

Immersing Users in Landscapes Using Large Scale Displays in Public Spaces

Giannis Drossis¹, Antonios Ntelidakis¹, Dimitris Grammenos¹, Xenophon Zabulis¹
and Constantine Stephanidis^{1,2}

¹Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH)
N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Crete, Greece
{drossis, ntelidak, gramenos, zabulis, cs}@ics.forth.gr

²Department of Computer Science, University of Crete, Crete, Greece

Abstract. This paper reports on the design and implementation of BeThere-Now, a public interactive information system where users are depicted immersed in various sceneries. The work is focused on the domain of info-tainment in public spaces using large displays and aims on short-time usage. The implemented system employs a mixed reality application through which users are informed about different sceneries and also create personalized digital postcards. This process is accomplished using computer vision algorithms in order to depict users and objects, while removing the background of the scene. Finally, the lessons learned from the long-term deployment of the system out-in-the-wild are presented, providing an insight on the users' actions and reactions and feedback on future research directions.

Keywords: Large scale displays, interactive public spaces, immersive user experience, background segmentation, green screen, interactive postal cards

1 Introduction

Public spaces form an important part of our everyday life – they create a sense of belonging, provide a place where we can socialize, relax, and learn something new [3]. A public space is a social space that is generally open and accessible to people, involving necessary, optional and social activities [5].

Public displays are for anyone to interact in a walk-up-and-use [10] manner. In public displays, a large proportion of users are passers-by and thus first-time users. Most of the research on public displays has been carried out by running installations in local communities, yet this research has only recently started. Public spaces such as the bus stop or the cafe can act as ‘encounter stages’ on which people negotiate boundaries of a social and cultural nature.

Info-tainment [13] is a growing domain which applies to public spaces. Apart from approaches that rely on handheld displays, such as [11, 16], large-scale displays are widely applied for immersive user experience in public displays [12]. Postcards are a widely adopted approach for visitors to state their presence in a specific location. Nevertheless, they are passive and fail to natively incorporate the users' presence in a

scenery. On the contrary, an interactive system that could mix the real world (the users) with the virtual (the sceneries) could act as an informative system which is more pleasurable and desirable to use.

In this context, this paper describes the design and implementation of a mixed reality system employed in large public displays, in which users are immersed in numerous sceneries. By employing computer vision algorithms, users are depicted standing before the various sceneries, within the landscapes and the vistas projected, as if they were at that place. The system aims to inform and entertain multiple visitors or passers-by in a straightforward manner, while being personal: users see themselves in a large display and are able to seamlessly switch between different landscapes.

2 Related Work

Mixed reality (MR), sometimes referred to as Hybrid reality (encompassing both augmented reality and augmented virtuality) refers to the merging of real and virtual worlds to produce new environments and visualizations, where physical and digital objects co-exist and interact in real time. Mixed reality encompasses augmented reality and augmented virtuality as it does not take place in the physical or the virtual world, but in a combination of the two. Mixed and Augmented Reality is applied in various contexts, including:

- Games: augmented reality is applied to augment tabletop games [19] in order to preserve the physical artifacts of the game, but also enrich the user experience.
- Advertisement: advertising employs Augmented Reality to impress and attract users. Such an early example is MINI's advertisement [6], where users may show a magazine to their webcam in order to view a 3D model of MINI. Furthermore, [7] create an advertisement game in order to promote traditional products in a large scale wall display.
- Cultural Heritage: Mixed reality is applied to augment physical exhibits and allow users to retrieve additional information regarding exhibits of their interest. Grammenos et.al. [8] use pieces of paper that host additional information upon placement over areas of interest. Furthermore, Barry et. al. [1] use mobile devices to augment a museum's physical world with 3D representations of ancient livings, bridging nowadays with the prehistoric times.

Large scale displays are a common approach for visualizing information in public spaces [2, 9, 17]. A common approach for the display of users within landscapes is the adoption of the green screen [4] (such as in television weather reports). This tactic, however, holds two major drawbacks: firstly, user intervention is required in order to achieve satisfactory results – especially in scenes with lightning variations - but more importantly holds color limitations, as any green objects in the actual world will not be displayed in the final result.

Most of the times users enjoy a personal user experience. Moreover, information sharing and more specifically photographs are a widely used practice [14, 15] which users appreciate. Furthermore, photographs act as keepsakes and allow users to take

home the presented information or share them with their friends or relatives (person-to-person information sharing).

3 Be There Now!

3.1 The Concept

The concept of BeThereNow is based on the idea of creating an immersive application, where users are depicted standing in various sceneries as if they were there. The objectives of our work can be summarized as follows:

- The system should work like a **magic mirror**, where users watch themselves in the display, as if they were standing in various sceneries
- Users (**Foreground**) should be depicted in front of various landscapes (**Background**)
- Straightforward usage
- Apply meaningful, elegant and aesthetically pleasing means of changing the displayed sceneries
- Effective and precise background subtraction
- Users should be able to capture digital photographs and send them by email
- The system should not be static, but provide a demonstration mode which intrigues the users to engage with it



Fig. 1. The setup as designed in a 3D model prototype (a depth sensor is located at the top, a touch screen at the left, an interactive cube at the right and the projection in the middle)

The system is used in a walk up and use manner while also offering users the ability to browse through different sceneries. Therefore, once the users step in front of the display they are instantly shown. Meanwhile, a physical cube (Fig. 1, Fig 4) resides next to the display, which can be rotated in order to switch through the various scen-

eries. When the cube is rotated, a synchronized virtual cube - having the different landscapes at its sides - is also rotated, creating a one to one mapping of the actual with the virtual object. This metaphor is also applied when the demonstration mode is applied, where the virtual cube starts rotating and the landscapes are therefore switched after a short interval (30 seconds). The demonstration mode is enabled when no users are near the system and disabled upon any user entering the effective area in front of it. Users can use a secondary touch display, residing at the side of the display, in order to either change the language of the displayed image description or take a snapshot of themselves in the currently displayed scenery. The system's overall rendering process is separated in two primary sections: **Background** and **Foreground**. Background refers to the pictures of the sceneries used to immerse the users into, while foreground contains the users or objects that stand in front of the display. The background is in essence the virtual cube, while the foreground is estimated using a background removal process.

An additional optional capability of the system is the addition of **Foreground Mask** to sceneries: some areas of the background can appear in front of the users (e.g. objects such as a desk). This option enhances the reality of the settings as users can be immersed in a room, standing, for instance, in front of a desk and appear as if they are studying. Additionally, another usage example involves showing theatrical costumes (Fig. 5), which the visitors can virtually 'wear' by standing in such a position that their bodies are completely hidden and only some of their parts are visible. The overall rendering process is illustrated in Fig. 2 below.



Fig. 2. Rendering order of different application layers: **Background** contains the pictures of the background sceneries, **Foreground** refers to the picture of the users and **Foreground Mask** includes all the parts of the background that appear in front of the users

3.2 Implementation

Computer Vision Algorithm.

Silhouette extractor. The vision module employs a typical RGB-D sensor (i.e. Microsoft Kinect or an Asus Xtion camera). RGB-D sensors provide two images: a conventional RGB image and a depth image registered to the former one. Through the depth image, the 3D coordinates of surfaces in the RGB image are measured.

In case that the RGB sensor is judged insufficient in terms of quality or resolution, an additional high resolution RGB camera is employed, which is rigidly mounted on top of the RGB-D sensor. Then, color images are provided by the additional camera, while color images from the RGB-D sensor are disregarded. The aforementioned was

mainly the case for early RGB-D sensors but is not required for more modern ones (i.e. Kinect 2) as they also provide a high end color camera.

The sensor is placed so that covers the scene including the ground plane (see Fig. 1) and thereafter calibrated so as to estimate its relative posture to the scene. A computer vision component is responsible for two things. First it detects objects (persons) in the depth image, finds their outlying contours, and maps these contours to the color image. Then, these contours determine the portion of the color images to be displayed as foreground in the large scale display.

Calibration. Calibration is a two-step process. The first is conventional intrinsic and extrinsic, grid-based calibration [18]. If an additional color camera is employed, it is this one that is calibrated instead of that of the RGB-D sensor. Using this calibration, the location of an imaged 3D point in the depth image, is found in the color image.

The second determines a cuboidal volume in the scene, aligned with the ground plane, which we call out “working volume”. Only within this volume objects (persons) are considered by the system. To achieve this, the ground plane is estimated first, by imaging an empty scene and extracting the 3D points of the ground plane. By least-squares fitting of a plane to these points, the ground plane is approximated. The two lateral to the camera planes of the cuboid are perpendicular to the camera axis. The remainder cuboid faces are defined from the aforementioned planes, so that they limit the working space according to the sensor’s range so that very distal surfaces are not considered.

Foreground segmentation. The method finds the foreground in the color image, using the 3D information that the depth image avails. At each frame, input to the method is the depth and color images. The system finds surface points within the working volume in 3D, as well as, the regions where these surfaces occur in the depth image as silhouettes. These silhouettes are mapped to the color image to the foreground.

Due to sensor limitations, the depth image often exhibits pixels with missing values. This hinders foreground detection because the effect is intensely pronounced upon outlines, due to the depth discontinuity that they image; and, in our application, also due to human hair.

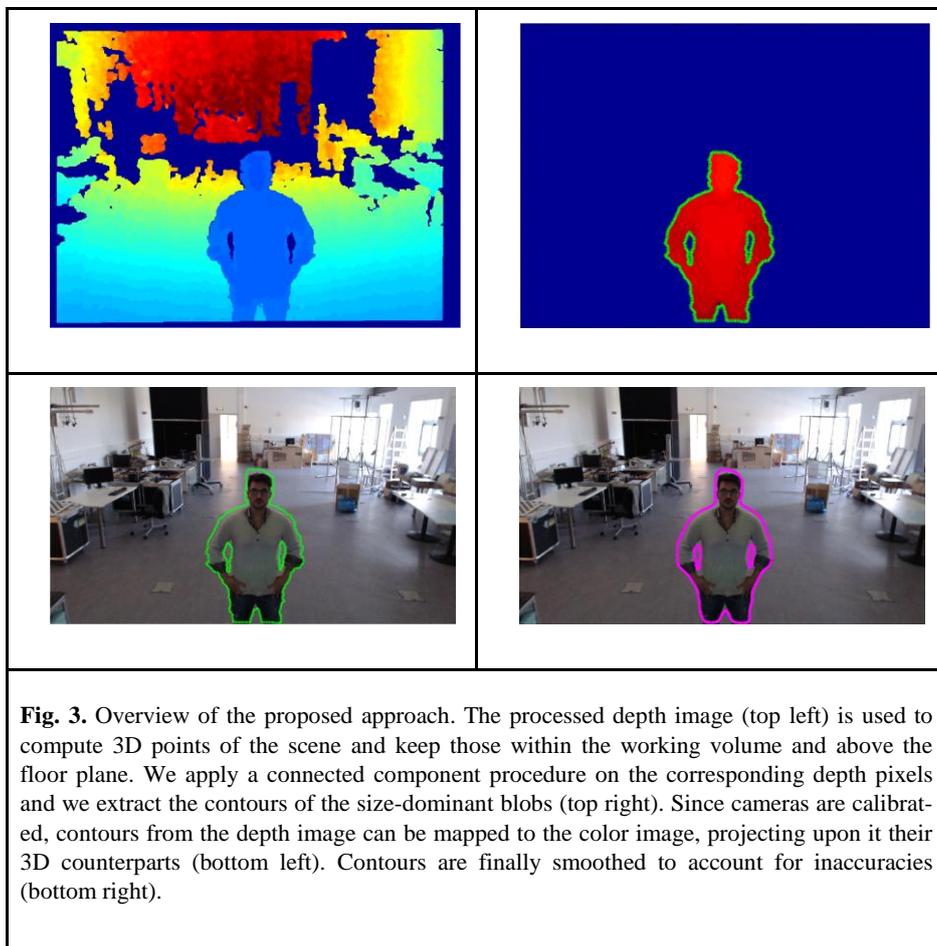
To reduce missing depth values, we apply nearest neighbour NN filling for such missing values. That is for an invalid pixel we assign the depth value of the nearest valid depth pixel with a neighborhood, if any, to the invalid pixel we currently evaluate. The output is the processed depth image which is henceforth called D .

We use image D to compute the 3D points of the scene within the working volume. These are foreground 3D points. The 2D pixels corresponding to these points define foreground mask M upon D . Next, a connected component procedure is employed on M in order to isolate the blobs corresponding to humans. Thus the blobs are filtered according to their size, excluding minute blobs that occur due to sensor noise. For each resultant blob we extract the external contours as well as the internal contours. Minute internal contours are attributed to sensor noise and are filled.

Using the calibration between the depth and color camera, the 3D counterparts of the contour pixels in the depth image are predicted in the color image. In this way, the

contours from the depth image are transferred to the color image. Note that due to the aforementioned NN preprocessing all such pixels have a depth value. We apply a 2D Gaussian smoothing on the transferred contours to account for possible inaccuracies. Areas in the color image encapsulated by the smoothed contours C map of the foreground in the depth image (see Fig.1 and Fig. 6).

The final step of the algorithm is to encode the result by creating an RGBA image. The RGB channels of this image are copied from the color image. The alpha channel contains the, aforementioned, mask of the foreground. Channel A is created by filling the external contours in C , but not the internal ones. In practice, the alpha channel is employed to restrict the color RGBA image so that only the foreground appears in it.



Technologies used.

The system was developed using Microsoft's XNA framework in combination with Windows Presentation Foundation (WPF). XNA is used for the rendering of the users

in the various landscapes, while the user interface is implemented using WPF. The computer vision component was developed using C++. Finally, an Arduino board equipped with a light sensor was used for the identification of the cube's orientation.

An optimization that was needed involved smoothing the persons' outline. Due to the sensor's inefficiency in accurately discriminating the foreground (the people's shapes) from the background (what lays behind the users in the real world), a custom smoothing algorithm was needed in order to fade out the boundaries of the foreground. An early implementation was developed in CPU but was later replaced by an HLSL shader, running in GPU, due to performance drop. The computer vision component provides the users' contours slightly inflated and thus containing a few pixels of the background as an outline. At rendering stage, the outline is faded out, using a neighbor-based algorithm, where each pixel's alpha value is the average of all its neighboring pixels.

4 Deployment in Public Spaces

BeThereNow is currently permanently installed in several public spaces, including ports, airports and museums. These installations include the Heraklion and Chania airports (since July 2013 and November 2013 respectively), Telloglion Foundation of Art (since January 2014) and Heraklion Municipal Info-point (since January 2015).

Additionally, it has been deployed in temporary public installations, including a travel trade show (ITB Berlin, 2014), promotional events (World Tourism Day 2013, Eleftheria Square, Heraklion, Crete) and conferences (HCII 2014, TEDx Heraklion, 2015).



Fig. 4. Heraklion Airport Installation

The visual output of the system was displayed using an ultra-short-throw projector, resulting in a projection 2.7m wide and 1.6m tall. Furthermore, a 12" touch monitor displaying a camera graphic was used to allow taking photographs. The touch monitor

was also used to change the language of the descriptions presented at the projection. The displays are accompanied by a physical cube, equipped with a rotation sensor, placed at the opposite side of the touch display (Fig. 1, Fig. 3). The cube also included magnets so as to provide a physical constraint and stabilize the rotations that the cube would end up to.



Fig. 5. Users interacting with the setup at Telloglion Foundation of art. Users are virtually wearing theatrical costumes (left) and sitting on chairs (right)

Sometimes people used their own cameras or smartphones in order to acquire a photograph, not noticing the system's related functionality. When more than one user was approaching the system concurrently the extra functionality was noticed by the users that were not distracted by having a camera in hand. In general, users in groups tended to focus on the entertaining part of the system, rather than exploring different sceneries. For example in Fig. 6, a girl is grabbing her friend to approach the system and take a photo together.



Fig. 6. A girl is insisting on taking a group photo, grabbing her friend by the arm and laughing after taking it.

In total, more than **21000 photographs** have been taken up to now using the system in permanent installations. The photographs provided extensive feedback on the

ways that people used the system. The following points summarize the ways that they used the system, sorted according to how often they were noticed. People:

- Tried to take a serious posture while looking as appealing as possible
- Made comical expressions or postures
- Pretended to interact with elements of the scenery (e.g. Fig. 5 - objects, animals, etc.)
- Stood still and serious, keeping a more disciplined posture
- Walked away from the system, hesitating to appear inside a photograph and resulting into an empty scenery

It is worth mentioning that children that interacted with the system always took more than one photograph, thus underlining the entertaining nature of the system. The same pattern was also noticed in the case of multiple users concurrently using the system, where not only they were photographed more than once, but also in different backgrounds.

5 Lessons Learned

On-site observation was used to gain an insight on people's actions and reactions using the system. The system was in general met with excitement by the users. People of all ages, ranging from young children to elderly, appeared to enjoy interacting with it. The vast majority of users could straightforwardly understand how to take photographs and manipulate the sceneries using the physical cube.

The main conclusions can be summarized as follows:

- Make interactive parts distinct and differentiate them from the system's physical setup, so that they are clearly noticeable
- Allow people to interact both with the system and with each other
- Users in groups tend to focus more on the entertaining and fun aspects of the system, rather than on information provision
- People share their personal information without second thoughts (e.g. email address) when they are going to receive something that includes their own image
- Keepsakes aid in creating a feeling of being personally engaged with the system

6 Conclusions

The response to the system by people of all ages was unanimously positive. The concept of immersing users in various landscapes created a positive emotion to both the people who used the system, but also to bystanders. Users were intrigued by the displayed sceneries to approach the system and were able to interact instantaneously, usually being surprised in a positive manner when seeing themselves on the large display.

The system's permanent installations served their purpose by both informing passers-by and entertaining them in various contexts, such as the airports' arrivals hall and museums. Users explored various local sightseeing locations and exhibited elements respectively, offering them the ability to become a part of the displayed vistas that would be otherwise unattainable, as if they were there. The system succeeded in drawing people's attention, both for urging them to engage with the system and for the bystanders, who watched other users being immersed in the projected landscapes.

Acknowledgements. The work reported in this paper has been conducted in the context of the AmI Programme of the Institute of Computer Science of the Foundation for Research and Technology-Hellas (FORTH). The authors would like to express their gratitude to Thanasis Toutountzis for his aid in manufacturing the rotating cube, as well as Manolis Apostolakis and Manolis Stamatakis who shaped the artistic concept and helped in the creation of the permanent installations.

References

1. Barry, A., Trout, J., Debenham, P., & Thomas, G. (2012). Augmented reality in a public space: The natural history museum, London. *Computer*, (7), 42-47.
2. Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT* (Vol. 3, pp. 17-24).
3. Carr S., Francis M., Rivlin L.G., and Stone A.M., *Public space*, Cambridge University Press, 1992.
4. Foster, J. (2014). *The green screen handbook: real-world production techniques*. CRC Press.
5. Gehl, Jan. *Life between buildings: using public space*. Island Press, 2011
6. *Geekology* (2008) <http://www.geekologie.com/2008/12/14-week>, Cool: Augmented Reality Advertisements, Dec. 19 (last accessed on 20/4/15)
7. Grammenos, D., Margetis, G., Koutlemanis, P., & Zabulis, X. (2012, October). Paximadaki, the game: Creating an advergame for promoting traditional food products. In *Proceeding of the 16th International Academic MindTrek Conference* (pp. 287-290). ACM.
8. Grammenos, D., Zabulis, X., Michel, D., Padeleris, P., Sarmis, T., Georgalis, G., Koutlemanis P, Veleni P. & Stephanidis, C. (2012). Macedonia from fragments to pixels: A permanent exhibition of interactive systems at the archaeological museum of thessaloniki. In *Progress in Cultural Heritage Preservation* (pp. 602-609). Springer Berlin Heidelberg.
9. Hornecker, E. (2008, October). "I don't understand it either, but it is cool"-visitor interactions with a multi-touch table in a museum. In *Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008. 3rd IEEE International Workshop on* (pp. 113-120). IEEE.
10. Izadi, Shahram, et al. "Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media." *Proceedings of the 16th annual ACM symposium on User interface software and technology*. ACM, 2003.
11. Keil, J., Pujol, L., Roussou, M., Engelke, T., Schmitt, M., Bockholt, U., & Eleftheratou, S. (2013). A digital look at physical museum exhibits. *Proceedings of the Digital Heritage*.

12. Lantz, E. (2007, August). A survey of large-scale immersive displays. In Proceedings of the 2007 workshop on Emerging displays technologies: images and beyond: the future of displays and interaction (p. 1). ACM.
13. Lucas, A. M. (1991). 'Info-tainment' and informal sources for learning science. *International Journal of Science Education*, 13(5), 495-504.
14. Neustaedter, C., & Fedorovskaya, E. (2009, April). Capturing and sharing memories in a virtual world. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1161-1170). ACM.
15. Nunes, M., Greenberg, S., & Neustaedter, C. (2008, February). Sharing digital photographs in the home through physical mementos, souvenirs, and keepsakes. In Proceedings of the 7th ACM conference on Designing interactive systems (pp. 250-260). ACM.
16. Madsen, C. B., Madsen, J. B., & Morrison, A. (2012, November). Aspects of what makes or breaks a museum ar experience. In *Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on* (pp. 91-92). IEEE.
17. Russell, D. M., Trimble, J. P., & Dieberger, A. (2004, January). The use patterns of large, interactive display surfaces: Case studies of media design and use for BlueBoard and MERBoard. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
18. Tsai R.Y., An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, FL, pp. 364-374, 1986.
19. Zidianakis, E., Antona, M., Paparoulis, G., & Stephanidis, C. (2012). An augmented interactive table supporting preschool children development through playing. *Proceedings of the AHFE International 2012*, 21-25.