SCIRES it
SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 4, Issue 2 (2014), 81-86
e-ISSN 2239-4303, DOI 10.2423/i22394303v4n2p81

DESIGN OF A DIGITAL INTERACTIVE CONFIGURATION, FLEXIBLE AND INEXPENSIVE, FOR MONITORING THE MOVEMENTS OF PEOPLE IN REAL TIME

Enrico Costa, Camillo Trevisan**

*Università Iuav di Venezia, Dipartimento di Culture del progetto – Venezia, Italy

Abstract

We aim to introduce here the idea and the first results of a still ongoing research regarding the setup of a flexible and inexpensive digital interactive configuration, which has the goal of tracking the position and routes of a certain amount of subjects inside a medium or large closed space. The configuration appears to be adequate especially for the monitoring of museums spaces, exhibitions and showrooms.

Keywords

Computer Vision, Blob Recognition, Gesture Recognition, Real-time Tracking, Motion Estimation, Camera Calibration

1. *Goals of the project*

This research[1] expands and summarizes the previous works on the topic of development and testing of digital interactive configurations for heterogeneous exhibitions like museums, expositions, showrooms (Borgherini, Garbin & Trevisan, 2012; Trevisan, 2012, 2011, 2010, 2008, 2007, 2005 and FSE 2012-13 research report *Interfacce naturali e trasparenti per la comunicazione e la rappresentazione degli artefatti - Spazi sensibili interattivi a basso costo per eventi in luoghi pubblici*). From the former studies and experiments, other than the obvious and proven effectiveness of a natural and transparent interface, both new ideas and problems emerged. First of all the need of design flexible but at the same time robust and inexpensive configurations, given the chronic scarcity of financial resources both in the context of public museums and private institutions and, on the other side, the need of a fast setup of the configuration, often along with suboptimal conditions and non-standard environments that require specific solutions.

Another important point is the *collaboration:* in fact, when taking in consideration an interactive space aimed to manage a large and heterogeneous public, as in the current case study, it is without doubts necessary to setup some single-user or multi-user communicative stations[2], but it is mandatory to have a clear, real-time and detailed amount of information about the movements of the people in the monitored space. The usages of this kind of information are multiple: first of all foreseeing the future actions of the users allows to organize and direct their interaction experience. Secondly, it makes possible to create interactions among groups of people and not just individuals. In this way, the environment monitored becomes a cohesive organism made of different but collaborating parts.

Moreover, the continue and solid knowledge of the position and the aggregation of people, of the speed and the direction of their movements,

[1] The research, that will end in April 2015, is financed by the Regione Veneto with an annual FSE research fellowship awarded to Enrico Costa, and with prof. Camillo Trevisan as the scientific overseer. The project includes a collaboration between Università Iuav di Venezia, Ashmultimedia company, located in Vicenza, as operative partner, the *Soprintendenza speciale per il Patrimonio Storico, Artistico ed Etnoantropologico e per il Polo museale della città di Venezia e dei comuni della Gronda lagunare* as network partner and the technical consulting of Andrea Albarelli of Università Ca' Foscari di Venezia.

[2] The singles interactive station can be provided with Microsoft Kinects© version 2, that allows one to identify up to six *skeletons*, each composed by 25 nodes. Other choices can be Leapmotion Leap3D© for high precision hands movements, and Creative Interactive Gesture Camera©, somewhat more precise than the Kinects but with less action radius.

of even their gestures, create the possibility of adapting and modifying the environment itself to suit at the best the monitored situation, through lights, sounds, projections and other media.

Finally, the obtained information can be used to setup small or big local 'events' based on interactions between users. In other words, a broad knowledge allows to define and modify interactively a global narrative plot through a reactive intelligent system. The single 'stations' of the system distributed in the space can therefore assume different meaning on the base of different predictable configurations. A station could be single-user or multi-user, an alternation between the two states or even a joining of more stations behaving like a single entity, communicating its content differently depending on the situation.

In order to achieve the described goals it is necessary to investigate the environment automatically, checking every tenth of a second the position, the direction and the speed of each person in the environment. It is useful as well to have additional information, such as the presence of groups (individuals, visitors, couples and families), macroscopic gestures and direction of the sight, not always coincident with the direction of the movements, especially when the users are not in proximity of a specialized digital station. Moreover, associations between people and their own personal wi-fi devices (smartphones and tablets) can provide more insights.

2. *Development of the project*

We set ad main system requirements flexibility and inexpensiveness (Domhan, 2010, Kumar, Varghese, Pavan, Narendra, Prashanth, Girish & Balamuralidhar, 2014; Remondino, Del Pizzo, Kersten & Troisi, 2012) then deciding to adopt the following configuration: a server of medium power, through a wireless router (in the current configuration TP-LINK TD-W8960N was proven to be sufficient), is connected [3] via wifi to up to 254 (the maximum router capability) Android smartphones positioned on the ceiling and in the corners of the rooms. Smartphones are the ideal basic modules for this type of need, because they are inexpensive (the ones tested up until now have an average cost of around 50 euros). Moreover, their characteristics are ideal because they are small, and equipped with a wide-angle camera. They have a good processor, a built-in wi-fi board, and they can operate for several hours without the need of an external source of power, that sometimes is quite difficult to hide and requires work and time in order to set up. It's also worthwhile considering extreme cases, like the possibility of using smartphones with a broken screen – an unnecessary element for the intended use – and therefore even less expensive. It is not mandatory for the smartphones belonging to the configuration to be same model, but they must have Android 4 in order to run the software.

The smartphones, disposed in such a way that their fields of view opportunely overlap between each other, are calibrated and oriented, and operate independently of each other. They constantly analyze the environment detecting *blobs* – silhouettes of entities moving on the background, and subsequently guessing the *ground point* (identified as the nearest blob pixel to the Principal Point if the view is zenithal, and as the central lowest blob pixel if the view is slanted), and the guessed *head point*, obtained through a calculation over similar right triangles obtained from the center of projection (see Fig. 2). The *head point* and the *ground point*, along with other interesting data mined from the images and a timestamp, are then translated locally to the global coordinates using an homography matrix and returned to the server via wi-fi. In such way only few *strings* of data occupy the band at the same time.

Once a sufficient amount of data is collected, synchronized and appropriately processed (given the overlapping fields of view, the temporary absence of information regarding a specific *blob* usually do not compromise irremediably the calculations), the server has enough information to know the instant position of each person in the monitored space. In order to guess the speed and the direction of movement it is necessary to compare the current and the previous data, and therefore it is crucial to keep tracking of a person associating constantly his identifier with a blob detected by one or more of the cameras over the time.

The identification starts at the entrance of the subject in the environment covered by the system, with the help of easily recognizable patterns like the clothing colors. The tracking proceeds from the first identification until the person leaves the space monitored, and it is

[3] The server that manages the data input is in connection with other computers whose role is to produce and send contents as video projections, sounds, light modifiers and environmental media.

possible mainly thanks to the availability of a certain amount of simultaneous views of the location where the person is. An appropriately redundant arrangement of the tracking devices allow therefore to adapt the configuration to complex and irregular spaces with obstacles that occlude the view from certain angles.

Macro gestures and direction of sights can be recognized using different approaches that are currently being refined.

3. *Cameras calibration*

As briefly described above, each camera needs a proper calibration and the knowledge of the internal and external orientation parameters. A first calibration is performed just one time, and it is focused to the correction of the camera lens. This phase usually requires filming a checkerboard or grid printed on a paper sheet and through a procedure identifying the parameters used to fix the distortion.

Currently, the simple but sufficiently effective procedure is to place some smartphones on the ground, such that they can be seen from various cameras. Each of the smartphones then emits a specific screen light pattern. When the calibration phase starts, each camera can map the *ground devices* positions (which world coordinates are known) with the image coordinates of the patterns detected.

An interesting aspect of this method is that the operation does not necessarily require all cameras to be calibrated in the same time, so the *ground devices* used to calibrate the cameras within a specific area can be used afterwards in another area.

It is currently under test a method that could simplify this phase, making unnecessary the direct measure of the ground coordinates corresponding with the devices positions.

4. *Identification and tracking*

The goal of this phase is to track the people moving in the space with enough accuracy. The only available data in this phase are the images recorded from the cameras, that are analyzed in two different phases, at first in the local context of the device that filmed the scene, and secondly in the global context of the server that gather the local devices data and compares them globally (see Fig. 1).

The technique used locally is to detect moving entities subtracting pixel by pixel the current frame from the previous frames. The result of the subtraction is zero if the pixels belong to the background, because it is not changing over the time. On the other hand, the subtraction does not return the value zero if pixels change color, an event that mostly happens when an entity covers the background with its silhouette. This technique is called *background subtraction* and in the current case it was necessary to take in consideration the history of a certain number of previous frames because of the possibility for an entity to stop and stand without moving for a few seconds.

Afterwards, it is necessary to filter the noise generated by the changes of sun light overtime, and distinguish the shadows of the entities from the entities themselves.

5. *Implementation choices*

Android devices were chosen as tracking system, therefore the software was developed using JAVA programming language, the only solution available on this specific configuration. The IDE (*Integrated Development Environment*) Eclipse offers a framework completed with all the necessary libraries for the production of a JAVA Android application, directly verifying their results on virtualized or physical devices. Another critical element to be added was OpenCV, a set of libraries *computer vision* oriented, with BSD license. OpenCV provides precise and fast algorithms, useful both for the calibration and the tracking phases. OpenCV is written in the *C++* programming language, therefore the Nvidia Corporation TEGRA suite, an Eclipse distribution that includes JAVA for Android and OpenCV wrappers was chosen as development framework.

The JAVA client is installed on the Android devices and uses *background subtraction* algorithms to identify blobs, and a JAVA socket to transmit data to the server and then receive information.

6. *Development phases*

The project implementation was divided in two phases. Firstly, a simulator based on some test videos was implemented, in order to work on the detection without the issues of a physical setup of the configuration, which can be resource

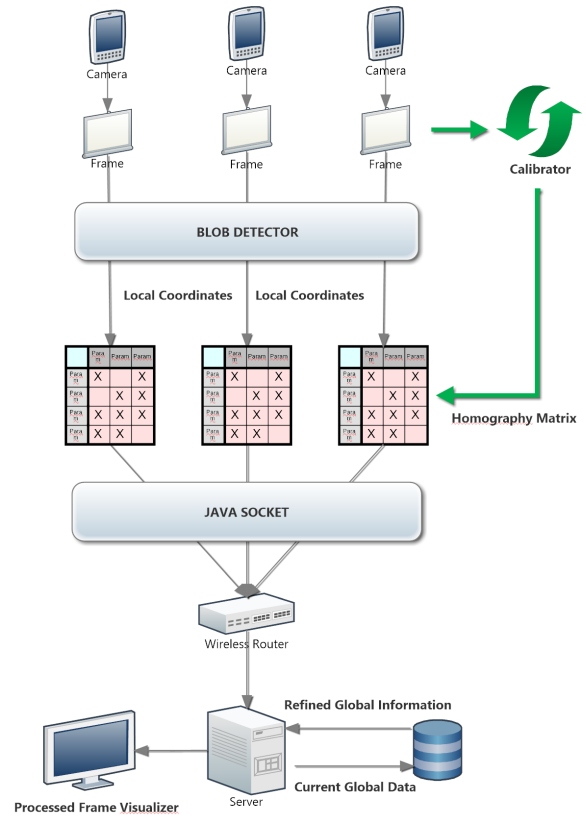
and time consuming. Each test video represents the same scene, seen from different points of view. Colored cylinders move around a closed space, stop, overlap and change speed and directions continuously. The simulator instantiate a class for each video, then the instances perform the blobs tracking independently from each other's and transmit the data to a server class that populates a table with the data, along with the video time in which they were gathered. The simulator constantly mines the table data and guesses the positions and the direction of movement of the blobs detected.

The second phase of the development focuses on the setup of a physical configuration that analyzes a real world environment, and therefore takes in consideration the cameras calibration and the most efficient network protocols, the noise suppression and the light changes over time.

7. *Future work*

The experiments performed lately underline as main criticality the low quality of the identified blobs, often affected by the object shadows and the background noise. More robust algorithms should be able to lower the uncertainty and eliminate the false positives. A possible solution could be the use of infrared illuminators provided with front pattern filters. The infrared light, visible from the smartphones if the wavelength is lower than 900 µm, provides many advantages. First of all, it is not sensible to the interferences of normal light (therefore, a person moving in a video projection cannot be confused with a physical moving entity). Moreover, being constant over time, it does not require a continue adaptation to mutating light conditions, especially if the environment is illuminated by sunlight, thus making the *background subtraction* way more efficient. Projecting an infrared pattern on the scene solves as well the problem of people standing without moving.

The possibility of using personal smartphones as useful tools in order to identify the position of the subject using them[4] still has to be tested.



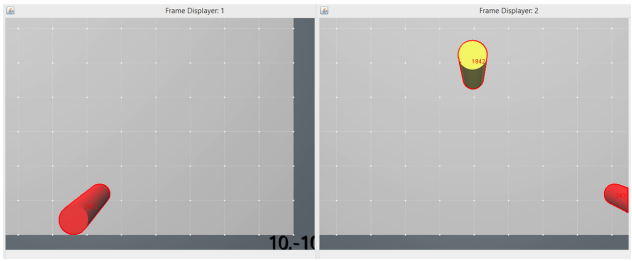


Fig. 1 (top): Process flow of the final configuration: each device camera send the acquired frames to its own blob processor, which detect the blobs local coordinates. Meanwhile, the calibrator check if the frames contains the calibrating pattern and produces homography matrices. The blob local coordinates are translated into world coordinates through the matrices and sent via wireless to the server along with timestamps and other significant data. Finally, the server queries the history database in order to refine the information on the current situation. At each iteration, the database content is updated, leading to an increasing knowledge of the environment monitored.

Fig 1 (bottom): Simulator at work on test videos. The two frames come from two different points of view, and are analysed independently at the same time. Blobs contours (red) are correctly detected, and each ID is shown in correspondence of the blob's barycentre. The red cylinder is detected in both frames, and the server will try to match the blobs using triangulation, movement history and other data, in order to understand if they detect the same entity.

[4] Another research (Gay & Trevisan, 2014) has the goal to produce a visitor's smartphone or tablet into an augmented reality device. In this case the main problem is to identify the highest possible accuracy of where the point of view is located as well as the direction of view of the camera. This is in order to send to the device proper images that can overlap the real scenes, giving more information to the user.

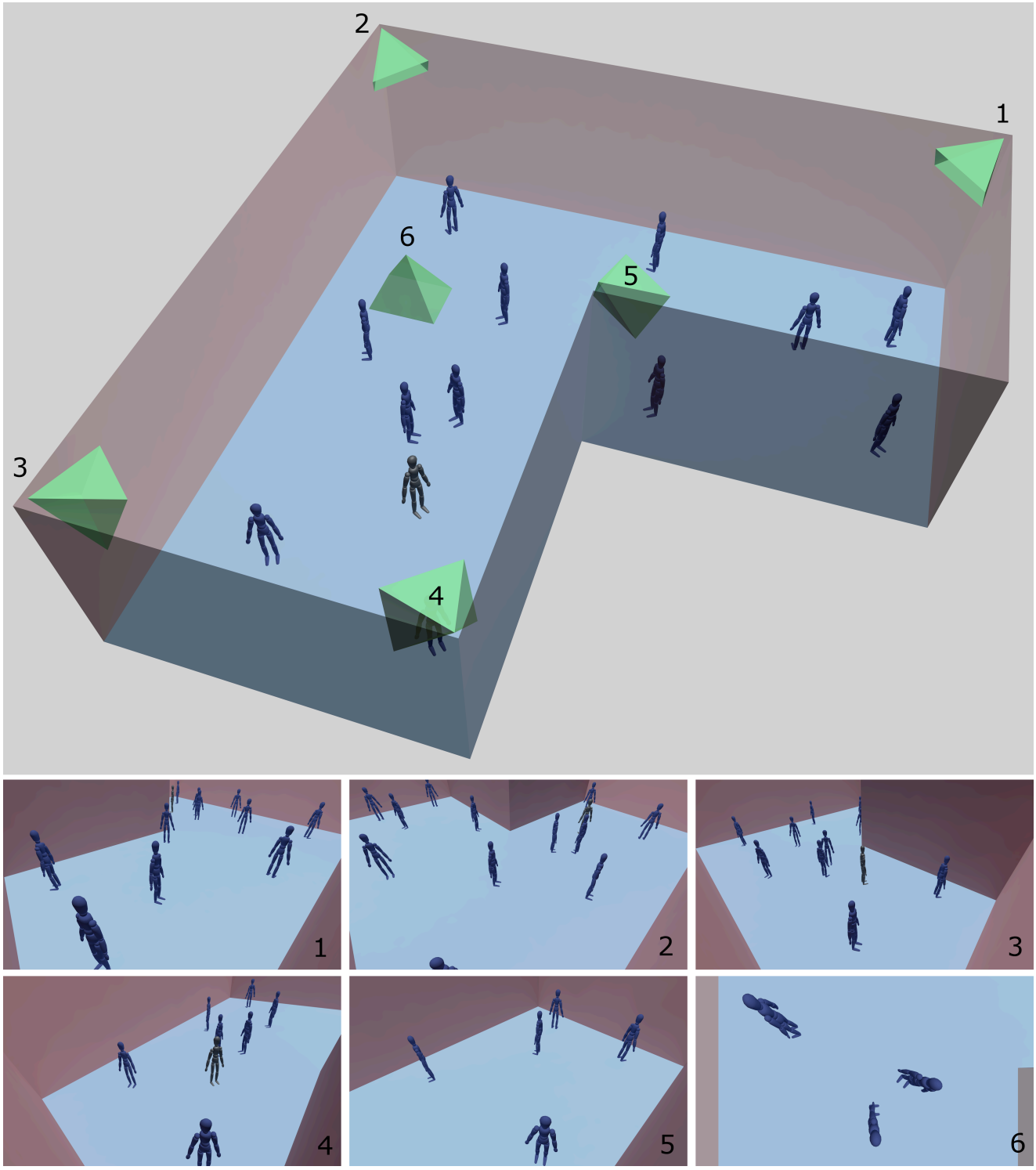


Fig. 2: Example of a possible arrangement of the cameras (smartphones) for an articulated space. Usually cameras placed at the vertices (cameras #1..5) are enough. In special cases, however, it is also possible to provide cameras with a vertical axis placed on the ceiling (for example, camera #6), to avoid blob's overlap. The use of additional wide-angle, even if it introduces a worsening of the accuracy, it is very useful to reduce the number of devices needed to overlay the entire space.
In any case are excluded connecting data cables to the server or for electric power: the data are transmitted via wifi, and the power supply is ensured by additional batteries.

REFERENCES

Borgherini, M. Garbin, E. & Trevisan, C. (2012). Una collezione di architetture digitali: i modelli digitali delle chiese di Andrea Palladio a Venezia. In G. Beltramini, M. Gaiani (Ed.), *Palladio LAB Architetture palladiane indagate con tecnologie digitali*. Vol. 11, (pp. 49-56), Vicenza: Centro Internazionale di Studi di Architettura Andrea Palladio.

Domhan, T. (2010). *Augmented Reality on Android Smartphones*. PHD Thesis DHBW Dualen Hochschule Baden- Württemberg, Stuttgart. Retrived from http://softwareforschung.de/fileadmin/_softwareforschung/downloads/WISTA/Tobias_Domhan_Studienarbeit.pdf

Gay, F. & Trevisan, C. (2014). Un museo di rilievi e un progetto della rappresentazione architettonica - A monumental museum and a project on architectural representation. In *Proceedings of Convegno UID Italian Survey & International Experience* (pp. 847-52). Roma: Gangemi Editore.

Kumar K., Varghese A., Pavan K., Narendra N., Prashanth S., Girish C., & Balamuralidhar P. (2014). An Improved Tracking using IMU and Vision Fusion for Mobile Augmented Reality Applications. *The International Journal of Multimedia & Its Applications (IJMA)*, 6(05), 13-29.

Remondino F., Del Pizzo S., Kersten T.P., & Troisi S. (2012). Low-Cost and Open-Source Solutions for Automated Image Orientation - A Critical Overview. In *Proceedings of Progress in Cultural Heritage Preservation, Lecture Notes in Computer Science* (pp. 40-54). Berlin Heidelberg: Springer.

Trevisan, C. (2005). Strumenti digitali per la comunicazione pubblica. *DIID. Disegno Industriale - Industrial Design*, vol. 16(05), 78-83.

Trevisan, C. (2007). Esplorazione interattiva di modelli digitali, Progetto di una stazione monoutente stereoscopica multimodale e multimediale. In *Proceedings of Sistemi informativi per l'architettura* (pp. 582-587). Firenze: Alinea.

Trevisan, C. (2008). Stereoscopia e interfacce naturali nell'esplorazione interattiva di modelli digitali. In R. Migliari (Ed.), *Prospettiva dinamica interattiva. La tecnologia dei videogiochi per l'esplorazione di modelli 3D di architettura* (pp. 166-175). Roma: Kappa.

Trevisan, C. (2010). Comunicazione digitale interattiva dell'architettura in ambito museale. In Borgherini M., Guerra A., Modesti P. (Ed.), *Architettura delle facciate. Le chiese di Palladio a Venezia. Nuovi rilievi, storia, materiali* (pp. 239-251). Venezia: Marsilio.

Trevisan, C. (2011). Pareti Digitali. In *La ricerca nel disegno di design* (pp. 144-149). Santarcangelo di Romagna: Maggioli.

Trevisan, C. (2012). Gesture 3D per la definizione della giacitura di un sistema cartesiano e il controllo di sei gradi di libertà: tre rotazioni e tre traslazioni lungo gli assi cartesiani. Patent filed to Patent Office in Bern, Switzerland, N. 2301/12.

Trevisan, C. (2012). Descrizione di tre configurazioni digitali utili per la rappresentazione e l'interazione con modelli digitali 3D. In A. Casale (Ed.), *Geometria Descrittiva e Rappresentazione Digitale, Memoria e Innovazione*. Vol. 2 (pp. 235-252). Roma: Kappa.