A General-Purpose Sensing Floor Architecture for Human-Environment interaction

ROBERTO VEZZANI and MARTINO LOMBARDI and AUGUSTO PIERACCI and PAOLO SANTINELLI and RITA CUCCHIARA,
Softech-ICT, University of Modena and Reggio Emilia

Smart environments are now designed as natural interfaces to capture and understand human behavior without a need for explicit human-computer interaction. In this paper, we present a general-purpose architecture that acquires and understands human behaviors through a sensing floor. The pressure field generated by moving people is captured and analyzed. Specific actions and events are then detected by a low-level processing engine and sent to high-level interfaces providing different functions. The proposed architecture and sensors are modular, general-purpose, cheap, and suitable for both small- and large-area coverage. Some sample entertainment and virtual reality applications that we developed to test the platform are presented.

CCS Concepts:•**Human-centered computing** → **Interaction devices;**•**Computing methodologies** → *Activity recognition and understanding; Object detection;*

Additional Key Words and Phrases: Sensing Floors, Florimage, Human-Environment Interaction

ACM Reference Format:
R. Vezzani, M. Lombardi, A. Pieracci, P. Santinelli, R. Cucchiara. 2015. A General-Purpose Sensing Floor Architecture for Human-Environment interaction *ACM Trans. Interact. Intell. Syst.* 1, 1, Article 1 (March 2015), 25 pages.
DOI: 0000001.0000001

1. INTRODUCTION

In the last decades, research in multi-sensing environments and perceptual spaces to support new Human-Computer Interaction (HCI) methodologies has achieved impressive results.

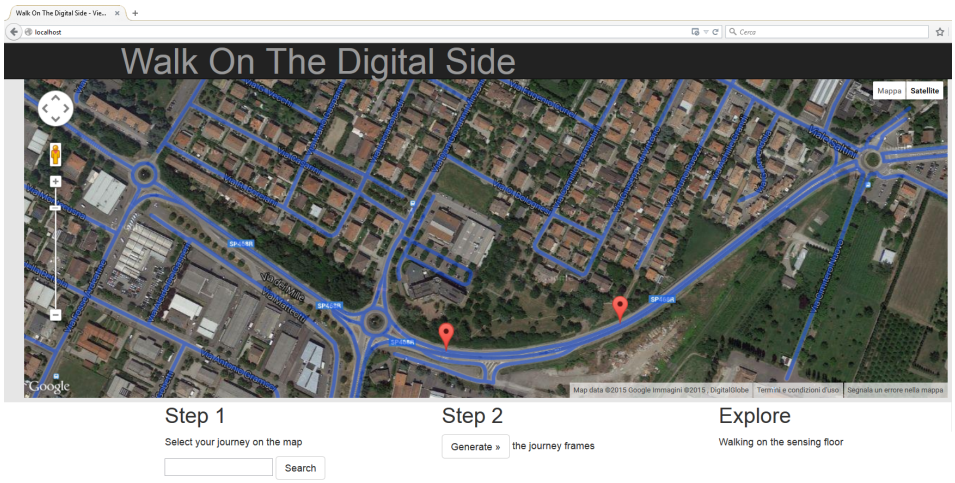
In addition to explicit HCI interfaces based on voice, vision and touchscreens, sensing environments are becoming an effective and feasible way to allow humans either to interact with virtual worlds. Applications of sensing spaces are suitable for many fields, such as intelligent building automation, e-health, surveillance, entertainment, edutainment (educational entertainment [Sakamura 1999]), and smart factories.

Sensing environments bypass the use of non-natural input devices such as keyboard and mouse, which force users to learn the computer interface paradigm. In addition, they can also avoid the need of semi-natural interfaces like touchscreens and make humans completely free to move and interact with virtual or physical items and with other humans.

In pursuit of the aim of providing hands-free natural interfaces, we propose an innovative sensing floor able to capture and measure the pressure field exerted by people or objects. Instead of optical images, the surface generates a "*floor image*", in which each "*pixel*" corresponds to a spatial portion of the floor and the "*pixel value*" is related to the pressure applied on top of it. Floor images are analyzed using traditional computer vision and pattern recognition techniques in order to detect people and their behaviors. The system allows a plethora of applications, ranging from entertainment to surveillance, from multimedia content access to medical rehabilitation.

Authors' address: Dipartimento di Ingegneria "Enzo Ferrari", University of Modena and Reggio Emilia, Via Vivarelli 10 41125, Modena - Italy; emails: {roberto.vezzani, martino.lombardi, augusto.pieracci, paolo.santinelli, rita.cucchiara}@unimore.it.

DOI: 0000001.0000001



(a)

(b)

Fig. 1: The Street View application: a walk on the sensing floor becomes a virtual journey. a) The journey selection and b) the virtual tour.

The idea of developing sensing floors is not new since first attempts are dated 1997 at MIT [Paradiso et al. 1997]; and a representative subset of them is described in Section 2. However, the majority of them presented only a specific prototype for specific applications. In addition, none of them fulfills all the following requirements, which led our development to the proposed solution:

— low cost for hardware devices: the cost of the sensing elements should be comparable to traditional floors to be affordable in many contexts;
— high scalability: the sensors should be integrated into a hierarchical network, in order to allow the coverage of narrow rooms as well as of wide areas;

— high reliability and robustness: breakable and fragile elements should be avoided or limited to protected packages;
— temporal and spatial resolutions: they should be high enough to allow people detection and tracking, even in presence of multiple targets, or modular depending on the applications;
— non-invasive and invisible devices: for design issues, the sensing layer must be invisible to the users and the floor should appear similar to traditional floors to avoid the "observer effect".

The last requirement is one of the key aspects of the sensing floor technology. For instance, RGB-D sensors (e.g., Microsoft Kinect) or stereo cameras allow a more precise and reliable people detection and localization at a considerably lower cost. However, they suffer from typical occlusion problems as well as from privacy related issues. The formers can be partially solved using networks of sensors, while the user's perception of privacy awareness is always limited when visual sensors are installed. On the contrary, other low-cost solutions such as PIR and proximity sensors do not guarantee the same spatial resolution of the proposed system.

Preliminary descriptions of the sensing floor have been presented in [Lombardi et al. 2013a] and [Lombardi et al. 2013b]. In addition to providing more details and implementation remarks, here we have (1) correctly formalized the acquisition model, (2) described and tested the low level processing steps, (3) defined and test high level applications such as the virtual walk.

The paper is structured as follows: in the subsequent Section some related works of other experiences and prototypes with touch and sensing surfaces will be described,highlighting similarities and differences. In Section 3, we will provide an overview of the general-purpose hardware and software architecture, giving details on the capture elements and the communication model. The input device and the low-level processing system will be described in Section 4 and Section 5, respectively. In Section 6, we will propose some developed applications presenting, in particular, a virtual journey based on Google Street View[Anguelov et al. 2010] and the proposed sensing floor. Section 7 will deal with an experimental evaluation of all the system components, from the physical sensor to the high level application. Finally, Section 8 will summarize the main novelties, limitations and some future improvements of the system.

2. RELATED WORK

Some prototypes of sensing floors for human-based action detection and identification have been designed in the past. Different physical characteristics have been adopted; among others, the measure of pressure and the proximity effects related with the electrical properties of human body are the most widely exploited features. In the following paragraphs, we will provide a brief overview of previously developed prototypes, focusing on their differences and their capabilities in terms of cost, scalability, modularity, and generality of the architecture. A summary is provided in the Table I.

Lee Middleton *et al.* [Middleton et al. 2005] developed a floor sensor mat specifically conceived for gait recognition. The sensor is made of a set of binary switches and uses a simple design inspired by computer keyboards. Each switch is made by separating a pair of wires with a deformable foam material. Wires come into contact and the switch is closed when a pressure is applied. These switches are disposed on a grid and their state is read using the same approach of computer keyboards. The design is simple and easily scalable, even if the adopted switches are binary and they do not provide a response commensurate with the pressure.

Subsequently, Chen-Rong Yu *et al.* [Yu et al. 2006] proposed a localization system to accurately estimate human's positions. It performs single person and multiple people tracking in a home environment. The Condensation algorithm is exploited to locate residents' position via multi-camera and sensory floor approaches. This work, based on a sensor floor and four video cameras, is almost completely focused on the proposed data processing techniques which are thoroughly explained. Conversely, no details concerning the nature of the forty pressure cells employed in the sensing floor are provided, neither on the architecture of the system used to collect the sensor data, nor on the size of the sensing area.

Table I: Sensing Floors.

Authors	Application	Technology	Low cost	Scalability	Invisibility	Reliability	Temporal and Spatial resolution
[Middleton et al. 2005]	gate recognition	custom switch metrics	y	n	n	n	y
[Shen and Shin 2009]	human localization / multi-cameras & floor	off-the-shelf load cell / cameras	n	n	n	n	n
[Savio and Ludwig 2007]	human tracking / gait	custom capacitive / electronics textile based	n	y	n	n	y
[Valtonen et al. 2009]	unobtrusive 2D human positioning	custom capacitive	n	n	y	y	n
[Srinivasan et al. 2005]	Interactive media application / human dance movement	off-the-shelf FSR sensor based mat	n	n	n	y	y
[Anlauff et al. 2010]	spatially resolved force sensing floor surface	custom "Paper FSR"	y	n	n	y	y
[Rajaling-ham et al. 2010]	tracking	off-the-shelf FSR	n	n	y	n	y
[Vera-Rodriguez et al. 2013]	biometrics	piezoelectric	n	n	n	y	y

In their study, Domnic Savio *et al.* [Savio and Ludwig 2007] described three processing techniques to track the gait of human walk. They developed a smart carpet that can be laid on the floor. The sensor set forms a self-organizing sensor network. To identify the footstep, clustering algorithms based on Maximum Likelihood Estimate and Rank Regression have been applied. The proposed approach is scalable and commercially viable even if the binary nature of the embedded nodes does not provide a response commensurate with the applied weight. The authors developed a 240 cm by 200 cm smart carpet composed of 180 interconnected capacitive nodes forming the sensor network. Each node is equipped with a 16 bit micro-controller. This solution is very innovative and smart though, at the same time, very expensive. Besides, it does not seems to be not reliable and durable enough, due to its complexity and fragility and to the lack of a protective layer, being directly exposed to the human activities. As well the invisibility requirement is not fulfilled.

A further step was made by the research of Miika Valtonen *et al.* [Valtonen et al. 2009] in which an unobtrusive two-dimensional human positioning and tracking system based on a low-frequency electric field was described. The capacitance between multiple floor tiles and a receiving electrode was measured. Their method was based on the fact that the human body is able to conduct a low-frequency signal. The authors did not handle, however, problems related to the large scale deployment, nor to the scalability of the hardware architecture. The proposed method, in fact, was based on the capacitive floor and did not provide any information related to the human weight. The lack of weight information represents a drawback to perform human action detection blocking any possibility to detect non-conductive objects, regardless of the weight.

A portable high-resolution sensing floor prototype was proposed by Prashant Srinivasan *et al.* [Srinivasan et al. 2005]. The sensor the developed measures the pressure field generated by people

walking on it. The system consisted of several sensor mats, each one is composed of a 42x48 grid of pressure sensors with size of 48.8 cm x 42.7 cm. The system was entirely built employing off-the-shelf components. The sensor elements of the mat were made using a pressure sensitive polymer between conductive tracks on sheets of Mylar and they changed the resistance with the applied pressure. On the one hand the proposed architecture was suitable for research-oriented application but, on the other hand, it did not show the required scalability and it was not cheap enough to allow the coverage of wide areas. A high number of expensive components were used for each mat,as well as a multi-sensor element (model-5315 by Tekscan [1]) and a Rabbit Ethernet-enabled controller.

Afterwards, Jan Anlauff *et al.* [Anlauff et al. 2010] presented a prototype of a floor surface based on sensing elements made out of conductive black art paper and grouped into modules forming a grid of resistors able to measure quasi-static forces. The suggested approach was a low cost alternative for spatially resolved tactile sensing. However the employed signal conditioning system was based on the matrix arrangements of the sensing elements thus the proposed solution suffered from mutual interference between different sensors in the matrix. The authors don't provide a detailed description of the scalability of the communication infrastructure; moreover, the invisibility and the physical implementation requirements were not considered.

A probabilistic approach to the tracking and estimation of the lower body posture was presented in the work by Rishi Rajalingham *et al.* [Rajalingham et al. 2010]. Their sensing floor had limited sizes, it was not easily scalable and commercially viable. It employed off the shelf resistive force sensors and consisted of a 66 array of rigid tiles, 30 cm on each side. Tiles were equipped with four resistive force sensors, which were located at the corners. An array of six small-form-factor computers was used for data processing. The studied solution was focused on data processing and the exploited sensing floor did not seem to be feasible for creating wide area implementations. Several reasons can be mentioned, *i.e.* the employed sensing elements are too expensive (several USD per piece), unfeasible for large scale applications; there is lack of details regarding modularity, scalability and physical implementation.

Finally, Ruben Vera-Rodriguez *et al.* [Vera-Rodriguez et al. 2013] presented a work focused on footstep signals as a biometric with a comparative analysis and fusion of spatio-temporal information of the signals for person recognition. The described experiments were carried out on a footstep database. The contribution of this work is the assessment of footsteps in time, in space, and in a combination of the two. The performance obtained for the two domains showed error rates in the range of 5-15% for each domain, and in the range of 2.5-10% for their fusion. The sensor arrangement used by the authors is based on two (45 x 30 cm) sensor mats. Each mat is equipped with 88 piezoelectric sensors and the sampling frequency for each sensor is quite high, 1.6 kHz. Piezoelectric sensors need more sophisticated conditioning hardware compared with the sensors we proposed, they are more expensive, weaker, and they need high sampling rates to exploit the dynamic nature of their electrical response. The work does not provide any details related with the physical architecture of the presented system and its scalability.

As shown in Table I, the majority of the proposals previously introduced produced very interesting research prototypes, which could be integrated in a multi-sensing environment. However, none of them represents a good solution for covering large areas with a scalable and low cost technology.

3. SYSTEM ARCHITECTURE

In this paper, we propose a complete and general purpose architecture for Human-Environment interaction based on smart floors. Therefore, network communication and processing components are handled in addition to the specific hardware device.

Figure 2 shows a schema of the whole system architecture, in which the three main components are highlighted, *i.e.* the *input device*, the *low-level processing unit* and the pool of *high-level applications*.

[1] http://www.tekscan.com/5315-pressure-sensor

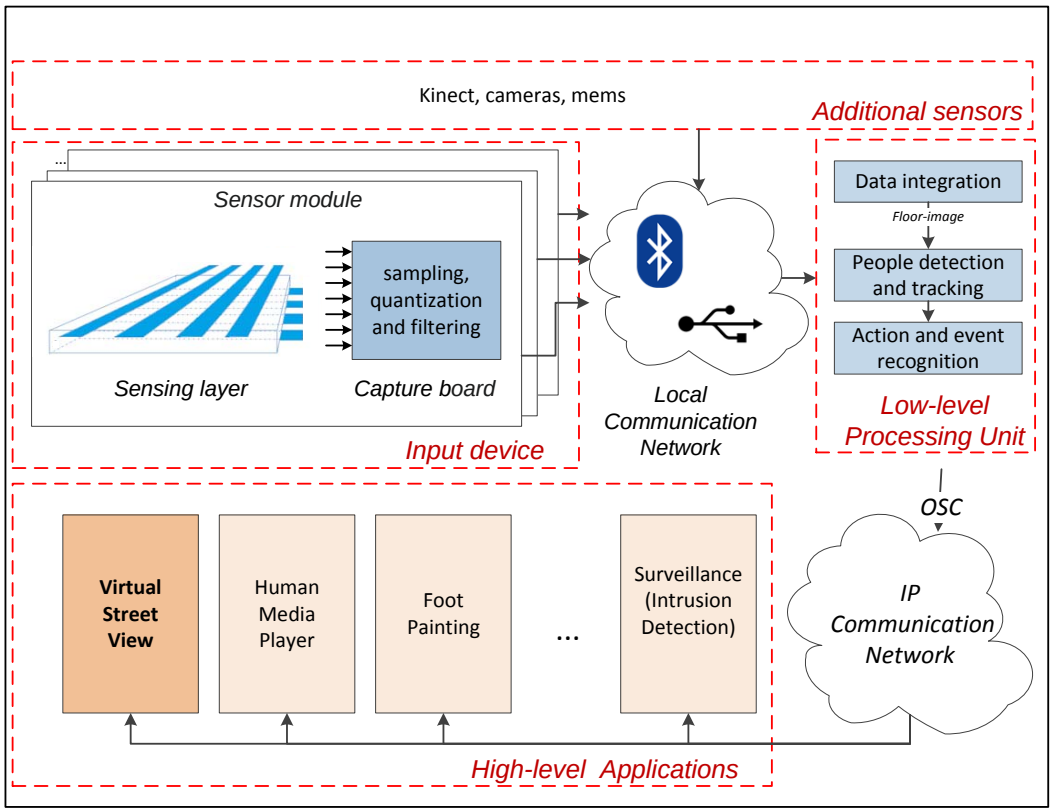


Fig. 2: The main components of the floor image architecture: the capture device, the low-level processing unit and the client applications.

The **input device** is obtained arranging a set of *sensor modules*. The modules can be freely disposed on the floor to cover areas with different shape and size. However, a grid distribution of the modules leads to an easier integration and processing of the data. Each module is a standalone sensor device and is composed of a fixed-size rectangular *sensing layer* and a *capture board*. The former is a passive sensor, whereas the latter is able to acquire, digitize and transmit the sensed values. A unique identifier has been assigned to each capture board.

The integration and combination of data coming from different sensor modules is the first task assigned to the **low-level processing unit**. The capture board identifier is exploited to this aim. As a result, the data integration step generates a sequence of floor images, which are subsequently processed to detect and track people on the floor. In addition, people actions and other important events can be recognized by means of specific software classifiers included in the low-level processing module.

Finally, a pool of **high-level applications** are connected to the low-level processing server, which delivers the required events according to the application needs and goals.

The framework includes two different communication networks. Sensor data are sent from the input device to the low-level processing unit using a serial protocol over private point-to-point Bluetooth or USB links. Additional sensors (e.g., Microsoft Kinects, cameras, or mems) can be connected to the low-level processing unit through this local network. On the other hand, the high-level applications receive commands and events by means of a public IP network.

Each of the framework components will be detailed in the following specific Sections.

Table II: Mechanical and electrical characteristics of the resistive polymer.

Characteristic	Min	Max	Unit
Thickness	2.5	3.0	mm
Compressive strength at 10 % of relative deformation	15	25	kPa
Compressive deformation at 100 kg for 60x60 cm tile	6	10	%
Temperature range	- 20	+80	^{o}C
Surface Resistance	1.0×10^5	1.0×10^{10}	Ω/sq
Volume Resistivity	1.0×10^4	1.0×10^{10}	$\Omega \times cm$

4. THE INPUT DEVICE

The *input device* is the first and most important part of the architecture and it represents the main contribution of this work. As before mentioned, each sensor module is composed of a sensing pad and a capture board. The module has a rectangular shape and is 2x1 meters large. Superficially, the module is covered with a commercial paving technology, called *ceramic floating floor*. In particular, we use the SLIM4 tiles produced by the Italian company Florim Ceramiche SPA[2]. A commercial floating floor does not need to be nailed or glued to the sublayer and, thus, it might be constructed over a sub-floor or even over an existing floor. It consists of a polymeric layer holding up the tiles. Therefore, every tile can move perpendicularly to the floor plane and independently from its neighbors, so it can transmit pressure to the sublayer due to the presence of weight on it. The VELCRO® attachment system is employed to keep the tiles connected to the floor.

4.1. The Sensing layer

A layer of conductive polymer obtained as a foam of Polyethylene and Carbon substitutes the original electrically insulating material. The layer is an Electrostatic discharge (ESD) polyethylene extrusion with anti-static properties, medium density (32 Kg/m^3) and closed-cell foam. The obtained element is durable, lightweight, flexible, and solid. It meets the EIA 541 requirements [EIA 1988] for static decay and surface resistance (see Fig. 3(a)). Many types of conductive polymeric material with different resistivity, elasticity and temperature range are commercially available. The requirements imposed by the application are specified in Table II. Among others, the ESD roll shown in Figure 3(a) can be easily supplied at affordable costs.

The anti-static characteristic is achieved through conductive chemical additives. These additives are usually incorporated into the foam during the manufacturing process. Such a technique guarantees an even distribution of the conductive elements throughout the block, which is mandatory to have a uniform electrical resistance property.

A sequence of parallel aluminum stripes has been unrolled on the two sides of the conductive polymer, lengthwise on the top and crosswise on the bottom of the layer (see Fig. 3(b) and 3(c)). As a result, a *sensing element* $S_{r,c}$ is created at each intersection between a lengthwise and a crosswise stripe.

When a pressure is applied on the top of the tiles, the rough surface of the conductive polymer is compressed onto the electrodes surface [Weiss and Worn 2005]. The electric resistance between the two stripes is related to their physical compression rate on the intermediate polymeric layer. In particular, the contacting area between polymer and electrodes is increased whilst the resistance between them is proportionally reduced. In other words, the sensor converts the applied pressure to an electric resistance, which is measured with the capture board described in Section 4.2.

The stripes cannot be glued to the conductive element, otherwise the contact resistance will not change during the working conditions. Therefore, two thin polyethylene sheets have been included in the sensor stack as a support for the adhesive aluminum stripes (Fig. 3(b)). The whole sensing module turns out to be composed of four layers: a polyethylene sheet with the horizontal conductive stripes on the bottom, the conductive polymer, another polyethylene sheet with the vertical conductive stripes from one side and the Velcro on the other side, and ceramic tiles on the top.

[2]http://www.florim.it

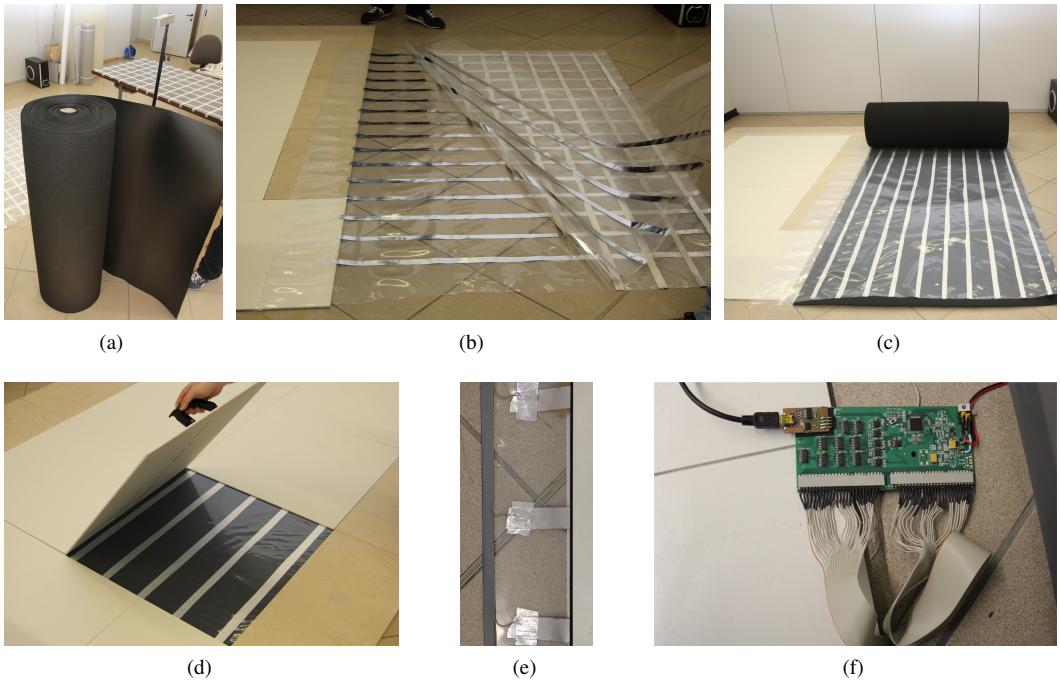


Fig. 3: Hardware components of the sensing floor prototype. (a) a roll of the conductive polymer, (b) the aluminum stripes glued on polyethylene sheets, (c) the layer overlapping, (d) the ceramic floating tiles placed on the top of the sensors, (e) the external wires from the sensor to the board, and (f) the capture board.

The spatial resolution of the sensor is defined by the distance between two consecutive stripes. A step of 125 mm has been adopted in the developed prototype, so that each rectangular module contains 8x16 sensing elements. This size is a reasonable tradeoff between a sufficient spatial resolution (to detect each foot of a walking person) and the sensor cost.

The connections of the stripes to the capture board have been provided by ribbon cables manually soldered to the stripe endings (see Fig. 3(e)). Connection obtained with lines of conductive painting is expected to be used in the future commercial version of the sensors.

4.2. The capture board

Each sensor module is equipped with a custom capture board (Fig. 3(f)), which measures the electric resistance of each sensing element $S_{r,c}$ (cross between a vertical and horizontal aluminum stripe). In addition, the board digitizes and transmits the values $\psi_{r,c}$ to the processing unit.

The board is equipped with a 32-bit ARM micro controller (STM32F103RB [3]). The signal conditioning is done using a custom circuit composed of an array of trans-resistance amplifiers. The amplifier outputs a voltage response V_o proportional to the conductance G of the corresponding sensing element. The base configuration of each cell is reported in Figure 4. The sensing element is depicted as a resistance R_S. Due to the inverting configuration, a constant voltage of $V_{cc}/2$ is applied to the sensing element, inducing a current I_s through it:

[3]http://www.st.com

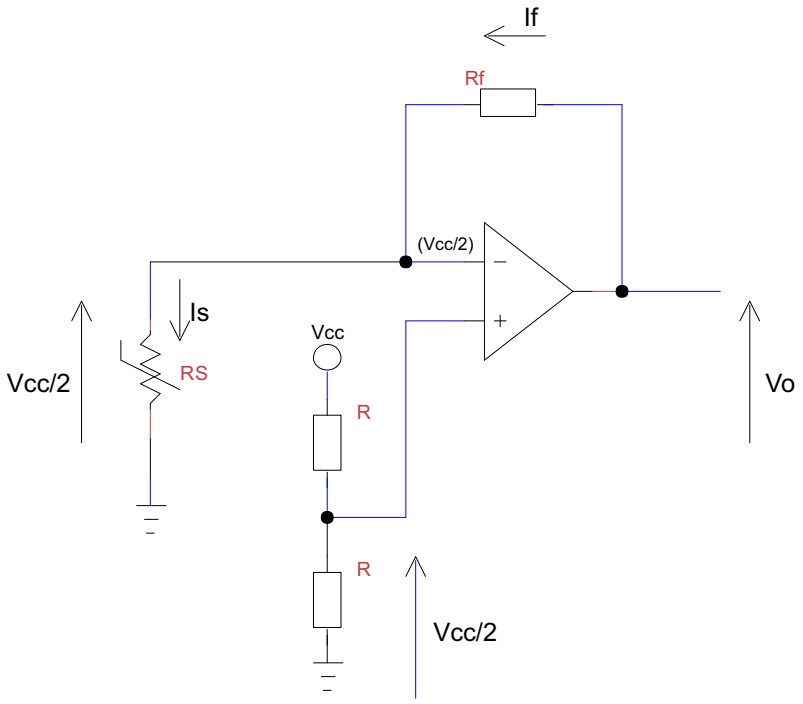


Fig. 4: Base configuration of the signal conditioning circuit.

$$I_s = \frac{V_{cc}/2}{R_S} \tag{1}$$

Since the electric currents I_f and I_s must have the same intensity, the output voltage Vo becomes:

$$Vo = V_{cc}/2 \cdot R_f \cdot G, \tag{2}$$

with Vo ranging from $V_{cc}/2$ (when $R_S >> R_f$) to V_{cc} (when $R_S \leq R_f$). Reference resistances R_f in the range of $[20 \ldots 100]$ $K\Omega$ have been exploited in our experiments.

Thanks to the grid distribution of the sensing elements, the overall configuration of the amplifier array follows a matrix layout, as shown in Figure 5. An inverting amplifier has been connected to each column. Selecting one row at a time, the amplifier senses the electrical conductance of the cell placed in the cross between the selected row and column.

The row selection is provided by an array of digital switches. If the row is connected to ground, the corresponding cell is exposed to a positive voltage and its behavior is the same of the schema in Figure 4. On the other hand, the other rows are connected to $V_{cc}/2$, and thus all the corresponding cells are subjected to a zero voltage. This solution has a good reliability against crosstalk effects [Shimojo et al. 1991], reduces the number of components and requires only a single supply.

The output pins of the amplifiers are connected to the Analog-to-Digital Converters (ADC) of the micro controller, which also triggers the analog switches of the row selection.

Each captured value is encoded using 16 bit unsigned integers. The device is able to sample each value in about 10μs, allowing a scan of the whole sensor matrix and the corresponding data encoding and transmission at up to 20 frames per second. Two operational modes have been implemented, *i.e.* the single scan and the streaming mode. In the first case, a single capture is done and sent for each

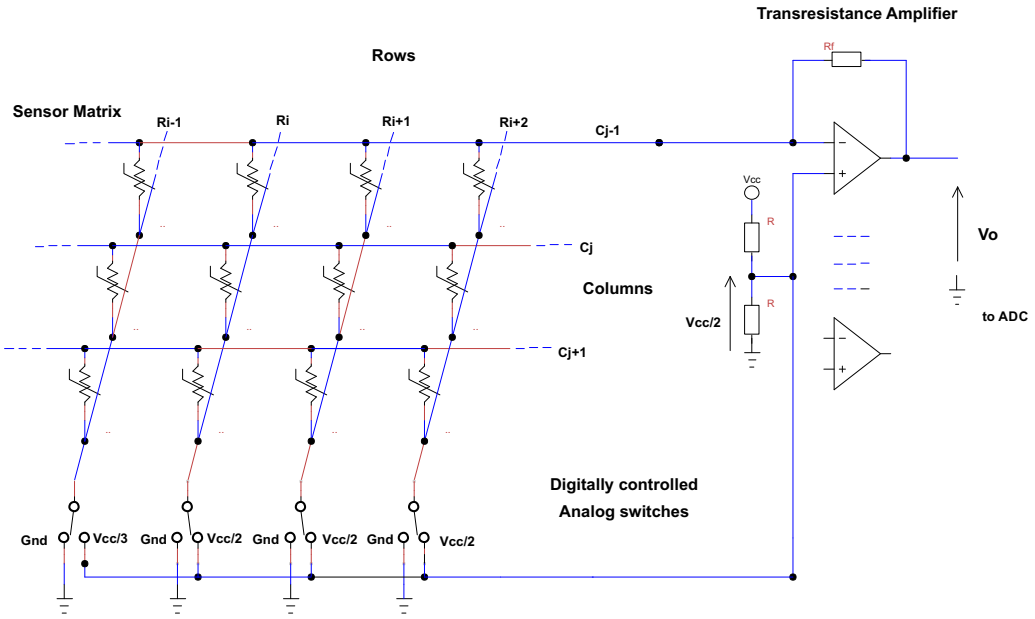


Fig. 5: Matrix version of the signal conditioning circuit.

request, whereas in the second one the sensor data are continuously captured and sent as a stream, until a stop command is received.

4.3. The Ceramic tiles

The sensitive area of the described stack does not cover the overall surface, but is limited to the regions of intersection between vertical and horizontal aluminum stripes. Therefore, only objects or people located within these regions can be detected. This drawback is solved thanks to the ceramic tiles, which are placed on the top of the sensors.

First, the tiles have been included in the system to obtain a complete and stylish floor solution, which should fulfill the requirements described in Section 1. In addition to aesthetic and design reasons, the ceramic tiles play an important role in the sensor device, since they act as a blurring filter. Ideally, the pressure applied on a specific point is distributed on the entire floor, with an effect inversely proportional to the distance. As a consequence, the value $\psi(r, c)$ acquired by the sensor $S_{r,c}$ is related to all the pressure stimuli applied to the floor.

The relation between the pressure field applied on the floor and the sensor response can be modeled with a kernel convolution:

$$\psi(r, c) = K_{r,c} \cdot \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_0 + x, y_0 + y) \cdot w_1(x, y) \;\; dxdy \tag{3}$$

where (x_0, y_0) is the spatial location of the sensor $S_{r,c}$ on the floor, $f(\cdot, \cdot)$ is the pressure field orthogonal with the floor and $w_1(\cdot, \cdot)$ is a weighting function.

The kernel term $w_1(x, y)$ depends on the distance between the sensor and the object/person above the floor. The normalization constant K is related to several building factors and, thus, its value should be calibrated or learnt for each specific sensor $S_{r,c}$.

The response of the sensing layer (without the ceramic tiles) can be obtained using the weighting kernel $w_1(x, y)$ of Equation 4:

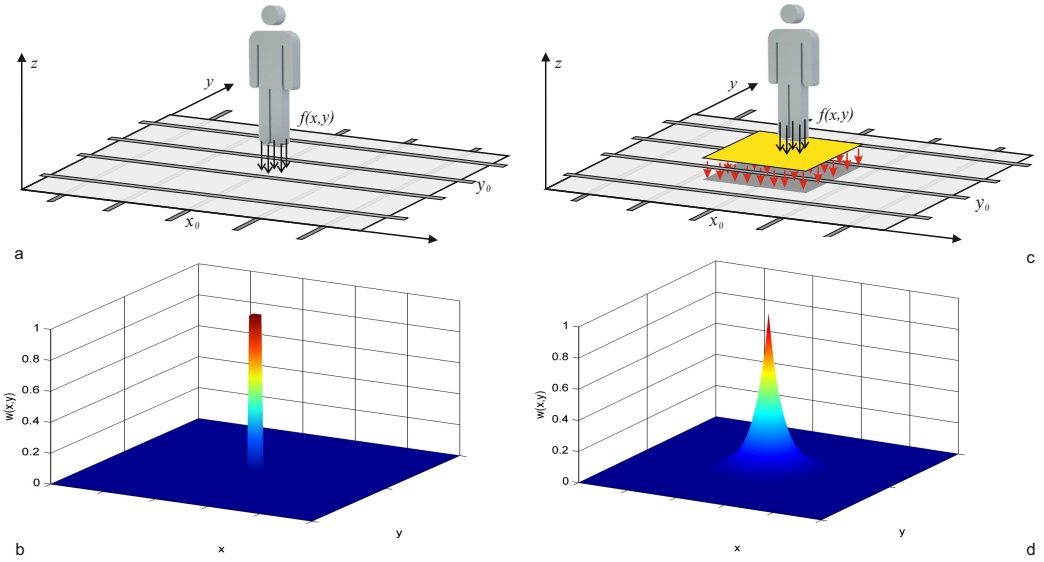


Fig. 6: Measuring the pressure field. a) the pressure field $f(x, y)$ generated by a person on the floor is captured by the sensor only when placed on the top of it. c) Using the ceramic tiles, the pressure is diffused over the whole tile extent. b) and d) shown the weighting functions without and with the tiles respectively.

$$w_1(x, y) = \begin{cases} 1 & |x| < l \wedge |y| < l \\ 0 & otherwise \end{cases} \tag{4}$$

where l is the width of the aluminum stripes.

The sensing layer without the ceramic layer can be effectively used only with a very high spatial frequency of the stripes. This principle is similar to the one used in several previously developed carpets, where discrete arrays of FSR sensors are placed for detecting pressure. Since they are sensible only in the exact position where the pressure is held, many sensors must be used to avoid insensitive zones. As a result, the system becomes too expensive for covering large areas.

Pairing the sensing layer with very slim ceramic tiles, the response of the sensor with respect to the object distance follows a Gaussian law. The corresponding kernel $w_2(x, y)$ is reported in Equation 5:

$$w_2(x, y) = \exp\left[-\left(\frac{(x-u)^2}{\sigma_x^2} + \frac{(y-v)^2}{\sigma_y^2}\right)\right] \tag{5}$$

Figure 6 shows a visual explanation of the blurring effect provided by the ceramic layer.

4.4. The Local Communication Network

The capture board described in Section 4.2 is equipped with two serial ports for data communication, namely a main port and an auxiliary port. RS422 or RS232 serial standards are supported by both of them. The role of the main port is to provide a direct communication channel to the low-level processing unit, whereas the auxiliary port can be used to establish a chain of neighbor capture boards. In this case, each board receives data packets from the auxiliary port and forwards them to the next board of the chain or to the processing unit, through the main port. The data packet contains

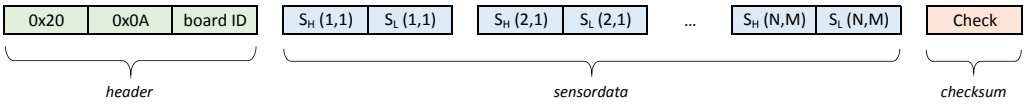


Fig. 7: Data packet format.

a header, which also embeds the unique identifier of the board, a body with the sensor data, and a checksum byte at the end. Sensor data are arranged by rows. The data packet format is depicted in Figure 7, where $S_H(i,j)$ and $S_L(i,j)$ are the most and the least significant byte from the sensor $S_{i,j}$, respectively.

The main port is also equipped with a serial-to-USB adapter to provide an easier connection with the low level processing unit. A serial-to-Bluetooth adapter is also available if a wireless connection is required. However, scalability and interference issues may arise with a high number of modules [Schurmann et al. 2011]. In this case, a hierarchical communication network is also required [Lombardi et al. 2013b].

The wired connection is preferable since it is more stable and, together with data cables, power lines can be installed toward the board as well. At the same time, a wireless connection drastically reduces the network and installation costs, but a battery power supply should be used.

The adopted serial protocol does not exploit compression algorithms to reduce the computational load on the capture board. The following analysis of *bandwidth requirements* confirms the scalability of the system even without data compression. As reported in Section 4.1, a sensor module contains 128 sensors and covers a region of two square meters. The corresponding data packet has a fixed size of 260 Bytes. We have estimated the required bandwidth in three different hypothetical setups, *i.e.* a single module, a sensing floor for a typical room of 16 square meters, and a very large area of 10.000 square meters. Three different frame rates have been also considered. Table III summarizes the results. Even in the worst case, the bandwidth is lower than 30 MiB/s, which is more or less the real limit of a single USB 2.0 link. Therefore, the implementation of compression algorithms is not required. In addition, the total number of sensors in the very large room is comparable with the number of pixels of a video camera. The further processing algorithms should fulfill similar constraints imposed on real time surveillance systems. As a consequence, it is reasonable to assume that a single processing unit supports a whole sensing floor in most of the real setups.

5. THE LOW LEVEL PROCESSING UNIT

The low-level processing unit collects input data coming from different sensor modules, filters the noise, detects and tracks moving people, and finally recognizes events. Then, pre-processed raw data as well as detected events are made available to the pool of high-level applications.

The overall processing sequence has been split into three steps: the generation of floor images from the set of sensor values, the detection and tracking of people, and the further recognition of events, people actions and behaviors. The last two steps are provided using common image and

Table III: Bandwidth estimation in three different hypothetical setups

	Single module	*Typical room*	*Large area*
Area [m^2]	2	16	10,000
N. modules	1	8	5,000
Sensors	128	1,024	640,000
Bytes for frame	260	2,080	1,300,000
Bandwidth @ 5fps [B/s]	1,300	10,400	$6.5 \cdot 10^6$
Bandwidth @ 10fps [B/s]	2,600	20,800	$13 \cdot 10^6$
Bandwidth @ 20fps [B/s]	5,200	41,600	$26 \cdot 10^6$

video processing techniques opportunely adapted and applied to floor images. A conceptual schema of the overall framework is reported in Figure 8, while each step is detailed in the next subsections.

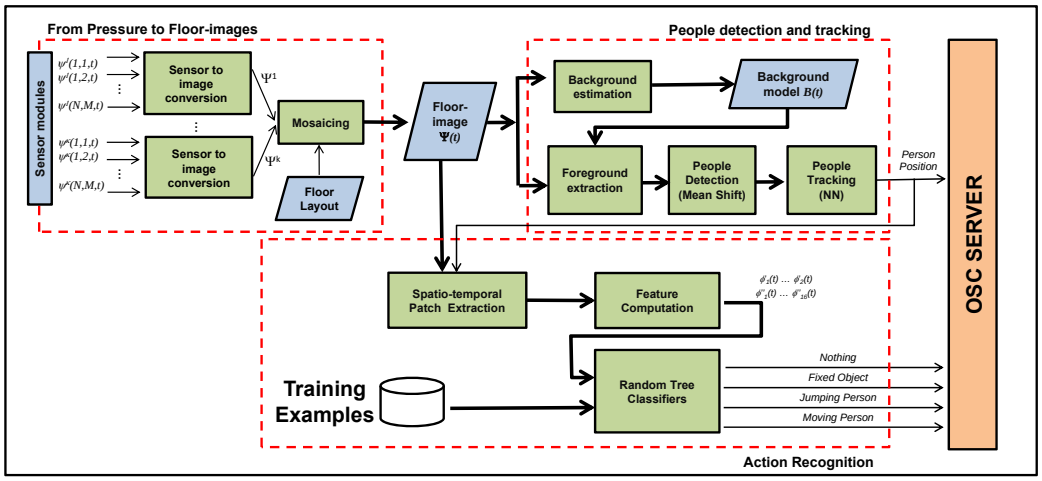


Fig. 8: Schema of the low level processing engine.

5.1. From pressure to floor images

The generation of the so called *floor images* is the first task of the processing unit. Thanks to the regular distribution of the sensing elements on a grid, a floor image Ψ is represented as a common dense image, where each pixel $\psi(i,j)$ corresponds to a sensor state and the spatial neighborhood relations are preserved.

If the system is composed by multiple sensor modules, the corresponding sub-images Ψ^k are merged as tiles of a mosaic. A floor layout configuration is needed to correctly map each Ψ^k — corresponding to the k-th sensor module — to the global floor image Ψ. Consecutive temporal samples $\Psi(t)$ are handled as frame sequences and generate a sort of floor-video.

As an example, the prototype system developed and assembled in our lab is shown in Figure 9. The input device is composed of two sensor modules. The corresponding sub-images are mosaiced together to obtain the whole floor image $\Psi(t)$ as in Figure 9(e), where the two 8x16 pixel matrices are highlighted.

Let $\Psi(t) = \{\psi(i,j,t)\}, i \in [1,W], j \in [1,H]$ be the floor image available at time t. Let (W,H) be the floor image size. The pixel value $\psi(i,j,t)$ is proportional to the corresponding value $sensordata[r,c]$ digitized by the capture board and scaled to fit the range [0-255] of 8-bit gray-level images.

5.2. People detection and tracking

Floor image pixel values include contributions due to deadweight, tile weight and noise, in addition to the pressure applied by an object or a person. Moreover, the presence of background objects in the scene (e.g., a chair or a bag left on the floor) makes the sensor calibration unfeasible. Thus, floor images are firstly pre-processed in order to filter out those undesired contributions. The operation is akin to the background subtraction task included in computer vision systems for people detection with fixed-cameras [Piccardi 2004; Hassanpour et al. 2011].

The foreground image $F = \{f(i,j,t)\}$ is computed as follows:

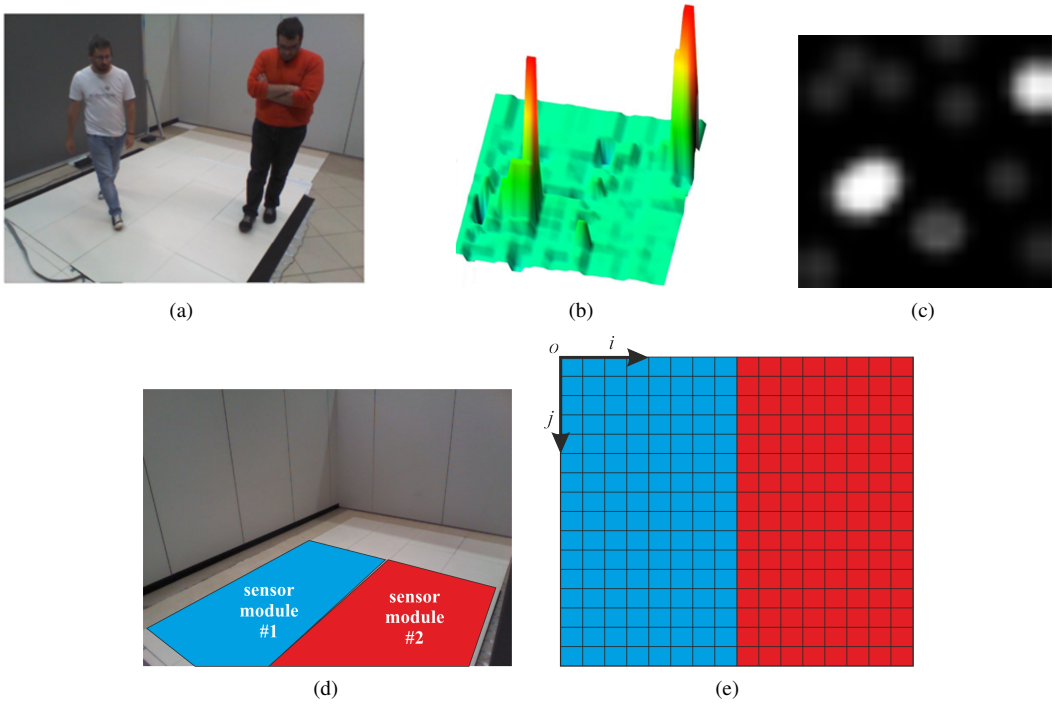


Fig. 9: Prototype sensing floor used for the experimental evaluation. (a) Picture of the floor with two standing people, (b) graph of the sensor values, (c) the integrated floor image obtained mosaicing the two sensor modules placed as in (d) and (e).

$$f(i,j,t) = \begin{cases} \psi(i,j,t) - b(i,j,t) & \text{if } (\psi(i,j,t) - b(i,j,t)) > Th \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $b(i,j,t)$ are the pixels of the background model B available at time t (floor-background) and Th is a fixed threshold.

In the following experiments, we have adopted a statistical background model with a very low updating rate. Each background pixel $b(i,j,t)$ is estimated as the mean value of N_b samples. A uniform sampling within a sliding window is performed to speed-up the background estimation step and to reduce storage requirements, as defined in Equation 7.

$$b(i,j,t) = \text{mean}\left\{\psi(i,j,t-\Delta t), \ldots, \psi(i,j,t-N_b\Delta t)\right\} \tag{7}$$

where N_b is the number of samples used and Δt is the temporal interval between two consecutive samples. In the following experiments we have set $N_b = 10$ and $\Delta t = 30$ s.

The spatial resolution of sensing elements is set to 125 mm, as described in Section 4. It is reasonable to presume that at least one sensor has a non-zero response when a person is walking on the floor. On the contrary, each person usually stimulates more than one sensor for each foot. The non-zero values of the foreground image are then clustered using a Gaussian mean shift [Comaniciu and Meer 2002], starting from all the local maxima as seeds.

The variance parameter σ^2 of the kernel has been set to 50^2 cm^2, which is large enough to cover a typical human step and to consider all possible data contributions due to both the feet of a walking person (see Figure 10). The selected σ value allows to filter out noisy peaks within the Kernel

radius. If two or more people are too close each other (the inter-person distance $\overline{P_A P_B}$ of Figure 10 is comparable with the σ value), the mean-shift algorithm will generate a single detection for the whole group.

The association of detections to people (i.e., the tracking algorithm) is based on positions only. Given the detections D_i at frame t and the current set of tracks T_j, we first compute the Euclidean distance matrix $M(i,j) = D_2(D_i, T_j)$. The detection to track association is provided using the schema proposed in [Rangarajan and Shah 1991]. For each frame, some detections may be assigned to tracks, while other detections and tracks may remain unassigned. The assigned tracks are updated using the corresponding detections. The unassigned tracks are marked invisible. Finally, unassigned detections begin new tracks. Each track keeps count of the number of consecutive frames where it remained unassigned. If the count exceeds a specified threshold, the tracking algorithm assumes that the object left the floor and it deletes the track.

Since the applications described in this paper are not required to work in very crowded situations, the implemented tracking algorithm handles neither groups (i.e., people closer to each other than a foot step) nor abrupt position changes (e.g., people leaping around). The reader can refer to [Smeulder et al. 2014] for more complex tracking schemes, if required by the application.

5.3. Action and event recognition

The recognition of people actions and behaviors using video cameras has been deeply addressed in the past [Poppe 2010]. A very common solution is based on the classification of spatio-temporal descriptors (STD) [Laptev 2005; Gorelick et al. 2007]. We applied the same approach to floor images. In video analysis, spatio-temporal patches can be obtained by stacking a set of rectangular sub-images of interest, which have been cropped from each frame. Similarly, the selection of the rectangular spatial regions on floor images may follow two different approaches, called *region-based* and *person-based* in the following lines. In the region-based approach, floor images are always cropped on the same fixed region, such as the square area covered by a tile. In this case, the region is manually selected independently from its content. As a consequence, the patch may contain a person, an object or even nothing. If a person or an object is located within the region, its position with respect to the patch center is not fixed. This first method does not require people detection and tracking.

In the person-based approach, instead, the region of interest follows and is centered around the estimated position of a detected and tracked person. Differently from the previous case, the person

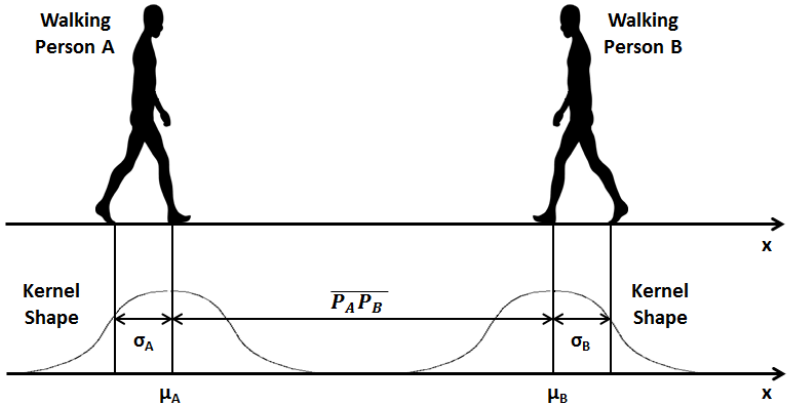


Fig. 10: The Gaussian Kernel used in the mean shift clustering of people positions. The inter-person distance $\overline{P_A P_B}$ between two users should be greater to the σ kernel parameter in order to be correctly distinguished.

is always in the center of the patch. This assumption leads to the selection of different feature sets and to the recognition of different types of action. We have addressed the region-based action recognition approach in the previous work [Lombardi et al. 2013b], while in this section we describe the person-based method.

Generic people actions depend both on the spatial movements of the person over the floor and on the type of interaction with the floor itself. For example, a person walk can be recognized using the trajectory on the floor, while a jump also requires the temporal analysis of the data. For this reason, two corresponding categories of features have been extracted, called *tracking-based* and *patch-based*, respectively. Tracking-based features Φ' are defined as follows:

$$\begin{array}{l} \Phi' = \{\phi'_1, \phi'_2\} \\ \phi'_1 = \sum_{k=t-\Delta t}^{t+\Delta t} |x_{k+1} - x_k| + |y_{k+1} - y_k| \\ \phi'_2 = |x_{t+\Delta t} - x_{t-\Delta t}| + |y_{t+\Delta t} - y_{t-\Delta t}| \end{array} \tag{8}$$

where (x_t, y_t) is the position of the person on the floor at time t and $2\Delta t$ is the patch size along the temporal axis.

The two features ϕ'_1 and ϕ'_2 take into account the motion of the person on the floor. ϕ'_1 is proportional to the length of the person trajectory, while ϕ'_2 is related to the distance between the source and the target position of the person within the considered patch. For example, the combination of both the features allows to distinguish among a static person, a person walking straight and one swinging on the same position.

Patch-based features Φ'' are computed on the three dimensional matrix $\mathbf{P}(i, j, t)$, which contains all the foreground values included in the selected spatio-temporal patch.

Let $M(t)$ be the mean value of $\mathbf{P}$ at time t and $b(t)$ the coordinates of the barycenter of the slice of $\mathbf{P}$ extracted at time t. Since the patch is extracted around the person's position estimated by the tracker, the barycenter $b(t)$ will be close to the center of the patch. However, $b(t)$ is also influenced by the person pose and the corresponding non-uniform distribution of his weight. $C(t)$ is the covariance matrix of the vectors $\{x, y, P(x, y, t)\}$. Let $\{e_1(t), e_2(t), e_3(t)\}$ be the three eigenvalues of the matrix $C(t)$.

The features $\Phi'' = \{\phi''_1, \ldots, \phi''_{16}\}$ are extracted as average, min or max of the previous defined values as follows:

$$\begin{array}{l} \phi''_1 = \frac{1}{n} \cdot \sum_{t=1}^{n} M(t) \\ \phi''_2 = \frac{1}{n} \cdot \sum_{t=1}^{n} \left(M(t) - \phi_1\right)^2 \\ \phi''_3 = \frac{1}{n} \cdot \sum_{t=2}^{n} |M(t) - M(t-1)| \\ \phi''_4 = \frac{1}{n} \cdot \sum_{t=2}^{n} \left(|M(t) - M(t-1)| - \phi_3\right)^2 \\ \phi''_5 = \frac{1}{n} \cdot \sum_{t=2}^{n} ||b(t) - b(t-1)|| \\ \phi''_6 = \frac{1}{n} \cdot \sum_{t=1}^{n} (||b(t) - b(t-1)|| - \phi_5)^2 \\ \{\phi''_7 \ldots \phi_9\} = \frac{1}{n} \cdot \sum_{t=1}^{n} \{e_1(t) \ldots e_3(t)\} \\ \{\phi''_{10} \ldots \phi''_{12}\} = \min_t \{e_1(t) \ldots e_3(t)\} \\ \{\phi''_{13} \ldots \phi''_{15}\} = \max_t \{e_1(t) \ldots e_3(t)\} \\ \phi''_{16} = \frac{1}{n \cdot M \cdot N} \cdot \sum_{t=2}^{n} \sum_{x=1}^{M} \sum_{y=1}^{N} |P(x, y, t) - P(x, y, t-1)| \end{array} \tag{9}$$

where ϕ''_1 and ϕ''_2 are the mean and variance of $M(t)$ over the temporal interval; ϕ''_3 and ϕ''_4 are the mean and variance of the $M(t)$ variations. ϕ''_5 and ϕ''_6 take into account the movement of the barycenter. Finally, ϕ''_7 to ϕ''_{15} evaluate the average, the minimum and the maximum three eigenvalues of the covariance matrix. Finally, ϕ''_{16} is the Mean of Absolute Differences (MAD) between consecutive values sensed by the same sensor. The concatenation $\Phi = \{\Phi', \Phi''\}$ of the tracking-based and patch-based features is used as input to a set of supervised Random Forest classifiers.

Random forests [Breiman 2001] are an ensemble learning method for classification (and regression) that operate by constructing a multitude (forest) of decision trees at training time. Each tree is

learnt on a random subset of the training data, generating a weak classifier. The output class of the forest is the mode of the classes output generated by the individual trees.

Since the proposed framework is conceived for human-computer interactions, a small set of natural and basic actions has been selected in this work. In particular, we aim at simulating common input devices such as a multi-touch screen or a mouse. People locations on the floor define the positions of the input commands, while jumps are mapped to *click* events. The *lying on the floor* class has been included in the action recognition classifier to enable surveillance applications. Finally, an *empty* class has been added to handle errors of the detection and tracking system. A qualitative and a quantitative analysis of the selected features and the implemented classifiers is provided in Section 7.

6. HIGH-LEVEL APPLICATIONS

Once the low-level processing unit has recognized actions and events, it maps them into generic commands. The high-level applications are basically unaware of the physical details of the input device. To this end, the Open Sound Control (OSC) standard protocol [Wright and Freed 1997] has been used to deliver triggers and command from the low-level processing unit to the client applications. OSC is suitable for networking sound synthesizers, computers, and other multimedia devices. Among others, OSC's advantages include interoperability, accuracy, flexibility, and documentation. OSC communications are based on a predefined set of string messages, which are delivered from the server to the clients. Person movements, actions and events are mapped to specific OSC commands. A list of the implemented messages is available in [Lombardi et al. 2013a].

This general and flexible architecture enables a plethora of applications, ranging from surveillance to entertainment [Lombardi et al. 2013a]. The external design of the device plays a fundamental role in the development of both collaborative and non-collaborative tasks (See Table IV).

Table IV: Sensing floor application taxonomy.

Non collaborative		**Collaborative**	
Data Collection	*Real time Event Detection*	*Basic Interaction*	*Full Interaction*
— Offline processing — Statistical analysis — Log data	— Surveillance — People counting — Statistical Flow analysis	— Single person position detection — Simple event detection [e.g., jump]	— Human Behavior Understanding — Multiple people interactions

The collection of data for offline processing, statistical analysis and logging are the basic features required by non-collaborative applications, such as people surveillance or flow analysis inside marketplaces or exhibitions. Introducing a more sophisticated processing level, specific events and information can be extracted in real-time to trigger alarms or actions. For example, the automatic detection of a queue at the counter may alert additional cashiers.

In the collaborative case, instead, the user explicitly interacts with the floor, expecting a real-time feedback in response to the action or movement performed on the floor. We have divided the applications into two categories, depending on the level of interaction and on the corresponding complexity of the processing engine. Basic applications require the instantaneous position of the people only as a control input. Future aims of the described project are more ambitious and, will include the integration of floor data with other sensors and the understanding of complex people behaviors and interactions.

In this section, we provide an overview of some applications developed to test the floor capabilities. In addition, we present and discuss a single-user interactive application for virtual street view navigation, which requires the detection of the basic events proposed in Section 5.3.

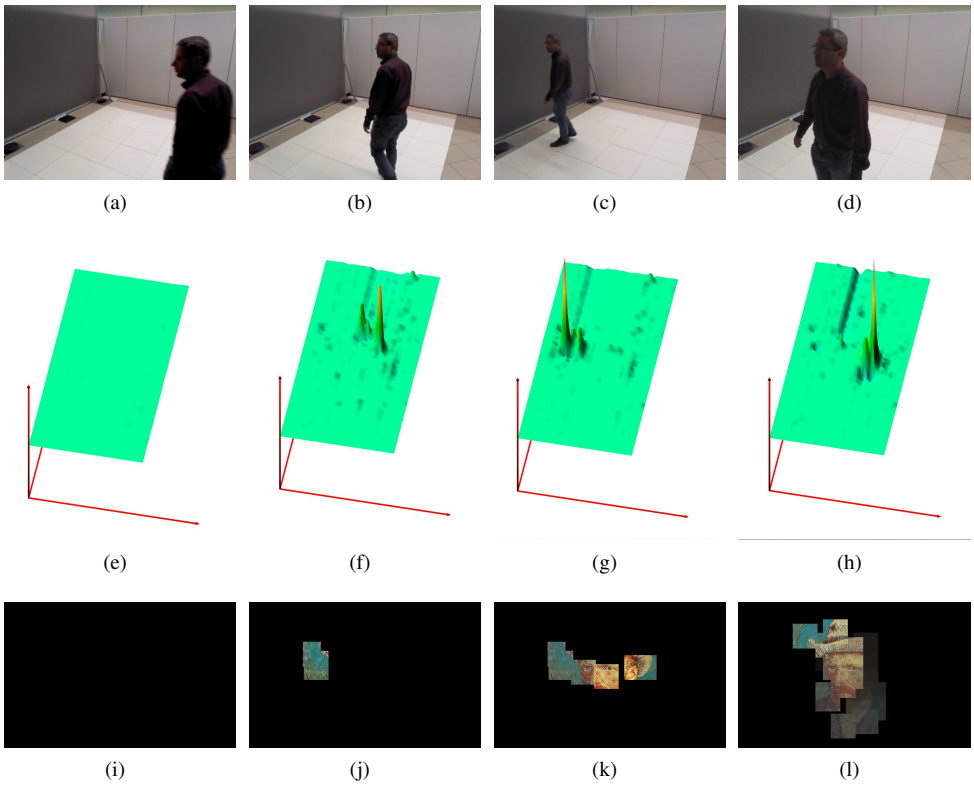


Fig. 11: Four frames from the foot painting demo. Top row: input image from a camera. Central row: a graph of the sensor values. Bottom row: the application output.

Foot painting: a projector screen is placed near to the sensing floor. A famous painting or a picture is selected and partitioned into small rectangular portions. Each of them is associated to a region on the floor. The position of a detected person triggers the projection of the corresponding image portion for few seconds. The user is thus encouraged to quickly walk all around the floor to unveil the entire image. Some snaps from the described application are depicted in Figure 11. The top row shows pictures of the sensing floor and of a user walking on it. The graphs in the central row highlight the sensor responses, while the output of the revealed painting is reported on the last row. Since the last frame (d) has been acquired few seconds after the previous one (c), the painting areas activated in (j) and (k) are no more visible in the corresponding final output (l).

Human Media Player: the setup is the same of the Foot painting application. A video or a picture sequence is projected on the screen. The position of the user on the floor controls the current media playing position.

Surveillance: the floor allows the detection of people within the boundaries of a safe area. Positive events may trigger alarms or start camera recordings. The proposed system has many advantages over traditional CCTV systems. The second ones are composed by a network of visual sensors and one or more processing and storage nodes. Despite their large diffusion, two major drawbacks still characterize vision-based systems. First of all, the complete coverage of wide areas is reached only using a redundant number of cameras. Placement constraints and the handling of occlusions due to furniture, objects and people usually impose the adoption of multiple views, especially in indoor environments. Furthermore, privacy issues strongly limit the usability and also user acceptability of surveillance systems. Completely automatic infrastructures only mitigate the problem by storing

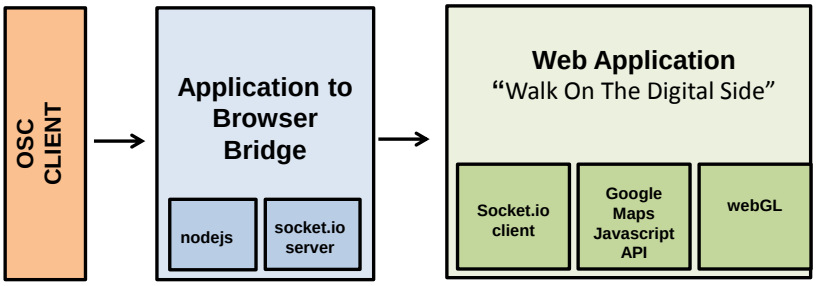


Fig. 12: The Virtual Street View modules: the OSC client and the web based application, connected through the node.js bridge.

and exploiting anonymised or aggregate data. This alternative solution obtained with sensing floors is cheap enough to allow the coverage of wide areas; in addition, the sensors can be integrated unobtrusively. Finally, the collected data do not contain identifying elements such as faces or biometric details, assuring a high respect of the user's privacy.

6.1. Virtual Street View

The setup of the virtual street view is composed by a pad made with the sensing floor technology surrounded by projection screens. The position of a user within the floor controls a "Virtual Street View" journey. After the selection of a starting and destination position on a Google Map (see Fig. 1(a)), frontal and lateral frames of the entire route are captured from Google Street View. On the one hand,a user walks toward the frontal screen, the projected frames depict a place closer to the path destination (see Fig. 1(b)). On the other hand when he/she walks back, the views are collected from places closer to the starting position.

The application is based on three modules: the OSC client, the web based application, and the bridge within the two previous units (see Fig. 12). The OSC client interacts with the OSC server (i.e., the low-level processing unit) to receive the user location on the floor. In turn, the client forwards the current position and corresponding events to the web browser thanks to the bridge, which exploits the *node.js*[4] and the *socket.io* frameworks[5] in order to establish suitable socket ports.

The web based application recovers the journey frames using the Google Maps Javascript APIs[6] and visualizes the output on a set of browser windows using the emerging web-standard *WebGL*.

7. EXPERIMENTAL EVALUATION

A prototype version of the described sensing floor architecture has been implemented and tested in our laboratory. As above mentioned in Section 5, the system is composed of two sensor modules (see Figure 9), covered with squared tiles of 0.6 m by 0.6 m each. The device generates floor images of 16x16 pixels. Low-level and high-level processing modules are running on the same computer, which also controls the projector screens. The capture board and some details of the hardware components adopted are shown in Figure 3.

Sensor data have been captured at 20Hz and stored for offline processing. A Microsoft Kinect 2.0 device has been installed and used to capture data streams synchronized to the main floor stream. In particular, the RGB color images and the 3d positions of people joints have been exploited with the twofold aim of annotating the sensor data and creating an automatic ground-truth for the people detection and tracking algorithm.

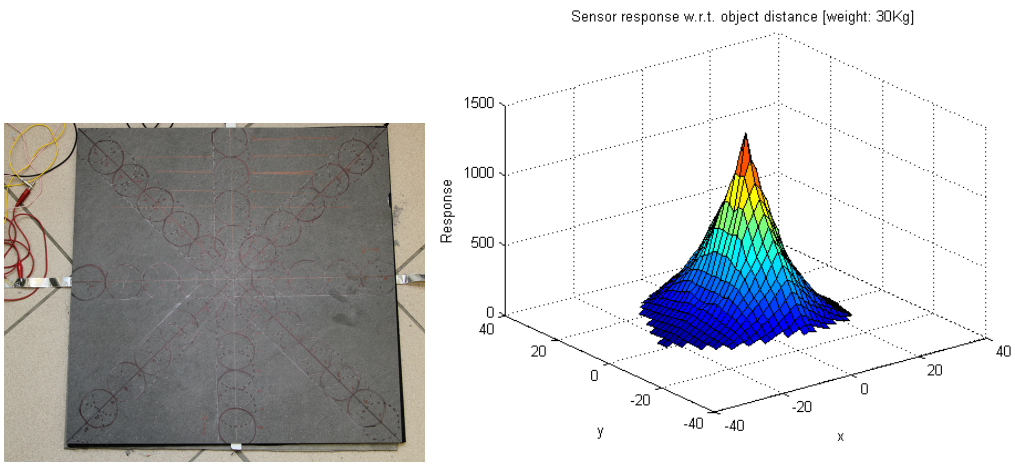
[4]http://nodejs.org
[5]http://socket.io/
[6]https://developers.google.com/maps/documentation/javascript

The extrinsic calibration parameters of the Kinect device with respect to the floor have been computed to obtain the geometric transformation between the original Kinect reference system and the floor reference coordinate system. At each frame captured by the floor device, the joint positions of people walking on the sensing area are extracted. Then, the central position between the feet is estimated and projected on the floor image ground plane. A preliminary dataset is available on the project website [7].

The experimental evaluation has been divided into four stages to test the sensor response, *i.e.* the low level processing chain, the action recognition algorithm, and the Street view application, respectively.

7.1. Experiments on the sensor response

The first experiment set is devoted to test and measure the behavior of the sensing layer coupled with the ceramic tiles. First, we have considered the response of one sensing element only and we have placed a reference weight at different positions with respect to the sensor. The Figure 13(a) shows a picture of the sensor and the tile used for this experiment. The object positions have been marked on the tile and the sensing element has been placed in the middle of it. The sensor response as a function of the sensor-to-object position is plotted in Figure 13(b). The graph highlights the smoothing behavior provided by the ceramic tile (see Section 4.3). In addition, the response is related to the distance between the sensor and the object only, independently from the directions of the two conductive stripes.



(a) (b)

Fig. 13: (a) The single-sensor tile used to test the system response. (b) Plot of the sensor response with respect to the position of the object.

In the second experiment, we have measured the output value provided by the sensor when exposed to different weights. Figure 14 plots the sensor response as a function of the object weight. More specifically, the plot reports the conductance G of the element with respect to the applied pressure. In the pressure range typically exerted by human feet (0.. 400 kPa, as reported in [Birtane and Tuna 2004]), the conductance is mostly proportional to the pressure. In addition to the contact

[7] http://imagelab.ing.unimore.it/go/sensingFloor

conductance, the applied pressure increases the internal resistance of the layer. This second phenomenon is less significant and appears only at very high pressures, when the contact resistance becomes almost stable (see the right part of the graph).

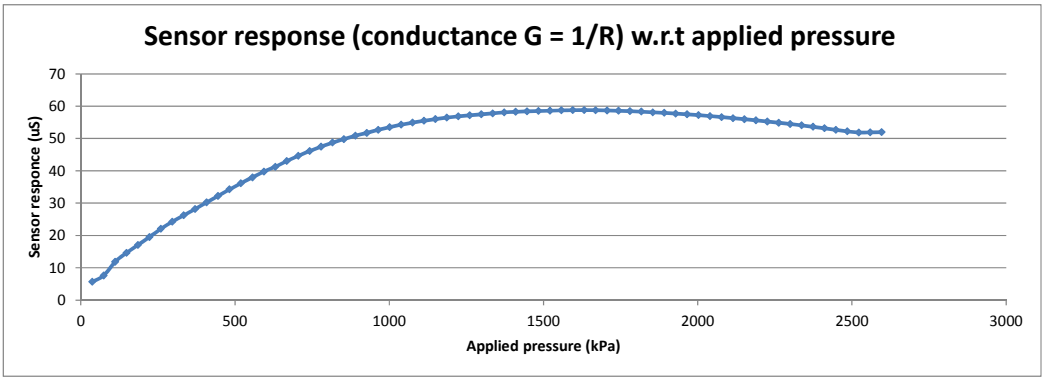


Fig. 14: Plot of the sensor response with respect to the pressure.

7.2. Experiments on the low-level processing unit

As above mentioned, the people detector, the tracker and the feet joint positions provided by the Kinect sensor $M(x, y, t)$ have been used as ground truth. The people detection and tracking algorithms of Section 5.2 have been evaluated in terms of True Positive (TP),True Negative (TN), False Positive (FP) and False Negative (FN) values [Nghiem et al. 2007]. A detection is considered as a True Positive if the Kinect sensor provides a corresponding person position closer than 30 cm. The precision $Pr = TP/(TP + FP)$ and accuracy $Ac = (TP + TN)/(TP + TN + FP + FN)$ metrics are computed as usual.

Table V reports the values of all the performance parameters estimated on eight different sequences from the dataset. The subscript indicates if the metric has been computed after the detection (D) or the tracking stage (T). The mean distance MD between the detection and the ground-truth positions is also reported in the last column. In all cases, this value is comparable to the sensor spatial resolution, indicating an optimal precision of the people localization step.

Table V: Accuracy results of People Detection and Tracking

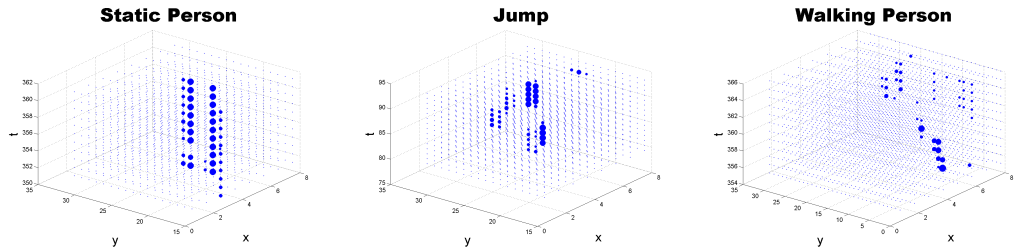
Seq	$\#fr$	Pr_D	Ac_D	Pr_T	Ac_T	TP_D	TP_T	FN_D	FN_T	FP_D	FP_T	$MD[cm]$
#1	1663	0.97	0.93	0.94	0.91	1263	1277	83	69	34	81	19.07
#2	760	0.98	0.88	0.96	0.96	509	578	80	11	12	23	16.84
#3	2470	0.97	0.89	0.94	0.90	853	890	235	198	28	54	20.12
#4	592	0.89	0.97	0.70	0.89	94	62	6	38	12	27	16.91
#5	1098	0.98	0.94	0.94	0.96	332	373	60	19	8	24	16.97
#6	655	0.94	0.90	0.92	0.95	232	271	49	10	15	25	15.09
#7	1735	0.97	0.88	0.96	0.94	1106	1227	175	54	37	56	15.27
#8	2062	0.98	0.89	0.97	0.95	1708	1863	201	46	31	54	13.92
	11035	0.97	0.90	0.94	0.93	6097	6541	889	445	177	344	17.03

Aggregate results reported in the last row of the Table show promising results of the system both in terms of precision and accuracy. In particular, the tracker improves the system accuracy by estimating the correct person position even when the floor is not able to. For example, the floor is not able to detect the person during jumps or fast steps, when both the feet are not pressing the sensors.

Conversely, the tracker contributes to the number of false detections, especially when a person exits the floor and the tracker still "remembers" the last seen position.

7.3. Experiments on the action recognition algorithm

Three visual examples of spatio-temporal patches are depicted in Figure 15. The size of the plotted points is proportional to the corresponding sensor value. The graphs show a promising classification capability of the spatio-temporal patches. The three plotted examples are captured in correspondence with a person standing (the sensor values do not change over time), jumping (the signal has a sort of break during the jump), and walking (shift of the firing elements during time).



(a) (b) (c)

Fig. 15: Spatio-temporal patches for three different actions: a) standing, b) jumping, and c) walking.

Table VI shows the accuracy of the action recognition system, as described in Section 5.3. The Table also reports the accuracy of a Support Vector Machine classifier as reference. For instance, the *jumping event* is detected with an accuracy close to 96% using random forest classifiers and close to 92% using SVM.

Table VI: Accuracy results of the action recognition algorithm.

Event	SVM	Random Trees
Empty	98.7%	**99.7%**
Standing	93.8%	**93.6%**
Walking	92.4%	**97.0%**
Jumping	91.0%	**95.2%**
Lying on the floor	96.6%	**98.3%**

7.4. Experiments on the Google Street view Application

Virtual reality frameworks are difficult to evaluate quantitatively, since it is not possible to define a ground-truth and a corresponding metric. Systems are usually evaluated by means of user interviews, where the *presence* of interaction is measured with subjective rates. We have followed the definition and the protocol provided by Witmer and Singer [Witmer and Singer 1998], extracting 11 questions among the 32 proposed. The unselected questions are referred to applications with audio interactions, which are not included in our setup. The survey has been submitted to a set of 50 users (having different weight, age and sex), after an interaction session with the prototype. The users provided a score on a scale from 1 to 7 to each question, where a lower score corresponds to a worse presence. Table VII reports the eleven questions together with the average scores assigned by the users.

The questions mainly belongs to two categories. On the one hand, the first category contains questions related to the control capabilities and are marked as *c* in the third column of the table.

Table VII: Presence questionnaire

ID	Question	Type	Avg Rate
Q1	How much were you able to control events?	*c*	6.3
Q2	How responsive was the environment to actions that you performed?	*c*	6.4
Q3	How natural did your interactions with the environment seem?	*c*	6.5
Q4	How much did the visual aspects of the environment involve you?	*p*	3.9
Q5	How natural was the mechanism which controlled movement through the environment?	*c*	6.5
Q6	Were you able to anticipate what would happen next in response to the actions that you performed?	*c*	6.5
Q7	How compelling was your sense of moving around inside the virtual environment?	*p*	5.7
Q8	How involved were you in the virtual experience?	*p*	4.1
Q9	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	*c*	6.6
Q10	How much did visual display quality interfere or distract you from performing assigned tasks or required activities?	*p*	5.9
Q11	How much did the control devices interfere with the performance of assigned tasks or with other activities?	*c*	2.1

Type: *c* control, *p* presence.

These questions evaluate how easy and naturally the user can control the application using the floor. The second category of questions, on the other hand, are more related to the subjective experience and are marked with the symbol *p* in the table.

The high average rate of questions related to the control confirms the performance of the whole system and in particular of the proposed sensing floor. Moreover, the score of question Q11 is 2.1, which indicates the low interference of the device with the virtual performance. However, the users have indicated a moderate presence rate, as highlighted by the corresponding scores to Q4 and Q8. Probably, the low-cost and limited quality of the visual setup have negatively influenced the virtual immersion (see questions Q7 and Q10).

8. CONCLUSIONS

In this paper, we have described an innovative sensing floor architecture. We aimed at creating a new form of natural interaction between humans and physical or virtual worlds. The architecture is based on a new sensing floor device, which acquires matrices of pressure points and converts them to gray level images. The developed prototype will be soon available as a commercial product. It allows to create low cost setups with standard ceramic tiles and can be easily integrated with data coming from other sensors, such as 3D cameras.

The system can be used as input device in a plethora of applications, ranging from entertainment to surveillance. The low cost of production and the scalability make the proposed solution very promising also from the commercial point of view, allowing wide installations in both private and public spaces (see the list of requirements reported in Section 1).

The prototype version of the floor installed in our lab has also shown encouraging reliability features. After one year from the installation of the hardware device, the surface resistivity of the foam and the corresponding variations when exposed to weight have not changed. The floor is still able to detect people and filter out the noise, as proved in Section 7. The reliability of the other components, such as the capture device, the connectors and the processing unit should be guaranteed by the industrial process.

The complete system can be easily moved thanks to the innovative characteristics of the floating floors. Temporary exhibitions, shopping centers, and stores may exploit this technology to monitor people flows and crowds instead of common vision-based surveillance systems. The floor system is not affected by occlusions, as vision systems are. If many people are walking on the floor, each person will generate a peak on the sensor data. A corresponding detection and track will be generated

by the low-level processing system. Of course, there are limits defined by the spatial resolution of the sensor. If two people are closer than this threshold, they will be merged in a single detection. Nevertheless, the same problem affects traditional vision system based on motion, when groups of people are walking close together.

Spatial resolution is currently set around 125 mm. This value has been selected not only to reduce the hardware equipment, but also because higher frequencies are denied by the ceramic layer. This resolution limits the application of the proposed floor in some fields, such as medical rehabilitation and postural analysis.

Moreover, it also reduces the amount of data captured and processed. For example, a sensing floor covering a room of one hundred square meters generates a floor image of 6400 pixels only, while a very large environment of ten thousand square meters produces an image smaller than 1 Megapixel. Real time processing and storing of floor images is thus reasonable even on low cost processing units. Finally, data transmission does not have tight bandwidth requirements, even in very large installations.

ACKNOWLEDGMENTS

This work was supported by Florim Ceramiche S.p.A. (Italy) and within the Regional Operational Programme POR FESR 2007-2013 of Softech-ICT. Some educational applications are under developing in collaboration with the Italian Company Vision-e s.r.l and are partially supported by the PON R&C project DICET-INMOTO (PON04a2-D). We finally thanks Dr. Federico Tasso (Accademia Belle Arti, Brera, Milan) for the development of the Street View application.

REFERENCES

D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. 2010. Google Street View: Capturing the World at Street Level. *Computer* 43, 6 (2010), 32–38.

J. Anlauff, T. Großhauser, and T. Hermann. 2010. TacTiles: a low-cost modular tactile sensing system for floor interactions. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM, 591–594.

Murat Birtane and Hakan Tuna. 2004. The evaluation of plantar pressure distribution in obese and non-obese adults. *Clinical Biomechanics* 19, 10 (2004), 1055 – 1059.

Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32.

Dorin Comaniciu and Peter Meer. 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 5 (May 2002), 603–619.

EIA. 1988. Packaging Material Standards for ESD Sensitive Items. (1988).

L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri. 2007. Actions as Space-Time Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 12 (Dec. 2007), 2247–2253.

H. Hassanpour, M. Sedighi, and A. Manashty. 2011. Video Frames Background Modeling: Reviewing the Techniques. *Journal of Signal and Information Processing* 2 (2011), 72–78. Issue 2.

I. Laptev. 2005. On Space-Time Interest Points. *International Journal of Computer Vision* 64, 2-3 (Sept. 2005), 107–123.

Martino Lombardi, Augusto Pieracci, Paolo Santinelli, Roberto Vezzani, and Rita Cucchiara. 2013a. Human Behavior Understanding with Wide Area Sensing Floors. In *Proceedings of the 4th International Workshop on Human Behavior Understading (HBU2013)*. Barcelona, Spain.

Martino Lombardi, Augusto Pieracci, Paolo Santinelli, Roberto Vezzani, and Rita Cucchiara. 2013b. Sensing floors for privacy-compliant surveillance of wide areas. In *Proceedings of the 10th IEEE International Conference on Advanced Video and Signal-Based Surveillance*. Krakw, Poland.

L. Middleton, A.A. Buss, A. Bazin, and M.S. Nixon. 2005. A floor sensor system for gait recognition. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*. 171–176.

A.-T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin. 2007. ETISEO, performance evaluation for video surveillance systems. In *Proceedings of IEEE International Conference on Advanced Video and Signal-Based Surveillance 2007*. 476 – 481.

J. Paradiso, C. Abler, K. Hsiao, and M. Reynolds. 1997. The magic carpet: physical sensing for immersive environments. In *Extended Abstracts on Human Factors in Computing Systems*. 277–278.

Massimo Piccardi. 2004. Background subtraction techniques: a review. In *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4. 3099–3104.

Ronald Poppe. 2010. A survey on vision-based human action recognition. *Image and Vision Computing* 28, 6 (2010), 976 – 990.

R. Rajalingham, Y. Visell, and J.R. Cooperstock. 2010. Probabilistic Tracking of Pedestrian Movements via In-Floor Force Sensing. In *Canadian Conference on Computer and Robot Vision*. 143–150.

Krishnan Rangarajan and Mubarak Shah. 1991. Establishing motion correspondence. *Computer Vision, Graphics, and Image Processing: Image Understanding* 54, 1 (1991), 56 – 73.

Ken Sakamura. 1999. Guest Editor's Introduction: Entertainment and Edutainment. *IEEE Micro* 19, 6 (Nov. 1999), 15–19.

D. Savio and T. Ludwig. 2007. Smart Carpet: A Footstep Tracking Interface. In *Proceedings of International Conference on Advanced Information Networking and Applications Workshops*, Vol. 2. 754–760.

C. Schurmann, R. Koiva, R. Haschke, and H. Ritter. 2011. A modular high-speed tactile sensor for human manipulation research. In *Proceedings of IEEE World Haptics Conference (WHC) 2011*. 339–344.

Y.L. Shen and C. S. Shin. 2009. Distributed Sensing Floor for an Intelligent Environment. *IEEE Sensors Journal* 9, 12 (2009), 1673–1678.

M. Shimojo, M. Ishikawa, and K. Kanaya. 1991. A flexible high resolution tactile imager with video signal output. In *Proceedings of IEEE International Conference on Robotics and Automation, 1991*, Vol. 1. 384–389.

Arnold Smeulder, Dung Chu, Rita Cucchiara, Simone Calderara, Afshin Deghan, and Mubarak Shah. 2014. Visual Tracking: an Experimental Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (July 2014), 1442–1468.

P. Srinivasan, D. Birchfield, G. Qian, and A. Kidané. 2005. A pressure sensing floor for interactive media applications. In *Proceedings of the International Conference on Advances in computer entertainment technology*. 278–281.

M. Valtonen, J. Maentausta, and J. Vanhala. 2009. TileTrack: Capacitive human tracking using floor tiles. In *IEEE International Conference on Pervasive Computing and Communications*. 1–10.

R. Vera-Rodriguez, J.S.D. Mason, J. Fierrez, and J. Ortega-Garcia. 2013. Comparative Analysis and Fusion of Spatiotemporal Information for Footstep Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 4 (2013), 823–834.

K. Weiss and Heinz Worn. 2005. The working principle of resistive tactile sensor cells. In *Proceedings of IEEE International Conference on Mechatronics and Automation*, Vol. 1. 471–476 Vol. 1.

Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240.

M. Wright and A. Freed. 1997. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In *International Computer Music Conference*. International Computer Music Association, Thessaloniki, Hellas, 101–104.

C.R. Yu, C.L. Wu, C. H. Lu, and L. C. Fu. 2006. Human Localization via Multi-Cameras and Floor Sensors in Smart Home. In *IEEE International Conference on Systems, Man and Cybernetics, 2006.*, Vol. 5. 3822–3827.