

An Indoor Location-aware System for an IoT-based Smart Museum

S. Alletto, R. Cucchiara, G. Del Fiore, L. Mainetti, V. Mighali, L. Patrono, and G. Serra

Abstract— The new technologies characterizing the Internet of Things allow realizing real smart environments able to provide advanced services to the users. Recently, these smart environments are also being exploited to renovate the users’ interest on the cultural heritage, by guaranteeing real interactive cultural experiences. In this paper, we design and validate an indoor location-aware architecture able to enhance the user experience in a museum. In particular, the proposed system relies on a wearable device that combines image recognition and localization capabilities to automatically provide the users with cultural contents related to the observed artworks. The localization information is obtained by a Bluetooth Low Energy infrastructure installed in the museum. Moreover, the system interacts with the Cloud to store multimedia contents produced by the user and to share environment-generated events on his/her social networks. Finally, several location-aware services, running in the system, control the environment status also according to users’ movements. These services interact with physical devices through a multi-protocol middleware. The system has been designed to be easily extensible to other IoT technologies and its effectiveness has been evaluated in the MUST museum in Lecce, Italy.

Index Terms—Bluetooth Low Energy, Image recognition, IoT, Location-awareness, SoA, Smart Museum.

I. INTRODUCTION

Art and culture have always played an important role in human beings lives. Over the centuries, hundred of museums and art galleries have preserved our diverse cultural heritage and served as important sources of education and learning. Museums are nowadays a tool of entertainment such as theatres or cinemas. Today, museums and art galleries usually provide visitors either with paper booklets or with audio guides. Visits at museums are often considered boring, because it is hard for museums curators to catch the attention of tourists. In particular, it is difficult to define in advance a tour for all the visitors, because interests may vary from person to person. Interests are different from children to

adults, students group from single visitor, casual visitor to fond-visitor. Therefore, interactive and personalized museum tours need to be developed. In this perspective, a significant contribution can be given by the next Internet of Things (IoT), which involves the extension of the Internet to small and low-cost “things” that are thought to realize smart environments in order to provide new services to the users. IoT aims to create a better world for people, where smart objects around us know what we like, what we want and act accordingly without explicit gestures. To achieve this goal, the industrial world is strongly focused on adopting low-power and low-cost embedded technologies in everyday objects, which become real smart objects. Unfortunately, the widespread diffusion of these objects is still strongly limited by several weaknesses, such as high heterogeneity among devices. It leads to strictly vertical software solutions, meaning that they are focused on specific contexts and only on certain technologies. On the contrary, it would be much more interesting and meaningful to have systems capable of integrating the heterogeneous technologies that will characterize the future smart environment, enabling transparent interoperability among smart devices. This horizontal approach would enable rapid and efficient development of services and applications, since designers and developers should no longer worry about the low-level details.

Moving to the upper layers, another highly debated issue relating to the smart environment concerns their capability to evolve as autonomously as possible, with minimal human intervention. Although the voluntary interaction between users and physical devices is still important (e.g., for security reasons), the ability of the environment to automatically react to some events is even more attractive and challenging. In other words, the transition from the Human-to-Machine (H2M) communication paradigm towards the Machine-to-Machine (M2M) communication paradigm would definitely lead to an improvement of the user experience. To achieve this result, multiple applications (services) should constantly manage the interaction among physical devices, in accordance with predefined business logic. The current trend is to configure these services as location-aware services, i.e., applications driven by location information, in particular, by users movements in the environment.

Taking into account all these considerations, we propose a system able to address all the above-described issues. More in detail, the solution presented in this paper enables wearable devices, interacting with an IoT-based smart environment, to

Manuscript received on July 13, 2015. Revised on November 6, 2015.

G. Del Fiore, L. Mainetti, V. Mighali, and L. Patrono are with the Department of Innovation Engineering, University of Salento, Lecce 73100, Italy (e-mail: {luca.mainetti, vincenzo.mighali, luigi.patrono}@unisalento.it)

S. Alletto, R. Cucchiara, and G. Serra are with the Department of Engineering, University of Modena and Reggio Emilia, Modena 41125, Italy (e-mail: 72056@studenti.unimore.it, {rita.cucchiara, giuseppe.serra}@unimore.it).

act as museum guides, providing a real interactive cultural experience. The whole system becomes a generator of events, which can be used to enhance the user experience. For example, when a user is in front of an artwork, several details such as title, artist, historical context, critical review can be easily and automatically provided. The information can refer not only to the global artwork, but also to details or to the entire room. For instance, particular faces or sub-scenes of large painting or frescoes can be identified. The cultural contents could be sent individually to a specific user or made available through multimedia walls in the museum room. In addition, the user could store vocal comments about his/her extemporaneous feelings, so that s/he can live them again at a later stage. The cultural experience could also be shared on the social networks, by automatically posting information about the particular artwork that is being admired. Then, the environment itself should modify its status according to specific events (e.g., the number of users in a room) or to the visitors' personal profile. Such an augmented reality application could assist to appreciate art more deeply and make it more accessible to everyone. Finally yet importantly, the information collected from the environment could also be used for the management of the entire facility by the museum supervisors. For example, the number of users during the hours of the day could be exploited to reorganize the opening and closing times, whereas the capability to know the most visited rooms could be used for planning partial maintenance works.

To provide all these features, the user is equipped with a wearable device able to capture videos and images, whereas the actual business logic is managed by several location-aware services running on a processing center. More in detail, the wearable device accomplishes two main tasks: it continuously tracks the user by leveraging a Bluetooth Low Energy (BLE) infrastructure, and recognizes the artwork in front of the user through both its processing capabilities and localization information. The results of this twofold activity are sent to the processing center and then used by the location-aware services that are in charge to provide all the other features of the system. In particular, they (i) provide cultural contents to the visitors, (ii) communicate useful information to external users, and (iii) interact with heterogeneous technologies that control the status of the environment (e.g., a building automation system that manages lighting and thermoregulation of the museum). To accomplish the last task, they exploit a multi-protocol middleware that allows a transparent access to heterogeneous IoT technologies, hiding the low-level communication details. This middleware is designed to be easily extended to new technologies, in order to improve flexibility and scalability.

Regarding image recognition, we present a method for artwork recognition that gives to the visitor an automated description about the artworks. Our approach can deal with rapid changes in illuminations, significant camera motion and presence of occlusions. Local features are extracted from the input frame and matched with a candidate target in the

database. The RANSAC [1] algorithm is used to detect feature outliers. In particular, it is composed of two main parts: recognition of the painting from the dataset with a SIFT-based approach, and temporal and spatial reasoning to obtain a robust detection.

The effectiveness of the proposed architecture is evaluated in two successive phases. First, the performance of both the image recognition algorithm and the localization service is analyzed through specific stressing tests, whereas the whole architecture is evaluated in a real scenario staged at MUST museum in Lecce, Italy.

The rest of the paper is organized as follows. Section II summarizes the motivation underlying the proposed architecture. The related works are reported in Section III. Section IV provides a detailed description of the proposal. In Section V the results of the system validation are provided. In Section VI, a comparison with another architecture already in the literature is presented. Conclusions are drawn in Section VII.

II. MOTIVATION

Exploiting new IoT-enabling technologies to create smart environments able to predict users' desires is the current trend in both academies and industries. However, these technologies are not widespread in cultural environments, where the innovation process is growing at a slower pace. This situation entails that the cultural heritage is a prerogative of a restricted category of users, which are actually interested in this field. In order to involve other users, above all young people, it is necessary to provide for a significant technological improvement in the locations dedicated to the culture. Unfortunately, these places do not usually allow the installation of hardware infrastructures able to provide new services to the users, so any innovation should be as less invasive as possible. For this reason, the only advanced tools available in cultural locations, such as museums, are the classical audio-guides provided to the visitors or, at most, some "smart" objects, such as QR code or NFC tags near artworks, which allow receiving static information about the artwork itself on the user's mobile device.

This status quo is the principal motivation underlying the proposed system. The main objective is to realize a real smart environment, in which "cultural things" are able to speak to the visitor, and to do so in a personalized way. This means that each user should receive different kind of information based on his/her specific personal profile. Moreover, to improve the user experience as much as possible, this interaction should be fully automatic, i.e., without any explicit action of the user. Of course, this result has to be achieved in a very unobtrusive way in order to be actually feasible in a real scenario.

The integration of image recognition capabilities with a low-cost localization infrastructure allows achieving the desired goals. Furthermore, the possibility to involve Cloud services, such as Social Networks, for sharing the cultural experience can be a strong driving factor to approach young people to the cultural world.

III. RELATED WORKS

In the literature, there are several works addressing the aforementioned issues, but none of them provides a flexible and scalable solution that can solve all the problems in one system. As said, one of the key features of the proposal is represented by the indoor localization mechanism, which currently is an important and challenging research topic. In [2], the authors present a personalized smart control system that: (i) localizes the user exploiting the magnetic field of the smartphone, and (ii) controls appliances present at the user's location. Another example of location-aware services in a smart environment is reported in [3]. Here, the authors propose a system that collects information from the environment and then provides services to improve the lifestyle of the users, mainly from an energetic point of view. An interesting solution is provided in [4]. In this case, the authors propose an access control mechanism, which consists of an engine embedded into smart objects able to make authorization decisions by considering both user location data and access credentials. User location data are estimated using magnetic field measured and sent by the phone. Finally, in [5, 6] authors present location-aware systems able to localize users by exploiting an infrastructure based on the Bluetooth Low Energy technology. Then, the localization information is used to adapt the home environment to the users' needs.

With regard to the interaction with smart environments, some solutions in the literature focus on specific technologies and aim at simplifying the development and customization of user applications. For example, in [7], authors develop and validate an architecture, both hardware and software, able to monitor and manage a KNX-based home automation system. Instead, to solve the problem of devices heterogeneity, most of the approaches are based on the concept of middleware. It is an important architectural component able to present high-level interfaces to the upper layers, in order to mask the heterogeneity of underlying technologies. For example, in [8], the authors focus the attention on various integration styles for non-IP based devices already deployed in home and building automation systems. Instead, in [9], the readiness and compatibility of existing building automation system technologies with IPv6 are investigated, and the integration challenges and new opportunities of IPv6 for these technologies are presented. Finally, in [10], the authors leverage the principles of the Web of Things approach to implement a gateway able to expose capabilities of a KNX home automation system as Web Services.

Focusing on the museum environment, several solutions have been recently proposed for interactive guides enhancing cultural experiences. An example is the "SmartMuseum" system [11], in which visitors can gather information about what the museum displays and customize their visit based on specific interests. This system, which integrates PDAs and RFIDs, brought an interesting novelty when first released, but it has some limiting flaws. In fact, researches demonstrated

how the use of PDAs devices on the long term decreases the quality of the visit due to their users paying more attention to the tool rather than to the work of art itself. Other interesting examples focused on RFID technologies are [12, 13]. In these works, the authors describe smart systems able to provide users with cultural contents by exploiting the interaction between an RFID reader integrated on users' mobile devices and RFID tags placed near each artworks. Although interesting, these solutions require the use of mobile devices equipped with RFID readers, which are expensive and not so common. Moreover, the solution in [13] also needs the installation and maintenance of specific RFID infrastructures for users' tracking. In 2007, another system described by [14] aimed to customize visitors experience in museum using software capable of learning their interests based on the answers to a questionnaire that they compile before the visit. Similarly to the Smart Museum, one of the main flaws of this system is the need to stop the visitors and force them into doing something that probably s/he would not want to do. The museum wearable [15] is a storytelling device: it is a museum guide which in real-time evaluates the visitor's preferences by observing his/her path and length of stops along the museum's exhibit space, and selects content from a database of available movie clips and audio. However this system does not use any algorithm for visual analysis of understanding the surrounding environment. Furthermore the localization is based on an indoor infrared positioning system; this methodology allows the system to roughly estimate the position and therefore the content delivery may not be well suited for the visitor.

At the same time, research efforts have been made on the definition of techniques to automatically recognize objects, actions, and social interactions. Interest points and local descriptors were used by many authors and appear much more appropriate to support detection and recognition of object in real world images and video. In fact, local visual descriptors like MSER [16], SIFT [17], SURF [18], have been proven to be able to capture sufficiently discriminative local elements with some invariant properties to geometric or photometric transformations and are robust to occlusions. However, they suffer of high blur, a typical characteristic of video acquired by a wearable camera, and in this scenario achieve a low recognition accuracy.

IV. THE PROPOSED ARCHITECTURE

Fig. 1 shows the overall structure of the proposed system architecture. It is composed, as described below, of three main building blocks:

- The localization service: it is distributed between the wearable device and the processing center. The first one detects the current user's position and communicates it to the processing center. Here, the localization information is stored and made available to other services. The information is also used locally (on the wearable device) to speed up the image-processing algorithm.

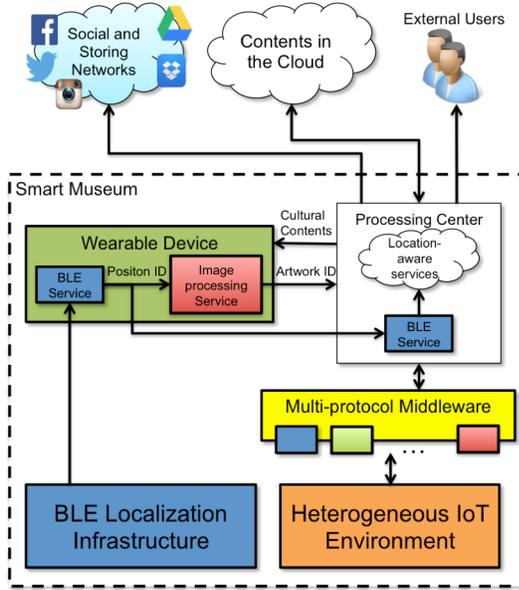


Fig. 1. The overall structure of the system.

- The image-processing algorithm: it is able to detect, in real-time, the artwork observed by the user. It can quickly analyze the video frames captured by the wearable vision device and identify the target object with high accuracy and reliability. The result of the processing activity is then sent to the processing center.
- The processing center: it is the core of the business logic. It accesses, in the Cloud, the cultural contents required by the users and smartly provides such contents on several interactive platforms. Then, it allows the execution of several location-aware services by providing them with the positioning information coming from the localization infrastructure. These services enrich the cultural experience of the users by immersing them in a real interactive environment controlled through the multi-protocol middleware. Moreover, they allow visitors to preserve their memories by interacting with external world and, finally, they also support the management of the museum.

The focus on wearable devices is due to the potential negative impact of mobile devices on the users' cultural experience, as said in Section III. Of course, to provide a real customized experience, a profiling process has to be performed before giving the wearable device to a user. To do so, several approaches could be exploited. In the proposed solution, the queued users can download a mobile application on their smartphone and compile a simple questionnaire for profiling purposes. At the end of the procedure, the application stores the result on the processing center and generates a QRcode on the display. At the ticketing desk, the box office worker scans it to associate the outgoing wearable device with the user's profile. If a user does not have a smartphone, s/he can follow the same procedure through a digital totem.

A. Indoor localization

As said, several system components depend on the localization service. It consists of three main elements: (i) an infrastructure of wireless landmarks that periodically send localization information, (ii) a service installed on the wearable smart gateway that collects the information of the landmarks to determine its location, and (iii) the service running on the processing center that receives the location of the user and provides it to the other services. More specifically, the network of wireless landmarks consists of embedded devices equipped with Bluetooth Low Energy (BLE) interface and placed individually in the different rooms of the building. The choice of BLE is mainly due to its low energy consumption in front of a communication range comparable with that of the traditional Bluetooth. In this way, the wireless landmarks can be battery powered, making the localization mechanism more flexible and less invasive. Each device of the BLE infrastructure sends its location identification (ID) together with the transmission (TX) power value. The service running on the user's wearable device collects location data from all the landmarks within its listening range and then determines the room in which it is located. To do so, it computes a proximity index d , for each landmark, using the corresponding value of the RSSI. The adopted equation is the following:

$$RSSI = -(10n \log_{10} d + A) \quad (1)$$

where A is the received signal strength at 1 m, n is a signal propagation constant depending mainly on the environment, and d is the distance from the sender [19]. The landmark that has the lowest value of d represents the user's current location. This information has a twofold function: (i) it is used locally to speed up the image processing algorithm, and (ii) it is sent to the service running on the processing center, which makes it available for any interested service. Fig. 2 shows the main steps of the user localization process.

B. Artwork recognition

Before any attempt at recognizing an artwork, a useful pre-processing step should be made. In fact, a typical ego-vision

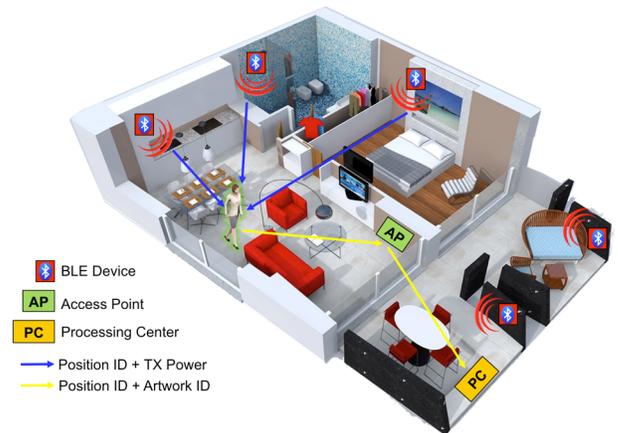


Fig. 2. The main components of the localization mechanism.

characteristic is that the user with the camera can have very fast head motion, e.g. when s/he is looking around for something. This leads to high blur in that part of the video sequence, which results in a very low quality. Thus, removing blurred frames from the processing not only can improve the quality, but also prevents the system from giving the user information s/he is not interested in.

This is done analyzing the amount of gradient in the scene. In particular, an equation that recognizes the blur degree in a frame f has been defined:

$$\text{Blur}(f, \theta_B) = \sum_I \sqrt{\nabla S_x^2(f) + \nabla S_y^2(f)}, \quad (2)$$

where $\nabla S_x(f)$ and $\nabla S_y(f)$ are the x and y components of Sobel's gradient in the frame. A threshold θ_B , learned by computing the average amount of gradient in a sequence, is used to discard frames with excessive motion blurriness.

Once a frame that satisfies Equation 2 is identified, the recognition process can be started. To match the framed artwork and its counterpart in the museum database, we extract Scale Invariant Feature Transform (SIFT) local descriptors [17].

The preliminary experiments showed that a detection step aimed at sampling keypoints only from a painting could not provide satisfactory results, since detection based on local appearance can often produce many false positives (windows, doors, etc.). To improve the match quality, a spatial verification by fitting a homography using RANSAC is performed. To further improve the matching results, a first thresholding step is carried out: using a threshold over the distances among SIFT descriptors θ_S , we remove the matches which have a large distance.

Since keypoints are sampled in the whole image, a second thresholding step is introduced for discriminating frames effectively containing an artwork from frames where many keypoints are detected on architectonic details. The threshold θ_D is defined over the ratio between matches that survived the previous pruning steps (RANSAC and θ_S , K_R and K_{θ_S} respectively) and the original amount of keypoints in the current frame K_0 , so that the set of accepted frames is :

$$F_a = \{f \mid (K_R + K_{\theta_S})/K_0 > \theta_D\} \quad (3)$$

Adjusting this threshold can render the method more robust to noise and occlusions.

Exploiting the location-awareness of our system, it is possible to greatly reduce the computational requirements of the matching process and increase its accuracy. This is done by analyzing the localization information obtained by the BLE infrastructure. The current frame is matched against the templates of the artworks that belongs to the room where the user currently is. It is worth noting that the image recognition algorithm also aims at significantly preserve the energy of the wearable device. Indeed, on the one hand, the BLE technology requires a very low communication power; on the other hand, the savings in terms of computational effort allows a further reduction of power consumption. A summarization of the painting recognition method is presented in Fig. 3.

C. Cultural contents delivery and Cloud services

The output obtained by the image processing algorithm, i.e. the unique identifier of the observed artwork, represents the key information for accessing the desired cultural contents. The wearable device sends this information to the processing center through the local WiFi network. There, a specific service is in charge of receiving all requests coming from users and analyzing them to start the proper procedure. More in detail, the interpretation of the artwork identifier can lead to two possible results:

1. An audio description of the artwork on the user's wearable device;
2. Multimedia cultural contents on interactive walls of the museum.

In the first case, an audio-streaming server application provides the interested clients with the audio contents related to the observed artwork. In the second case, the processing server exploits the wired local networks to send multimedia contents to interactive displays and totems in the museum, so that the same cultural information is simultaneously available to all the involved users. The processing center takes the decision depending on a threshold algorithm that constantly monitors the number of simultaneous visitors looking the same artwork. More in detail, when the number of visitors is below a predetermined threshold, the algorithm converges to the first alternative. On the contrary, if there is a large group of visitors sharing the same "visiting experience", the algorithm converges to the second alternative. It is worth

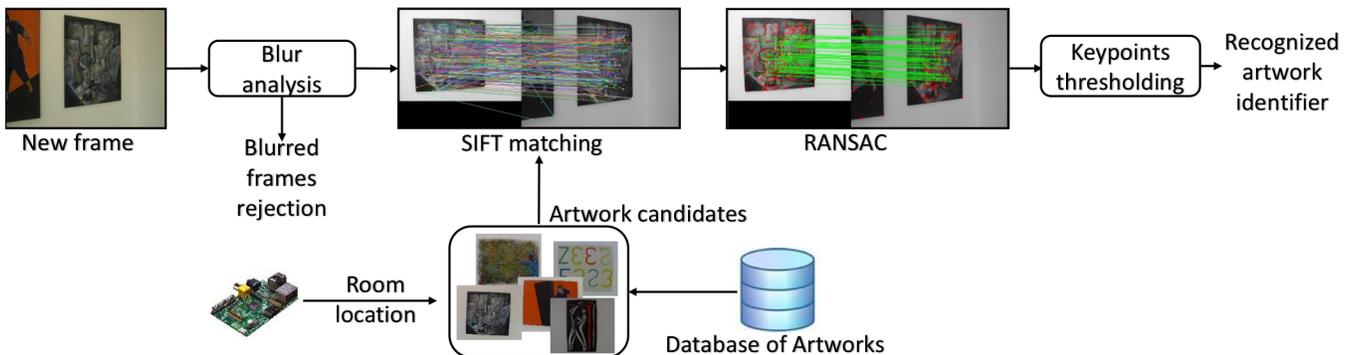


Fig. 3. An overall view of the painting recognition method.

noting that retrieving information and multimedia contents related to the cultural tour of each user is an expensive process from both the computational and memory point of view. For this reason, the Cloud seems to represent the solution that best suits this kind of needs, as its storing and computing capabilities allow to process data more efficiently. In particular, in the proposed system, the Cloud is accessed by the processing center whenever the running services need to retrieve cultural contents destined to the users.

The Cloud is also exploited to provide two other interesting services. The first one allows each visitor to store vocal comments about his/her cultural experience on a storage space in the Cloud, so that s/he can hear again them after the visit. To accomplish this task, the wearable device is equipped with a microphone and, of course, it is associated to a storage user account in the Cloud. More specifically, when the wearable device is provided to the user, a quick profiling procedure is performed to both associate the device to the user and create an account in the Cloud to store the vocal comments s/he generates during the cultural experience. The Cloud storing is realized exploiting the Amazon Simple Storage Service (Amazon S3) [20], which provides proper APIs for several client platforms. In this way, the wearable device becomes an Amazon S3 client that stores the audio files in the Cloud and generates the REST URIs to get these files back from the Web. Of course, the file URIs are stored in a database and are associated with the correspondent user ID. The second Cloud service concerns the social activity of the user. If s/he enables the “social option” when the wearable device is delivered at the beginning of the tour, every interesting event triggered by the environment is automatically shared on user’s social networks. Currently, the proposed system is able to share events on Facebook by exploiting the Facebook Graph APIs [21], which allow managing (e.g., post, delete, update, etc.) messages on a Facebook account. In more detail, when a meaningful event is detected from the environment, the application running on the wearable device authenticates the user on Facebook by exploiting the OAuth 2.0 authentication standard [22] and then shares the event through the Graph APIs.

Finally, the architecture exposes in the Cloud another useful service that provides statistical information about the busyness of the museum. Indeed, by exploiting the localization information, this service always knows how many visitors are moving in the museum and where they are. Therefore, this service can be used by external users to know in advance the length of queues in specific areas of the museum or which are the most admired artworks. Moreover, the information provided by this service could also be exploited by the museum supervisors to schedule partial maintenance works or to reorganize the internal spaces.

D. Interaction with IoT environment

One of the main tasks of the services running on the processing center is to adapt the status of the environment according to the information coming from the localization

service. More in detail, exploiting IoT-aware technologies, the environment could be modified in real-time in order to provide the user with a real interactive experience. As an example, imagine that the museum has a special room where an historical war is represented by a mechanical animation managed by several IoT actuators. To maximize the impact of this animation, the system could decide to activate it only when the number of visitors in the room is higher than a predefined threshold. In the same way, lighting, temperature and other physical characteristics of a room could be controlled to automatically perform special effects typical of a 4D cinema. Obviously, the IoT technologies able to provide these features could be extremely heterogeneous since they are often compliant to different standards and protocols. In order to efficiently interact with such a kind of technologies, the services of the processing center exploit a multi-protocol middleware that allows a transparent access to the underlying heterogeneous devices, hiding the low-level communication details. In particular, on the one hand, it provides the services with high-level RESTful APIs to communicate with the physical network, whereas, on the other hand, it is equipped with specific software modules, called *adapters*, which communicate with the IoT devices in accordance to specific standards and protocols. The modular structure that characterizes the middleware allows to easily extend it to new technologies, so guaranteeing flexibility and scalability. Actually, the middleware is equipped with three main adapters, which allow interacting with Costrained Application Protocol (CoAP), KNX and BLE compliant devices (Fig. 4). The initial choice of KNX, CoAP, and BLE is due to their diffusion in both commercial and academic solutions available in the literature. KNX is the worldwide standard for home and building control, CoAP is one of the most used application protocol in the IoT, which provides a lightweight access to physical resources, and BLE is more and more the leading technology in commercial smart devices. However, any other IoT technology can be integrated into the system.

V. SYSTEM VALIDATION

A. Test environment

The components used for the validation phase are: a wearable gateway, a wearable vision device, a processing center, a multimedia wall, and an infrastructure that is able to ensure the indoor localization service.

The smart gateway is realized through an embedded computer, namely an Odroid-XU [23]. It is a single ARM

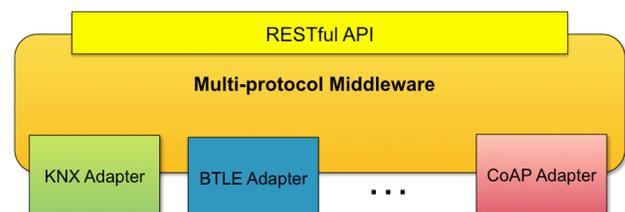


Fig. 4. The architecture of the multi-protocol middleware.

board measuring just 94x70x18 mm, it has a Samsung Exynos5 Octa Core processor (Cortex-A15) of 1.2 GHz with PowerVR SGX544 graphics card MP3 and it is equipped with 2 GB of DDR3 RAM.

The image acquisition is performed through an Odroid USB-cam 720p, which has a resolution of 1280*720 HD, an USB 2.0 plug-and-play interface and supports up to 30 fps. Fig. 5 shows the hardware used to realize the wearable device.

The device that receives and processes the cultural content sent by the processing center when there is a large number of visitors has been realized through a Raspberry Pi Model B [24]. It is equipped with 512 MB of RAM and is based on the Broadcom BCM2835 system with a chip (SoC) that includes an ARM1176JZF-S 700 MHz processor and a VideoCore IV GPU. Furthermore, the system has MicroSD sockets for boot media and persistent storage. The Raspberry Pi board is connected to an interactive wall in order to display the cultural contents received by the processing center.

Same boards were also exploited to realize the BLE landmarks that make up the indoor localization infrastructure.

B. Results

The artwork recognition method was tested on the real and unconstrained dataset acquired with a head-mounted camera at MUST museum (Lecce - Italy). The dataset contains more than 2000 frames at 640×480 resolution annotated with the current visible artworks and their room location.

This amount of frames represents a challenging sequence characterized by different types of artworks, different levels of light, and blur due to motion and occlusions. The recognition capability was evaluated in terms of detection precision and recall, and classification accuracy. The first two metrics represent the detection capability: a high precision means that frames containing artworks are correctly detected, whereas a high recall means that few frames containing artworks are missed. The accuracy metric measures the matching performance, showing how many artworks are correctly classified.

Since the proposed method is based on two different thresholding steps, the results concerning detection and

recognition are shown separately. This allows analyzing how different values can influence the performances of the proposed approach (see Fig. 6).

Note that θ_S is responsible to filter bad matches. In fact, setting it to a low value yields a high precision, but achieves very low recall and thus low accuracy, discarding most of the correct matches. On the other hand, if its value is too high, it cannot filter bad matches and therefore the overall accuracy decreases (see Fig. 6.a). Similarly, θ_D controls the detection performance: increasing its value the precision improves but, at some point, recall sharply drops ($\theta_D > 0.1$). This is due to the fact that the method starts discarding too many frames. Based on these considerations we fixed these thresholds θ_S and θ_D to 270 and 0.06, respectively. It can be seen how the chosen values represent the best compromise in terms of overall performance.

In the second experiment, it is shown how the location-awareness of the proposed system impacts its performance. Table I presents the results of the proposed method in terms of precision, recall and accuracy. Here it is compared to a baseline obtained using a standard SIFT matching technique [17]. Table I also reports the performance of the approach without the use of the user localization.

As Table I shows, our system, exploiting localization information, provides the best results. It is closely followed by its own variant that does not rely on localization. The small gap in recognition performance is mainly due to the robustness of the method. In fact, the proposed solution has good discriminative capabilities and can effectively identify the correct artwork regardless the number of pieces in the museum dataset. The baseline achieves 100% recall, since it treats every frame as an artwork lacking the detection component. However this performance is the result of a significant number of false positives that lead to a significant loss in terms of accuracy, validating the use of a detection threshold.

While the impact of user localization on accuracy is limited, it can greatly influence in the computation performance. In fact, execution times of the proposed method can be roughly divided into two main components: descriptor computation and matching. Given the magnitude of these two steps, the impact of other components such as BLE communication is too small to be significant.

For each frame that needs to be processed, a sparse SIFT descriptor must be first computed. Using the hardware and image resolution described above, this step requires on average 1200 ms. This process cannot be avoided and is not influenced by the localization. Once the current frame descriptor has been computed, it has to be matched against the artwork templates. Since their descriptors are pre-computed,

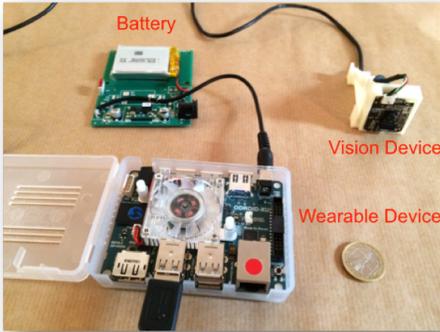


Fig. 5. The wearable device.

TABLE I
PERFORMANCE OF THE PROPOSED PAINTING RECOGNITION ALGORITHM

Solution	Precision	Recall	Accuracy
Our System	0.783	0.946	0.436
Our System w/o localization	0.769	0.979	0.429
SIFT baseline	0.760	1	0.201

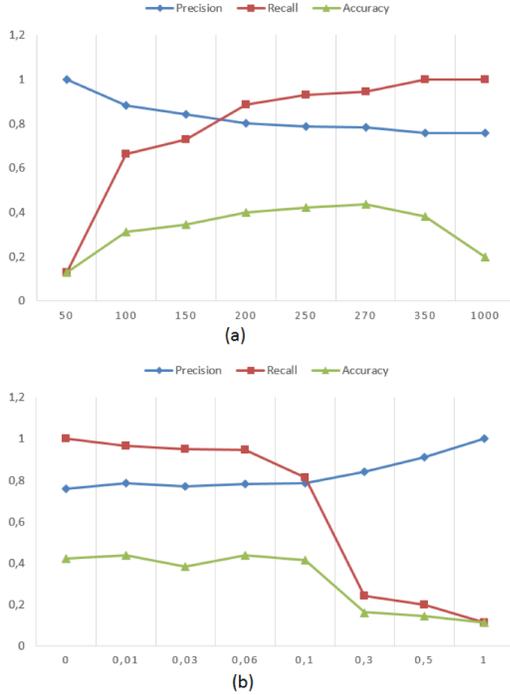


Fig. 6. Precision, recall and accuracy of our painting recognition solution under different threshold values: a) θ_S ; b) θ_D .

matching against one image takes an average of 100 ms. It clearly emerges how reducing the number of database entries can greatly decrease the computation effort. In fact, if no localization information is used, the time required by the matching component increases linearly with the number of artworks in the museum. As Fig. 7 shows, assuming an average of 5 artworks per room and a total of 10 rooms, exploiting localization results in an average processing time per frame of 1700 ms. On the other hand, matching against the entire museum database results in an average of 6200 ms (a 365% increase in required computational times).

Having an algorithm that is more than three times faster than its counterpart, which does not rely on localization, is of paramount importance in this setting. Aiming at a real-time wearable application, a faster algorithm means lower framerate subsampling and results in increased responsiveness. Furthermore, the proposed system is battery-

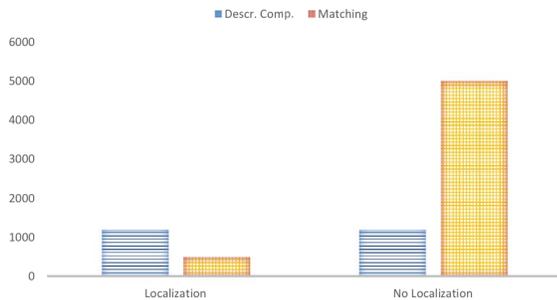


Fig. 7. Average time consumption of descriptor computation and matching in ms.

powered, and reducing the amount of computation needed in each frame can greatly improve the battery life allowing for longer visits.

Finally, considering the granularity of a room, the correct visitor's localization has been tested in two different cases. In the first case, referred as *best case*, the BLE devices characterizing the localization infrastructure have been placed on the partition wall between the two rooms, so that they are not in line of sight of each other. In the second case, referred as *worst case*, the devices have been placed in line of sight at 5 meters from the separating door. The results shown in the Fig. 8 were obtained by positioning the wearable device at three different distances from the door. The measurements allowed us to identify the Successful Localization Probability, which is the probability of correct localization of the visitor inside the room. The Fig. 8 shows how the results obtained are optimal in the best case and next to the ideal ones in the worst case.

VI. DISCUSSION

As highlighted by the state-of-art analysis proposed in Section II, very few solutions address the problem of delivering personalized content during a museum visit. Moreover, to achieve this goal, they differ in both hardware and software architectures.

To provide a comparison between the proposed framework and these solutions, namely SmartMuseum [11] and MuseumWearable [15], a description of the adopted technologies is firstly provided. First of all, since a key feature of every smart museum system is the ability to follow the tourist during his/her visit, all the solutions analyzed in this section exploit mobile technologies. Either PDAs or wearable devices are used in order to provide the visitor with a small and light-weighted device. Beyond this common feature, the compared architectures show deep differences in the technologies designated to identify artworks, localize the user and implement communication (see Table II). Furthermore, they provide different additional services to users.

In [11], the system adopts two different approaches in order to get information about the artwork the user is looking at. In more detail, the visitor can either type a number related to the artwork or scan a RFID using the provided PDA. Once the artwork has been selected, the visitor receives personalized

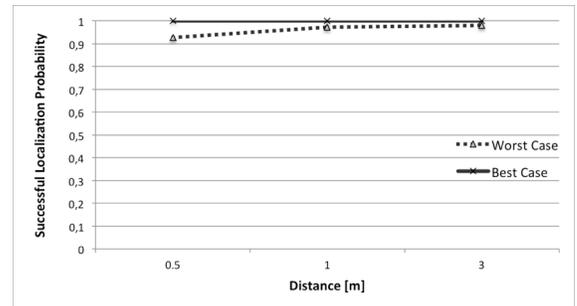


Fig. 8. Successful Localization Probability with the worst and best disposition of the BLE devices.

TABLE II
COMPARISON BETWEEN THE PROPOSED ARCHITECTURE AND SIMILAR SYSTEMS PRESENTED IN LITERATURE

Reference work	Technologies			Services provided					Communication
	CV	RFID	Mobile	Audio Information	Personalized Information	Interaction-less Requests	Location Awareness	Video Information	
[11]		✓	✓	✓	✓		✓		None
[15]			✓		✓	✓	✓	✓	Infrared
[25]			✓	✓	✓		✓	✓	Not specified
Proposed	✓		✓	✓	✓	✓	✓		BTLE

information based on his/her profile, which s/he previously set up through a Web application. This architecture relies on the user manual interaction to achieve both artwork recognition and location awareness. In addition, no IoT architecture is actually present and no recent communication protocols are employed.

On the other hand, in [15], the system relies on a smart environment equipped with infrared location sensors, which provide the location awareness by interacting with a wearable IR receiver. Moreover, rich visual information in an augmented reality fashion is given to the visitor through a wearable, head-mounted display. Information personalization is achieved through a stochastic approach, by studying how the user reacts to each artwork, i.e. how long s/he stare at it. WiFi technology is used to implement communication between the wearable computer and the museum infrastructure, in order to coordinate the information retrieval. The architecture does not rely on direct user interaction since the personalization is achieved using a statistical algorithm.

The PIL [25] system architecture is based on a multi-agent communication network that enables different tourists to communicate among each other using a customized user interface and a PDA. A spatial information broker is employed in order to achieve location awareness inside the museum, a feature exploited to improve the communication. Being outside the scope of the paper, no explicit detail about the employed communication protocols is provided.

The architecture presented in this paper has some similarities to the aforementioned approaches, but presents some significant differences. All solutions rely on mobile devices and are capable of providing personalized information based on some sort of location awareness. After all, these features should be the core components of any smart museum device capable of enriching a visitor's experience. However, a fundamental characteristic that distinguishes the proposed system from the other solutions is the degree of interaction required to obtain such information. By exploiting a Computer Vision (CV) algorithm, capable of recognizing in real-time what the user is looking at, the proposed system eliminates the need for explicit interaction, so effectively reducing distractions that may arise from continuously checking on a user interface. Furthermore, the location awareness is achieved

with high precision and low costs by exploiting the recent BLE communication technology. Localization of the user through BLE devices instead of RFID tags tied to each artwork has the important advantage of being independent from the museum layout. In fact, while rearranging an exposition would require rearranging every tag in order to maintain a coherent localization, in our solution, only room information is broadcasted, so there is no need to worry about changes in the artworks' disposition.

VII. CONCLUSIONS

In this work, an indoor location-aware architecture for smart museums was designed and validated. In more detail, the proposed system relies on a wearable device equipped with image recognition and localization capabilities to automatically provide users with cultural contents related to the observed artworks. The ability to identify the user's position is guaranteed by an infrastructure of Bluetooth Low Energy transmitters. The architecture also consists of a processing center, where the actual business logic is in charge to (i) retrieve from the Cloud the cultural information related to the observed artworks and (ii) manage the status of the indoor environment in accordance to users' position. Finally, the system accesses the Cloud also to store multimedia contents produced by the user and to share environment-generated events on his/her social networks. Even if the proposed system could suffer from potential maintenance issues caused by hardware damages due to intensive use, it does not require the use of expensive devices or infrastructures, as well as it does not force the users to exploit their own devices. The effectiveness of the whole architecture has been evaluated in the MUST museum in Lecce, Italy.

References

- [1] M. A. Fischler, and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, no.6, 1981, pp. 381-395.
- [2] K.P. Subbu and N. Thomas, "Poster abstract: A location aware personalized smart control system," in *Proc. 13th International Symposium on Information Processing in Sensor Networks*, Berlin, 2014, pp. 309 – 310.
- [3] J. Wang, C. Zixue, L. Jing, O. Yota, and Y. Zhou, "A location-aware lifestyle improvement system to save energy in smart home," in *Proc.*

- 4th International Conference on Awareness Science and Technology, Seoul, Korea, 2012, pp. 109 – 114.
- [4] M.V. Moreno, J.L. Hernandez, and A.F. Skarmeta, “A New Location-Aware Authorization Mechanism for Indoor Environments,” in *Proc. 28th International Conference on Advanced Information Networking and Applications*, Victoria, 2014, pp. 791 – 796.
- [5] L. Mainetti, V. Mighali, and L. Patrono, “An IoT-based User-centric Ecosystem for Heterogeneous Smart Home Environments,” in *Proc. 2015 IEEE International Conference on Communications, IEEE ICC 2015*, London (UK), June 8-12, 2015.
- [6] L. Mainetti, V. Mighali, L. Patrono, “A Location-aware Architecture for Heterogeneous Building Automation Systems,” in *14th IFIP/IEEE Symposium on Integrated Network and Service Management, IM 2015*, Ottawa (Canada), May 11-15, 2015.
- [7] G. De Luca, P. Lillo, L. Mainetti, V. Mighali, L. Patrono, and I. Sergi, “The use of NFC and Android technologies to enable a KNX-based smart home,” in *Proc. 2013 International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2013*, 2013, Article number 6671904, pp.1-7.
- [8] M. Jung, J. Weidinger, C. Reinisch, W. Kastner, C. Crettaz, A. Olivieri, and Y. Bocchi, “A Transparent IPv6 Multi-protocol Gateway to Integrate Building Automation Systems in the Internet of Things,” in *Proc. 2012 IEEE Int. Conf. Green Computing and Communications*, Besancon, 2012, pp. 225 – 233.
- [9] M. Jung, C. Reinisch, and W. Kastner, “Integrating Building Automation Systems and IPv6 in the Internet of Things,” in *Proc. Sixth Int. Conf. Inn. Mobile and Internet Services in Ubiquitous Computing*, Palermo, 2012, pp. 683 – 688.
- [10] G. Bovet and J. Hennebert, “A Web-of-Things Gateway for KNX Networks,” in *Proc. European Conference on Smart Objects, Systems and Technologies*, Germany, 2013, pp. 1 –8.
- [11] A. Kuusik, S. Roche, and F. Weis, “Smartmuseum: Cultural content recommendation system for mobile users,” in *Proc. Fourth Int. Conference on Computer Sciences and Convergence Information Technology*, Tallinn, Estonia, 2009, pp. 477–482.
- [12] Y. Wang, C. Yang, S. Liu, R. Wang, X. Meng, “A RFID & handheld device-based museum guide system”, *2nd Int. Conference on Pervasive Computing and Applications (ICPCA07)*, pp. 308 - 313, July 2007.
- [13] L. Cavaglione, M. Coccoli, A. Grosso, “A framework for the delivery of contents in RFID-driven smart environments,” in *2011 IEEE International Conference on RFID-Technologies and Applications (RFID-TA)*, pp.45-49, 2011, doi: 10.1109/RFID-TA.2011.6068614.
- [14] University of Haifa, “Handheld ‘Smart’ Museum Guide Unveiled,” *ScienceDaily*, 21 March 2007.
- [15] F. Sparacino, “The museum wearable: real-time sensor-driven understanding of visitors’ interests for personalized visually-augmented museum experiences,” in *Museums and the Web*, 2002, pp. 17–20.
- [16] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” *Image and vision computing*, 2004, Vol. 22, no. 10, pp. 761-767.
- [17] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proc. of the seventh IEEE international conference on computer vision*. Kerkyra, 1999, Vol. 2., pp. 1150-1157.
- [18] H. Bay, T. Tuytelaars, and L. V. Gool. “Surf: Speeded up robust features,” in *Proc. of European Conference on Computer vision*, Springer, Berlin, Heidelberg, 2006, pp. 404-417.
- [19] E. Lau and W.Y. Chung, “Enhanced RSSI-Based Real-Time User Location Tracking System for Indoor and Outdoor Environments,” in *Proc. Enhanced RSSI-Based Real-Time User Location Tracking System for Indoor and Outdoor Environments Convergence Information Technology*, Gyeongju, 2007, pp. 1213 – 1218.
- [20] “Amazon S3”. Internet: <http://aws.amazon.com/it/s3/> [June, 2015].
- [21] “Facebook Graph APIs”. Internet: <https://developers.facebook.com/docs/graph-api/reference/> [June, 2015].
- [22] D. Hardt, “The OAuth 2.0 Authorization Framework”, RFC 6749, October 2012.
- [23] “Odroid-XU”. Internet: http://www.hardkernel.com/main/products/prdt_info.php [June, 2015].
- [24] “Raspberry Pi”. Internet: <http://www.raspberrypi.org> [June, 2015].
- [25] T. Kuflik, O. Sagy, J. Lanir, A. Wecker, and O. Mogilevsky, “A different kind of experience: Using a smart mobile guide for education and aging research at the Hecht museum,” in *Proc. of The annual conference of Museums and the Web*, Portland, OR, USA, 2013.

Stefano Alletto is currently a Phd student at University of Modena and Reggio Emilia. He received the Laurea degree in computer engineering in 2014 from the University of Modena and Reggio Emilia, Italy. His research interests include egocentric and embedded vision, focusing on social interactions and augmented cultural experiences. In 2015, working on these topics, he won the best paper award at the International Conference on Intelligent Technologies for Interactive Entertainment.

Rita Cucchiara received the Laurea degree in electronic engineering and the Ph.D. degree in computer engineering from the University of Bologna, Bologna, Italy, in 1989 and 1992, respectively. Since 2005, she is a full professor at the University of Modena and Reggio Emilia, Italy; she heads the Imagelab Laboratory and is the director of the Research Center in ICT SOFTECH-ICT. Her research interests include pattern recognition, computer vision and multimedia. She is the author of more than 200 papers in journals and international proceedings, and is a reviewer for several international journals. Since 2006 she is a fellow of IAPR. She is a member of the Editorial Boards of Multimedia Tools and Applications and Machine Vision and Applications journals and chair of several workshops and conferences.

Giuseppe Del Fiore graduated cum laude in Computer Engineering at University of Salento, Italy, in 2013. His thesis concerned the definition and validation of Cross-layer Approach to Minimize the Energy Consumption in Wireless Sensor Networks. Since November 2013 he collaborates with IDA Lab - Identification Automation Laboratory at the Department of Innovation Engineering, University of Salento. His research activities are mainly focused on the design and validation through test beds on real devices of innovative applications and protocols aimed to reduce power consumption in Wireless Sensor Networks. He is also involved in the study of new solutions for Internet of Things.

Luca Mainetti is an associate professor of software engineering and computer graphics at the University of Salento. His research interests include web design methodologies, notations and tools, services oriented architectures and IoT applications, and collaborative computer graphics. He is a scientific coordinator of the GSA Lab - Graphics and Software Architectures Lab, University of Salento.

Vincenzo Mighali received the "Laurea" Degree in Computer Engineering with honors at the University of Salento, Lecce, Italy, in 2012. Since January 2009 he collaborates with IDA Lab — Identification Automation Laboratory at the Department of Innovation Engineering, University of Salento. His activity is focused on the definition and implementation of new tracking system based on RFID technology and on the design and validation of innovative communication protocol aimed to reduce power consumption in Wireless Sensor Networks. He is also involved in the study of new solutions for building automation. He authored several papers on international journals and conferences.

Luigi Patrono received his MS in Computer Engineering from University of Lecce, Lecce, Italy, in 1999 and PhD in Innovative Materials and Technologies for Satellite Networks from ISUFI-University of Lecce, Lecce, Italy, in 2003. He is an Assistant Professor of Network Design at the University of Salento, Lecce, Italy. His research interests include RFID, EPCglobal, Internet of Things, Wireless Sensor Networks, and design and performance evaluation of protocols. He is Organizer Chair of the international Symposium on RFID Technologies and Internet of Things within the IEEE SoftCOM conference. He is the scientific coordinator of the IDA Lab - Identification Automation Lab at the Department of Innovation Engineering, University of Salento. He is author of about 90 scientific papers published on international journals and conferences and four chapters of books with international diffusion.

Giuseppe Serra is a senior researcher at the Imagelab, University of Modena and Reggio Emilia, Italy. He received the Laurea degree in computer engineering in 2006 and the Ph.D. degree in computer engineering, multimedia and telecommunication in 2010, both from the University of Florence. His research interests include image and video analysis, multimedia ontologies, image forensics, and multiple-view geometry. He has published more than 25 publications in scientific journals and international conferences. He has been awarded the best paper award by the ACM-SIGMM Workshop on Social Media in 2010.