

PRHOLO: Interactive Holographic Public Relations

Ricardo Alves, Luís Sousa, Aldric Negrier, João M.F. Rodrigues, Pedro J.S. Cardoso, Jânio Monteiro,
Miguel Gomes and Paulo Bica

Abstract—The expansion from regional to global markets is nowadays a normal step in the life cycle of many successful companies. Public relations are company's first contact with potential clients, requiring them to give a small explanation but yet accurately enough on what the company does and how the client can benefit from it. By combining a 3D depth sensor with a holographic representation, an installation was developed capable of natural interaction with users, which can be configured for the company's needs, while passing important information. Users are capable of interacting with the holographic form using intuitive gestures and the system is capable to record users' interactions, creating key statistics to return to the company about their main products, the attention given to each presented content, user's actions, favourite menus, etc., including the creation of a personalized visitor cards with the users face.

Keywords—Human-Computer Interaction, Applications, Gesture Recognition, Holography, Computer Vision.

I. Introduction

A public relations has the fundamental job, within a company, to spread information to potential and regular clients. For this reason, it is very important that public relations (PR) know the right way to capture the attention and give good first impression to potential clients, persuading them that the company has the right tools and products to fit their needs. A different approach from the human PR is a holographic PR, which can capture the attention of a client using manifold visual (and sound) effects, and at the same time allowing a natural interaction between the user and the computer.

Projections techniques are an interesting way to achieve it, since the created virtual character is not limited to screen size and can have a real size. On the other hand, holography [1][2] is one of the most realistic form of image and video display and can sometimes, be defined as a photographic image that is 2D but appears to have depth. Pepper's Ghost [3] is one of the most popular techniques to create a holographic illusion and can be combined with projection. This last case is implemented using a projector with a retention film, to hold the projected image, while an acrylic sheet (or Mylar foil) reflects that image to the spectator, which is placed at a 45 degrees angle to the retention film. The combined result gives the illusion of an image floating in the air while also giving depth perception; see e.g. [4][5].

On the other hand, one of the new paradigms for Human-Computer Interaction (HCI) are the three-dimensional sensors, such as the Microsoft Kinect [6], the Asus Xtion [7], the Leap Motion [8] or the Structure Sensor [9]. Those sensors can be used for instance to interpret specific human gestures, enabling a completely hands-free control of devices. Hence, with the appropriate software, most of them have also the capability to detect the user's skeleton and/or tracking a single or several users, while replicating with accuracy the hands and the user movements in a 3D mesh. One of the most well-known 3D sensors is the Kinect, due to the game industry. Nevertheless, there are many other applications where this sensor is used, as for instance enabling interaction with art installations [10], in robotics [11], for head pose classification [12], applied in assistive technologies, as the enhancing visual performance skills on older adults [13] or for the operation of wheelchairs [14]. More challenges and applications of the Kinect can be found e.g. in [15]. Of course, HCI can be done with other 3D sensors, such as the mentioned Leap Motion, an example can be found in [16] where the interaction is done with holograms for teaching technical drawing.

In this paper an interactive human size holographic PR from an industrial (commercial) installation is presented. The hologram can be represented by an avatar or by a video from a real human PR, both allowing to show different contents (text, image, video, maps, etc.). The interactivity is achieved using a Kinect sensor and very intuitive gestures, recording all the users' interactions, creating key statistics to return to the company about their main products, the attention given to each presented contents, users' actions, favourite menus, viewing direction, face and body extraction, etc.

Despite the already huge amount of applications that uses HCI, as far as we know none have all this characteristics mentioned above. The main contribution of the paper is real size customizable interactive PR installation, that benefits from a natural interaction with the user, granting more realistic human (hologram) to human interaction, while building statistics about users interactions.

Ricardo Alves, Luís Sousa, Aldric Negrier, João M.F. Rodrigues, Pedro J.S. Cardoso

Instituto Superior de Engenharia, LARSyS (ISR – Lisbon), CIAC
University of the Algarve
Faro, Portugal

Jânio Monteiro
INESC-ID (Lisbon) and Instituto Superior de Engenharia,
University of the Algarve
Faro, Portugal

Miguel Gomes, Paulo Bica
SPIC – Creative Solutions
Loulé, Portugal

II. PRHOLO system

The installation is divided in three main modules: The Users Data (a), the Kinect extracts spatial coordinates from users joints and colour image, to detect the interaction, create statistics and any other relevant information (e.g., user body and face image). The Database (b) is responsible for storing menus options and configurations (collected from the Back office), as well as storing statistics collected from (a). The Interface (c) is responsible for processing visual information, displaying menus and the hologram. For the hologram projection, the Pepper's Ghost technique was used. Figure 1 shows the basic illustration of the technique; a detailed explanation is out of the focus of the paper, for more information please refer to [1][2][3][4][5][16]. In the following sections all three modules are explained in more

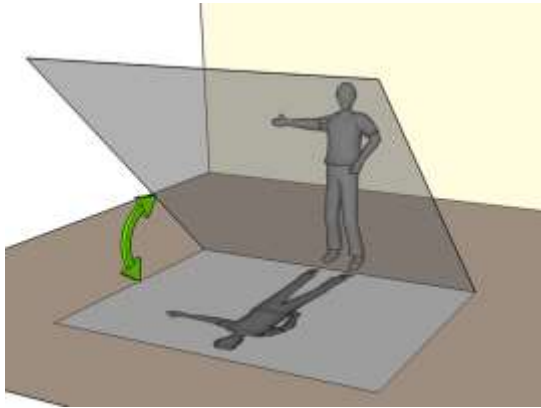


Figure 1. Implemented holographic scheme using Pepper's Ghost technique.

detail.

A. Users Data Module

Users Data module is responsible for handling and manipulating received data from Kinect (not Kinect 2). This sensor provides colour (RGB) and depth frames, 25 joints of 2 users and can track up to 6 users (not used here, it can also provide sound); for more details see [6]. The above data is manipulated in 5 main phases: (a) spatial information, (b) gestures recognition, (c) heat map, (d) user head direction, (e) body and face extraction:

(a) **Spatial Information** (x, y, z) of user's joints including the global position (P) , are stored using two FIFO (First In, First Out) list, one for the current (detected) users (Pc) and another for lost users (Pl) . Every time the Kinect has new information available about users, 3 different situations may occur: (i) *new users are available*, which are added to the end of the current users list (Pc) alongside all joints information and entry time (t) , but (ii) *if the user already exists*, then the new information is stored with the previously obtained information on the current users list, alongside the entry time of the new information. (iii) *If a user is lost*, then all information is updated in the database and the user is discarded from the current users list and added to the lost users list. Since users can sometimes block the Kinect's view, a simple occlusion

detection was built to recover a lost user. When a new user is detected, it is verified if, in the lost users list, there are lost users less than Δt seconds ago. For all of the last positions of all lost users (i) in the past Δt seconds, Pl_i , the current position of a detected user (u) , Pc_u , is considered to belong to a previous user if the Euclidean distance $d_{i,u}$ between Pl_i and Pc_u is less than 30 cm. If more than one lost user is closer than 30 cm to the position P_u then the closest distance d is chosen, recovering all information to the current users list (it was used $\Delta t = 5s$).

(b) **Gestures recognition** is responsible for inputs detection from users. After experimenting with different gestures, the "swipe gesture" and the "pose gesture" were chosen and implemented due to their intuitive nature for the users; Figure 2 left, illustrates those gestures, in the left the "pose" and on the right side of the image the "swipe". For the *swipe gesture* implementation, given minimum swipe distance d_s and the minimum swipe velocity v_s , a time window Δt_s can be calculated with $\Delta t_s = d_s/v_s$. By analysing only user information acquired from current instance t to $t - \Delta t_s$, a swipe was made if in any other sub-interval Δt_k , defined between t and $t - \Delta t_s$, with k the latest instance of Δt_k , all hand positions (h) minus its previous hand position, taking only into account the x component for the horizontal swipe, or the y component for the vertical swipe, has the same signal along the whole interval Δt_k , $|\sum_{j=k-\Delta t_k}^{k-1} sgn(h_{j+1,\{x/y\}} - h_{j,\{x/y\}})| = k - 1$, and if the total distance travelled, taking only into account the x/y component respectively for the horizontal and vertical swipe, between the first and the last point of that sequence is greater or equals to d_s , that is, $d_s \geq |h_{k,\{x/y\}} - h_{k-\Delta t_k,\{x/y\}}|$. Figure 2 right, illustrates a horizontal swipe. The minimum distance used was $d_s = 30$ cm and the minimum speed used was $v_s = 200$ cm/s. The *pose gesture* can be detected using the vector defined by the user's body $\vec{V}_b = (x_b, y_b, z_b)$, which is computed by subtracting the *shoulder centre* [6] position, P_{sc} , from the *spine* position, P_s , and using a vector defined by the arm, $\vec{V}_a = (x_a, y_a, z_a)$, computed by subtracting *hand (right / left)* position, P_h , from the *elbow (right / left)* position, P_e . The user is performing the menu pose if the two following conditions are true: (1) As the user must be facing Kinect, the distance between the *elbow* and the Kinect must be approximately the same as the *hand* and the Kinect. For this reason, only if $|z_h - z_e| \geq 0.6 \times d(P_h, P_e)$ is considered that the user is performing the gesture. (2) An angle θ between \vec{V}_b and \vec{V}_a can be measured with $\theta = \cos^{-1}((\vec{V}_b \cdot \vec{V}_a) / (|\vec{V}_b| |\vec{V}_a|))$. If the angle is $20^\circ \leq \theta \leq 160^\circ$, then the user is doing the pose

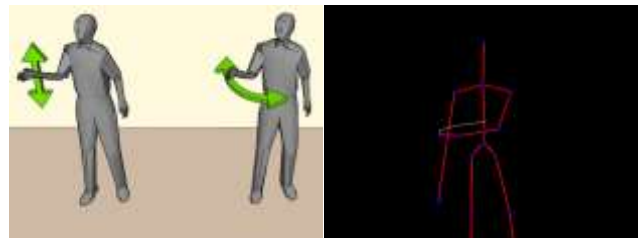


Figure 2. On the left, implemented gestures. On the right a swipe gesture being implemented.

gesture. To increase the reliability of the pose gesture, it is verified if at least 85% out of the previous detections made in the past 1 second are according to the above conditions. If the user is performing the pose gesture, it is considered - up - if the angle θ is less than 90° , and - down - otherwise.

(c) **Heat Map** of users is also calculated, useful for statistics about the most active locations of a user or a group of users. The users' global position (P) returned by Kinect, Figure 4 left, represents the distance in each axis between the user and the Kinect, with the x axis representing the horizontal distance and the z axis the vertical distance. In this application, it is assumed that the physical limits of the Kinect is 4 m in length and in width [6]. A matrix M with size of $N \times N$ can be created, dividing the map in squares of approximately $(4/N) \text{ m} \times (4/N) \text{ m}$. In the present case, it was considered $N = 26$, consequently each square as the area of 0.024 m^2 . A user position detected from Kinect on instance t is mapped to matrix M using $(x_t, y_t) = (-x_{k_t} \times (N/4) + N/2, z_{k_t} \times (N/4))$, with x_k and z_k the coordinates obtained from Kinect (Figure 4 right).

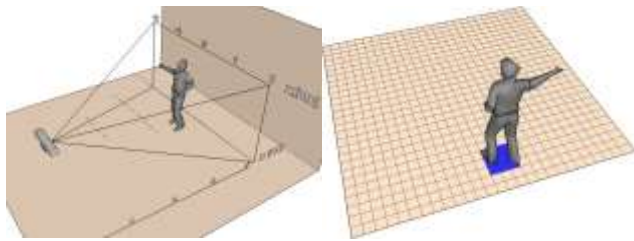


Figure 4. On the left, Kinect coordinate system. On the right, a user being mapped to matrix M .

Starting with every position of the matrix M equals to zero, every detection made at instance t increments its value, as well as its 8 neighbours: $M(x_t + i, y_t + j) = M(x_{t-1} + i, y_{t-1} + j) + 1$, with i and $j = \{-1, 0, 1\}$. After this step, the matrix M holds values of the positions that a user, or a group of users (optional). The matrix is then normalized, Figure 5 left, and converted to a colour map using JET colour map, visible on Figure 5 right.

(d) **User Head Direction** is used to estimate in which direction the user was looking and the time spent looking at the installation while interacting. Since Kinect SDK only allows two people for face tracking, only the first two users from the current users list are analysed. Getting a precise point where a user is looking to is difficult, mostly due to Kinect colour resolution. For this reason, reference points are used for the user's head to determine where the person is probably looking to, despite a person's head might be facing a direction while his/her eyes may be looking elsewhere. Using Kinect

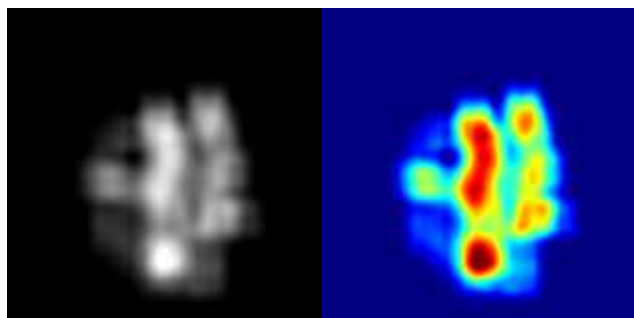


Figure 5. On the left, the normalized matrix M . On the right the same matrix converted to the JET colour map.

SDK the position of the users' eyes are obtained, in the colour frame, (Figure 3 red dots). This information is used, alongside the *shoulders* (S) position (left, centre and right; black dots in the same Figs) [6] where he/she might be facing. With the coordinates of the points, it is calculated the normal vector \vec{V}_s' to the vector \vec{V}_s , defining a line containing both shoulder left and right points in the colour frame. Vector \vec{V}_s is calculated using shoulder left (l) and right (r) points, $S_{l/r} = (x_{S_{l/r}}, y_{S_{l/r}})$, with $\vec{V}_s = (x_{S_l} - x_{S_r}, y_{S_l} - y_{S_r})$, thus being $\vec{V}_s' = (-V_{s_y}, V_{s_x})$. Applying \vec{V}_s' to the shoulder centre (c) Position (vertical or almost vertical blue line in the same Figs), three different scenarios can happen: (1) a user can have each eye on each side of the line, Figure 3 centre, meaning that the user is looking to the centre, (2) both eyes can be placed on the left side of the line, Figure 3 left, thus the user is looking to the right and finally both eyes are placed on the right side of the line, Figure 3 right, determining that the user is looking left. Depending on the position where the user is, extreme left and right directions were excluded. A timer was implemented to count the time spent looking in each direction.

(e) **Face and Body Extraction** is done and shown to the user as soon as he begins to interact with the installation. While the Kinect SDK does all of the job extracting the user face, Figure 6 first image, full body photo is not done directly. Every time



Figure 3. On the left, user looking to his right. On the centre, user looking to the middle. On the right, user looking to its left.

a new depth frame is available, the first three bits represents which user that pixel belongs to, from 1 to 6 and with 0 meaning it belongs to no user. A Boolean mask for each user is constructed, visible in Figure 6 second image. For that image the value for the highest and lowest $x_{min/max}$ and $y_{min/max}$ is found. The colour image is then cropped (U) using this points. An example can be seen in Figure 6 third image. Alternatively a Gaussian filter is applied to the Boolean mask ($\sigma = 2$), and again the image is cropped and a bitwise AND logic is performed with U , obtaining Figure 6 fourth



Figure 6. From left to right, extracted user's face, Boolean mask of the user, extracted user's body and extracted user's body with background removal.

image. These images (face and body) are used to insert the user in different backgrounds/situations and then offered to the user as “promotional gifts”.

B. Database Module

The database module stores two types of information: Menus configurations (i) enabling customization of the whole application changing the appearance of it to match companies' schemes, colours and logos. To adapt the interface to the company's needs, the interface can be entirely changed: (a) Content of the application can be customized. An application can have several menus, each menu with any number of contents, with an action associated with it. A content action can be: opening another menu, running a video or displaying an image. (b) To fit different types of information, 6 different menu layouts were build. (c) Design of the menus can also be changed. Images of buttons, font type, font colour and background images can be changed within a menu. Once a menu has been created, a company can change any of those parameters. (d) The avatar can also be changed to any 3D model of a character, or a video of a human PR.

Statistics (ii) can be quite useful to know users feedback, which information has been most requested or the time spent on each content. The following information is stored: (a) Users spatial position, alongside joints positions, enabling a replication of what the user has done while interacting with the installation. (b) Menus statistics are also saved. For each user it is counted the time he/she spent on it and also the number of selections. This is useful to determine which information is most requested by users. The user face and body photo are also saved and can be, afterwards, matched with the information requested to see what kind of product was a user most interested on.

C. Interface Module

The Interface module is responsible to read information stored in the database and automatically generate layouts, as well as generating avatar responses to the user input. To interact with the menus, in the present installation were used the horizontal swipe and the pose gestures (Sec. II.A). In resume, the application can generate the following layouts: (a) titles of menu contents useful to serve as bridge to other menus, visible in Figure 7 centre, (b) similar to (a), a layout that can fit small description if needed. In specific cases, (c) media content as image or video with description is useful to get a user the idea beyond a concept, generating a layout similar to (b). Alternatively, the same layout can be used but

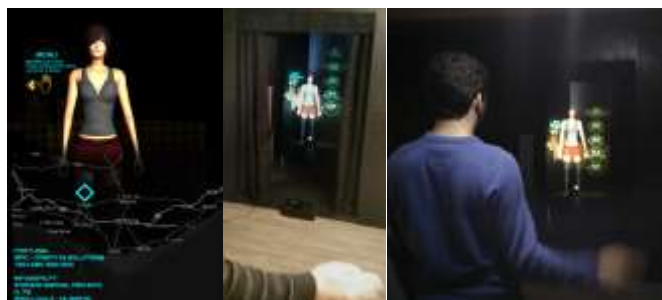


Figure 7. On the left, a menu example. On the centre and right, the installation.

displayed in a diagonal instead of vertical, visible in Figure 7 right. (d) One final option is available to display only images or video, Figure 7 left. All configurations are loaded when the application starts. The avatar trigger responses to inputs of the user: (a) If the interacting user with the installation changes, the avatar waves while showing that user's photo. (b) If the user selects an option, the avatar touches its content triggering a menu change. It is important to accentuate that despite the images in Figure 7 centre and right do not give that notion, the holographic installation has the height of ≈ 2 m and a width of ≈ 1.2 m and the lady avatar has the height of a human lady ≈ 1.7 m.

III. Conclusion

In this paper a PR installation was presented, interacting more naturally with users while enhancing their visual experience. Using a holographic technique, an avatar can interact with a user giving him/her information about a company. Using a Microsoft Kinect, users are detected and can interact with the avatar using gestures, while tracking all information about them, useful for statistics. Statistics about users' positions and body joints are stored in a database, being useful for heat maps as well as statistics about time spent on each content and requested information.

The results of swipe gestures were good, as previous hand positions were taken into account. The detected successful rate was 98%, being too difficult to tell exactly what did fail. Usually it was due to the user not making the gesture correctly, due to a lack of speed or gesture extend (distance). In respect of user looking estimation, as expected, the results were not as good. If a user is facing towards the Kinect, the result obtained works as expected, giving an estimation on where a person might be looking, despite his/her eyes might be looking elsewhere, but if the user is sideways to Kinect results are not reliable enough, although it is not a critical error for this application.

Future work includes improving recognition of previous user, not only when Kinect view is blocked but also when an user returns some hours later to ask for more information, as well as improving the user head direction.

Acknowledgments

This work was supported by FCT project LARSyS (UID/EEA/50009/2013), CIAC (PEstOE/EAT/UI4019/2013), PRHOLO QREN I&DT, nr. 33845, POPH and FEDER. We also thank our project leader SPIC - Creative Solutions [www.spic.pt].

References

- [1] S. Antonio, R. Herrera, E. Enriquez, “Projection’s panel of models for touch screen,” *Int. J. of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 9, Nov. 2013, pp. 2057–2064.
- [2] E. Mihaylova (ed) “Holography - Basic Principles and Contemporary Applications,” *InTech*, 2013.
- [3] J.C. Sprott, “Physics Demonstrations: A sourcebook for teachers of physics,” *Univ of Wisconsin Press*, 2006.
- [4] D’Strict, “3D sensing holographic installation,” <http://goo.gl/CyqKXe>, 2014 (retrieved 26/2/2014).

[5] J.M. Moser, "Tupac lives! what hologram authors should know about intellectual property law," <http://goo.gl/zMxTZP>, 2014 (retrieved 26/2/2014).

[6] Kinect, "Kinect for Windows," <http://www.microsoft.com/en-us/kinectforwindows/>, 2014 (retrieved 26/2/2014).

[7] Asus, "Asus Xtion Pro," <http://www.asus.com/pt/Multimedia/XtionPRO/>, 2014 (retrieved 26/2/2014).

[8] Leap, "Leap Motion," <https://www.leapmotion.com/>, 2014 (retrieved 26/2/2014).

[9] Struture, "Struture Sensor," <http://structure.io/>, 2014 (retrieved 26/2/2014).

[10] R. Alves, M. Madeira, J. Ferrer, S. Costa, D. Lopes, B.M. da Silva, L. Sousa, J. Martins, J. Rodrigues, "Fátima revisites: an interactive installation," In Int. Multidisciplinary Scientific Conf. on Social Sciences and Arts, SGEM, 2014, pp. 141–148.

[11] R.A. El-laithy, J. Huang, M. Yeh, "Study on the use of Microsoft Kinect for robotics applications," In IEEE/ION Position Location and Navigation Symp., IEEE, 2012, pp. 1280–1288.

[12] Y. Yun, M.H. Changrampadi, I.Y. Gu, "Head pose classification by multi-class AdaBoost with fusion of RGB and depth images," In Int. Conf. on Signal Processing and Integrated Networks, IEEE, 2014, pp. 174–177.

[13] I.T. Chiang, J.C. Tsai, S.T. Chen, "Using Xbox 360 Kinect games on enhancing visual performance skills on institutionalized older adults with wheelchairs," In IEEE Int. Conf. on Digital Game and Intelligent Toy Enhanced Learning, IEEE, 2012, pp. 263–267.

[14] F.A. Kondori, S.Yousefi, L. Liu, H. Li, "Head operated electric wheelchair," In Southwest Symp. on Image Analysis and Interpretation, IEEE, 2014, pp. 53–56.

[15] L. Cruz, D. Lucio, L. Velho, "Kinect and rgbd images: Challenges and applications," In SIBGRAP Conf. on Graphics, Patterns and Images Tutorials, IEEE, 2012, pp. 36–49.

[16] M. Figueiredo, L. Sousa, P. Cardoso, J. Rodrigues, C. Goncalves, R. Alves, "Learning technical drawing with augmented reality and holograms," In Int. Conf. on Education and Educational Technology, WSEAS, 2014, pp. 11–20.

About Author (s):



Ricardo Alves is currently finishing his masters degree on Electric and Electronic Engineering, Ricardo Alves is also a researcher for the University of the Algarve working with depth sensors. Ricardo, is one of the author of PoolLiveAid project, also spends some of his times developing other small electronics and programming projects.



Luís Sousa is a researcher at University of Algarve currently finishing his master degree in Electrical and Electronic Engineering. His major interests lies on electronic systems, embedded systems and computer vision, being PoolLiveAid one of his most recognized projects.



Aldric Negrier has a Masters degree in Eletronic and Telecommunication Engeniring by the University of Algarve, his main intrests are in the field of robotics, manufacturing processes and tecologies and in general software development.



Professor João Rodrigues graduated in Electrical Engineering in 1993, he got his M.Sc. in Computer Systems Engineering in 1998 and Ph.D. Electronics and Computer Engineering in 2008 from University of the Algarve, Portugal. He is Adjunct Professor at Instituto Superior de Engenharia, also in the University of the Algarve, where he lectures Computer Science and Computer Vision since 1994. He is member of LARSyS and the ISR in Lisbon. He participates in 13 financed scientific projects, and he is co-author more

than 100 scientific publications. His major research interests lies on computer and human vision, assistive technologies and human-computer interaction.



Professor Pedro Cardoso graduated in Mathematics/Computer Science from the University of Coimbra (Portugal) in 1996 , got his M.Sc. in Computational Mathematics from the University of Minho (Portugal) in 1999 and Ph.D. in Discrete Mathematics in 2007 from University of Seville (Spain). He is Adjunct Professor at Instituto Superior de Engenharia of the University of Algarve and a member of LARSyS (ISR - Lisbon). He is author of more than 30 scientific publications. His major research interests lies on operational research, databases, and human-computer interaction.



Professor Jânio Monteiro graduated in Electrical and Computers Engineering in 1995 from University of Porto, Portugal. In 2003 and 2010 he got a M.Sc. and Ph.D. degrees, respectively, in Electrical and Computers Engineering, from the Technical University of Lisbon, Portugal. He is an Adjunct Professor at Instituto Superior de Engenharia in the University of Algarve, where he lectures Telecommunication Networks since 1998. He is member of INESC-ID in Lisbon. He participated in several national and European scientific projects. His major research interests comprise Communication Networks, Video Transmission over IP and Smart Grid Protocols.



Miguel Gomes is a designer in SPIC - Creative Solutions, Loulé, Portugal [<http://www.spic.pt/>]



Paulo Bica is the CEO of SPIC - Creative Solutions, Loulé, Portugal [<http://www.spic.pt/>]