] M. Fabbri, F. Lanzi, R. Gasparini, S. Calderara, L. Baraldi, R. Cucchiara, Inter-homines: Distance-based risk estimation for human safety, arXiv preprint arXiv:2007.10243.

[2] J. Ribera, K. Tahboub, E. J. Delp, Automated crowd flow estimation enhanced by crowdsourcing, in: NAECON 2014-IEEE National Aerospace and Electronics Conference, IEEE, 2014, pp. 174–179.

[3] L. Gillooly, P. Crowther, D. Medway, Experiential sponsorship activation at a sports mega-event: the case of cisco at london 2012, Sport, Business and Management: An International Journal.

[4] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, S. Yan, Crowded scene analysis: A survey, IEEE transactions on circuits and systems for video technology 25 (3) (2015) 367–386.

[5] X. Zhang, W. Weng, H. Yuan, J. Chen, Empirical study of a unidirectional dense crowd during a real mass event, Physica A: Statistical Mechanics and its Applications 392 (12) (2013) 2781–2791.

[6] D. Helbing, P. Mukerji, Crowd disasters as systemic failures: analysis of the love parade disaster, EPJ Data Science 1 (1) (2012) 7.

[7] F. T. Illiyas, S. K. Mani, A. Pradeepkumar, K. Mohan, Human stampedes during religious festivals: A comparative review of mass gathering emergencies in india, International Journal of Disaster Risk Reduction 5 (2013) 10–18.

[8] A. Polus, J. L. Schofer, A. Ushpiz, Pedestrian flow and level of service, Journal of transportation engineering 109 (1) (1983) 46–56.

[9] G. Cuffaro, F. Becattini, C. Baecchi, L. Seidenari, A. Del Bimbo, Segmentation free object discovery in video, in: European Conference on Computer Vision, Springer, 2016, pp. 25–31.

[10] F. Marchetti, F. Becattini, L. Seidenari, A. Del Bimbo, Smemo: Social memory for trajectory forecasting, arXiv preprint arXiv:2203.12446.

[11] X. Liu, J. van de Weijer, A. D. Bagdanov, Leveraging unlabeled data for crowd counting by learning to rank, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[12] L. F. Chiara, P. Coscia, S. Das, S. Calderara, R. Cucchiara, L. Ballan, Goal-driven self-attentive recurrent networks for trajectory prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2518–2527.

[13] X. Shi, Z. Ye, N. Shiwakoti, D. Tang, C. Wang, W. Wang, Empirical investigation on safety constraints of merging pedestrian crowd through macroscopic and microscopic analysis, Accident Analysis & Prevention 95 (2016) 405–416.

[14] F. Marchetti, F. Becattini, L. Seidenari, A. Del Bimbo, Multiple trajectory prediction of moving agents with memory augmented networks, IEEE Transactions on Pattern Analysis and Machine Intelligence.

[15] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and localization in crowded scenes, IEEE transactions on pattern analysis and machine intelligence 36 (1) (2014) 18–32.

[16] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 935–942.

[17] A. Ciamarra, F. Becattini, L. Seidenari, A. Del Bimbo, Forecasting future instance segmentation with learned optical flow and warping, in: International Conference on Image Analysis and Processing, Springer, 2022, pp. 349–361.

[18] C. Lu, J. Shi, J. Jia, Abnormal event detection at 150 fps in matlab, in: Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720–2727.

[19] D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe, Learning deep representations of appearance and motion for anomalous event detection, arXiv preprint arXiv:1510.01553.

[20] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, R. Klette, Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes, Computer Vision and Image Understanding 172 (2018) 88–97.

[21] M. Cavallo, E. Guardo, A. Ortis, M. Sapienza, G. La Torre, Mass events monitoring through crowdsourced media analysis.

[22] Z. Xu, Y. Liu, N. Yen, L. Mei, X. Luo, X. Wei, C. Hu, Crowdsourcing based description of urban emergency events using social media big data, IEEE Transactions on Cloud Computing.

[23] M. Bertini, A. Del Bimbo, A. Ferracani, F. Gelli, D. Maddaluno, D. Pezzatini, A novel framework for collaborative video recommendation, interest discovery and friendship suggestion based on semantic profiling, in: Proceedings of the 21st ACM international conference on Multimedia, 2013, pp. 451–452.

[24] L. D. Divitiis, F. Becattini, C. Baecchi, A. Del Bimbo, Garment recommendation with memory augmented neural networks, in: International Conference on Pattern Recognition, Springer, 2021, pp. 282–295.

[25] S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli, User profiles for personalized information access, in: The adaptive web, Springer, 2007, pp. 54–89.

[26] A. F. Abate, C. Bisogni, L. Cascone, A. Castiglione, G. Costabile, I. Mercuri, Social robot interactions for social engineering: Opportunities and open issues, in: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, IEEE, 2020, pp. 539–547.

[27] S. Ouaftouh, A. Zellou, A. Idri, User profile model: a user dimension based classification, in: 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), IEEE, 2015, pp. 1–5.

[28] S. Alaoui, Y. E. B. E. Idrissi, R. Ajhoun, Building rich user profile based on intentional perspective, Procedia Computer Science 73 (2015) 342–349.

[29] M. Chen, A. A. Ghorbani, et al., A survey on user profiling model for anomaly detection in cyberspace, Journal of Cyber Security and Mobility 8 (1) (2019) 75–112.

[30] M. Mezghani, C. A. Zayani, I. Amous, F. Gargouri, A user profile modelling using social annotations: a survey, in: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 969–976.

[31] A. Farseev, M. Akbari, I. Samborskii, T.-S. Chua, " 360 user profiling: past, future, and applications" by aleksandr farseev, mohammad akbari, ivan samborskii and tat-seng chua with martin vesely as coordinator, ACM SIGWEB Newsletter (Summer) (2016) 1–11.

[32] S. Liang, X. Zhang, Z. Ren, E. Kanoulas, Dynamic embeddings for user profiling in twitter, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1764–1773.

[33] J. Chen, Y. Liu, M. Zou, Home location profiling for users in social media, Information & Management 53 (1) (2016) 135–143.

[34] G. Farnadi, J. Tang, M. De Cock, M.-F. Moens, User profiling through deep multimodal fusion, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 171–179.

[35] M. Bertini, A. Del Bimbo, A. Ferracani, D. Pezzatini, A social network for video annotation and discovery based on semantic profiling, in: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 317–320.

[36] A. Ferracani, D. Pezzatini, M. Bertini, S. Meucci, A. Del Bimbo, A system for video recommendation using visual saliency, crowdsourced and automatic annotations, in: Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 757–758.

[37] A. Ferracani, D. Pezzatini, M. Bertini, A. Del Bimbo, Item-based video recommendation: An hybrid approach considering human factors, in: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, 2016, pp. 351–354.

[38] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.