AUGMENTED GALLERY GUIDE

Zuzana Haladova[(a)], Csaba Boylos[(b)]

[(a)] Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia
[(b)]Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia

[(a)]zhaladova@gmail.com, [(b)]bladeszasza@gmail.com,

ABSTRACT

The possibility of creating museum guides utilizing not just audio or textual information leads to emerging of new augmented and multimedia solutions. In this work in progress paper we want to propose Augmented Gallery Guide developed for the common smartphone device with the Android operating system. The guide combines the audio with the augmented reality and creates an emerging user experience. This paper discusses several technical issues of the creation of the guide, especially the recognition of the exponates.

Keywords: museum guide, augmented reality, cultural heritage, smartphones, local features

1. INTRODUCTION

The progress in the field of Augmented reality and the context aware systems since the 90's induce the strong interest in cooperation of computer scientist and the museums and galleries. The possibility of presenting the cultural heritage through new and popular devices, such as the smartphones became a way of attracting young people. The research in this area is nowadays mostly focused on extending the information about museum exhibits with virtual textual or visual information.

In this paper we will propose a novel concept of augmented reality guide adjusted for the use in galleries. Our guide will consist of augmented reality with synchronized audio.

In the area of the museum/gallery guides we can recognize several types of guides: a person, a book, an audio guide, a visual (interactive) guide and an augmented reality guide. We will now focus not only on the augmented reality guide but also on the visual (interactive) guide as it is closely connected with the first one. Both of these types can possibly be multimedia guides when engaging audio or another media.

The difference between these two methods is the fact that the visual (interactive) guide does not fulfil all 3 conditions on AR system proposed by Azuma [Azuma 1997]:

1. Combines real and virtual
2. Interactive in real time
3. Registered in 3D.

The visual guide solution usually interactively recognizes different exponates and display the virtual content, but they don't register the virtual objects within the real environment.

Many different implementations of the systems of one of these two types have been published since the first content-aware system presented in [Abowd et al. 1997]. We will focus on these in the following section. For completeness, we have to mention that there is also some research on Augmented audio guides, with no visual information, for example [Zimmermann and Lorenz 2008].

This paper is organized as follows. In the second section we will define the area of museum/gallery guides, specify different aspect of such systems and present different previous approaches. In the third section we will focus on different methods for the recognition of the exponates. In the following section we will propose the new Augmented gallery guide concept based on conclusions from the previous section. Next section will evaluate the system and formulate a conclusion. In the last sections we will focus on the future work and the acknowledgements.

2. PREVIOUS WORKS

In the previous section we have defined two museum guide types we will be focusing on: visual (interactive) guide and augmented reality guide.

We can divide the hardware solutions of these two approaches together on 3 types:

1. Head mounted
2. Spatial
3. Handheld.

The Head mounted solution such as [Flavia 2002] uses the Head mounted display to provide the user with the immersive experience. This device is usually owned by the museum and can be borrowed by the user. Problem with HMD is the ratio between the ergonomic parameters of the device and the resolution of the displayed augmented reality. Although several different types of HMD are known (Optical see through, Video see through, HMD Projectors, Retinal displays - for the details see [Bimber and Raskar 2005]), we think that the future applications will be developed for the devices such as Google glasses. The main advantage of the HMD concept is the hands-free setup.

The spatial augmented museum guide was proposed by [Kusunoki et al. 2002]. The guide consisted of interactive sensing board and was capable of recognizing different objects using RFID technology, which will be discussed in the next section. Other types of spatial augmented museum solutions were investigated by the O. Bimber [Bimber et al. 2006; Bimber and Raskar 2005]. This category encloses all the different solutions utilizing the transparent displays, mirror beam combiners or holographic set-ups. The main advantage of the spatial solutions is that they are usually suitable for more users cooperation, provide hands-free set-up and have theoretically unlimited field of view. They seem to be the best choice for static applications.

The last category encloses all the handheld solutions including smartphones, PDAs, pocket PCs, Palmtops, Tablets, netbooks and small notebooks. Different systems implemented for such device can be found in [Fockler and et al. 2005; Bay et al. 2006; Abowd et al. 1997; Miyashita et al. 2008; Bruns et al. 2007]. Handheld devices became more and more common and popular and their performance is increasing rapidly. Nowadays smartphones are equipped with quad cores processors and with their popularity and availability became the best platform for the guiding systems. Main disadvantage (according to [Bimber and Raskar 2005]) is the non-hands-free setup and the relatively small field of view of the device.

3. RECOGNITION OF EXPONATES

In this section we will deal with recognition of the exponates and different methods proposed in the previous works. The previous augmented and visual museum guide solutions can be divided by the method of recognition of the exponates on:

1. Visually based
2. Outside-in, Inside out systems
3. Dead-reckoning systems
4. Combination of systems
5. User input based

3.1. Visually based

The visually based approaches utilize the image from the camera to recognize the exponate and estimate its proper 3D position (when creating Augmented reality). There are three different approaches to the visual recognition of the exponates. In the first case, system utilize binary markers (e.g. black and white ARToolkit tags) which has to be printed and placed (registered) near the exponate ([Wagner and Schmalstieg 2003]).

The second approach is based on the matching of the local features in the camera frame with the preliminary acquired database of photographs of exponates ([Bay et al. 2006]). The method consists of the detection of the interesting points in the image (frame), their description by the feature vectors and the matching of these feature vectors with feature vectors from the objects (exponates) in the database.

The local feature methods usually used are based on the SURF [Bay et al. 2006] or other methods such as SIFT [Lowe 1999], ORB [Rublee et al. 2011] or combinations of different detectors (FAST [Rosten and Drummond 2006], Harris corners [Harris and Stephens 1988]) and descriptors (BRIEF [Calonder et al. 2010]).

The third approach consists of recognition of the exponates using global features (for example colour histograms, histograms of gradients...). As the representative of this approach we can mention *PhoneGuide: museum guidance supported by on-device object recognition on mobile phones* [Fockler and et al. 2005] which uses the global features and the neural networks for the recognition of exponates.

The visual methods usually require more computations, and the recognition is slower and not hundred per cent precise. On the other hand they are very cheap and more portable to different museums as only the photographs of all exponates are required (except the first one with binary markers).

3.2. Outside-in, Inside out systems

The emitter-receiver (or sensor) based approaches are used for the visual museum guides purposes as they do not provide us with the exact 3D position of the exponate in the space. These solutions are usually very precise however there is a need to place additional components (such as the emitters, receivers or sensors) in the museum area. The typical implementation of this approach uses the Infrared emitters and readers ([Flavia 2002]) or the RFID tags and readers ([Kusunoki et al. 2002]).

3.3. Dead-reckoning systems

In the paper Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera [Kourogi and Kurata 2003] Kourogi and Kurata proposed the indoor user dead reckoning (calculating current position by using a previously determined position) tracking system composed of an accelerometer, gyroscope, magnetometer, camera and a head tracker. The System is not dependent on any external markers, chips or sensors and determines the user's relative position from the variation of the vertical and horizontal acceleration (caused by human walking locomotion). To estimate the absolute position they used additional method of matching the camera stream with the database of images (prepared beforehand) utilizing the Kalman filter framework [Kalman 1960].

A system like this can be used in the museum for standalone recognition of exponates based on user position (not very robust). Another possibility is to provide the user position as an additional information to the visual recognition system.

3.4. Combination of systems

The fourth category encloses all the solutions which combine some of previous approaches. For example in [Bruns et al. 2007] the authors utilize the rough user position from the Bluetooth emitters placed in every room with the combination of exponate recognition using global features and neural networks. This method averages all of the positives and negatives of previous solutions. It is more accurate and faster than the visual methods and also fewer components need to be placed in the museum. On the other hand there is still the necessity of acquiring photographs of all the exponates.

3.5. User input based

The last category encloses the oldest solutions which utilize the user input instead of an automatic recognition. This method is not appropriate for the augmented reality guides, because it cannot register the exponate in 3D. However it will allow for the easy and straight forward implementation of an interactive guide without special exponate recognition algorithms.

When comparing different systems it is necessary to take another things into account: the price of the solution for the museum, necessity of additional components, portability and necessity of the pre-acquired photographs.

Apart from the type of the guide and the discussed recognition method we can mention three additional important criteria:

- The type of the information displayed to the user (text, image, video, 3D object),
- The type of the device (common handheld device/special components required),
- The necessity of sending the computations to the external server.

4. AUGMENTED GALLERY GUIDE

As we mentioned before our goal is to create an Augmented Gallery Guide. The main reason, why we decided to focus on the galleries instead of the museums is the fact that in galleries majority of exponates are paintings. The main advantage of paintings when compared to other exponates is that they are planar and it will be easier to properly register an augmented layer on them.

Based on the previous conclusions we have created several criteria on the Augmented Gallery Guide:

1. It has to run on common smartphones on Android platform
2. It does not require the installation of any additional components in the museum area,
3. It combines the augmented reality (as defined in [Azuma 1997]) with the synchronized audio comments and all the computations will be carried on the device.

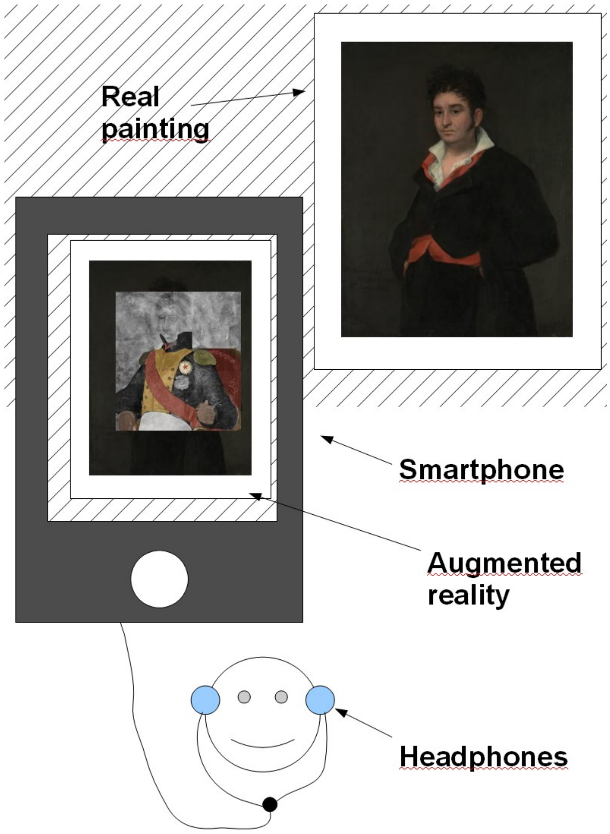

Figure 1: The scheme of the proposed Augmented Gallery Guide. The user holds a smartphone and wears the headphones connected to the smartphone. Camera on the smartphone streams recorded reality. The virtual information is augmented with the recognized and registered painting.

To fulfil all of the criteria we have to exclude all the spatial and head-mounted solutions. Also the system which recognizes the exponates with the sensor based or the combine solution and the visual solution using markers will be excluded. We have decided to use the local features for the recognition and registration of the exponates. We choose the FAST [Rosten and Drummond 2006] feature point detector and the BRIEF [Calonder et al. 2010] binary descriptor.

The main advantage of the BRIEF descriptor is the fact that it produces vectors of binary feature which can be easily matched using the Hamming distance metric (instead of the L2 norm commonly used for the matching). This causes that BRIEF descriptor can be matched very efficiently in comparison to SURF [Bay et al. 2006] or SIFT [Lowe 1999].

The main disadvantage however is, that BRIEF does not produce scale or rotation invariant descriptors. The rotational invariance is avoidable, and can be solved by utilizing the gyroscope of the smartphone. On the other hand the scale invariance is more important and can be solved by the storage of the paintings database in several scales.

The input of our system is a pre-processed database of the paintings we want to augment. For each painting we have a virtual image, object or video and the audio track stored. In the pre-processing phase we process the database of the paintings images and compute the keypoints using FAST detector and then feature vectors for all keypoints using BRIEF descriptor. For every image we have the file containing these feature vectors stored.

In the run time, our system works as follows: First the frame is grabbed from the video camera of the smartphone. Then the keypoints in the image are found using the FAST detector and the descriptors are computed using BRIEF. Afterwards these descriptors are matched with the database of the paintings' descriptors. A good match of two keypoints is estimated using the second nearest neighbour strategy. The painting in the frame is recognized as the database painting with most matches.

Then we use RANSAC algorithm [Rublee et al. 2011] to exclude the outliers and estimate the homography between the painting from the database and the one from the frame. We can then estimate the position of painting's corners and draw a quadrilateral on the frame. In this moment the user is allowed to trigger the (augmented) visual and audio content.

In the next frame after the recognition, the following loop starts. The frame is grabbed and the keypoints and the descriptors (feature vectors) of these keypoints are computed. These are then matched only with the descriptors of the database image recognize in the previous frame. RANSAC is used and homography is computed. This loop proceeds until the painting is still present in the image, i.e. the number of good matches exceeds a threshold.

When the painting is no longer presented in the frame, the visual content is no longer available. However the audio track is still proceeding until stopped by the user. The pipeline of the run-time can be found on Figure 2.

If we look closely on the application, in the first step the application's main thread starts two threads: one video thread which streams the input video from the camera, and one computing thread. The Video thread is needed to satisfy the user, because it shows always high frame per second (FPS) video.

Some attention has to be paid to the concept of the mixed augmented and audio solution. Our solution is mainly created for the augmentation of the paintings. If we want to provide the user with some additional interesting information on painting and also create an augmented experience on the small screen, we have to somehow cover the display with the text frames.

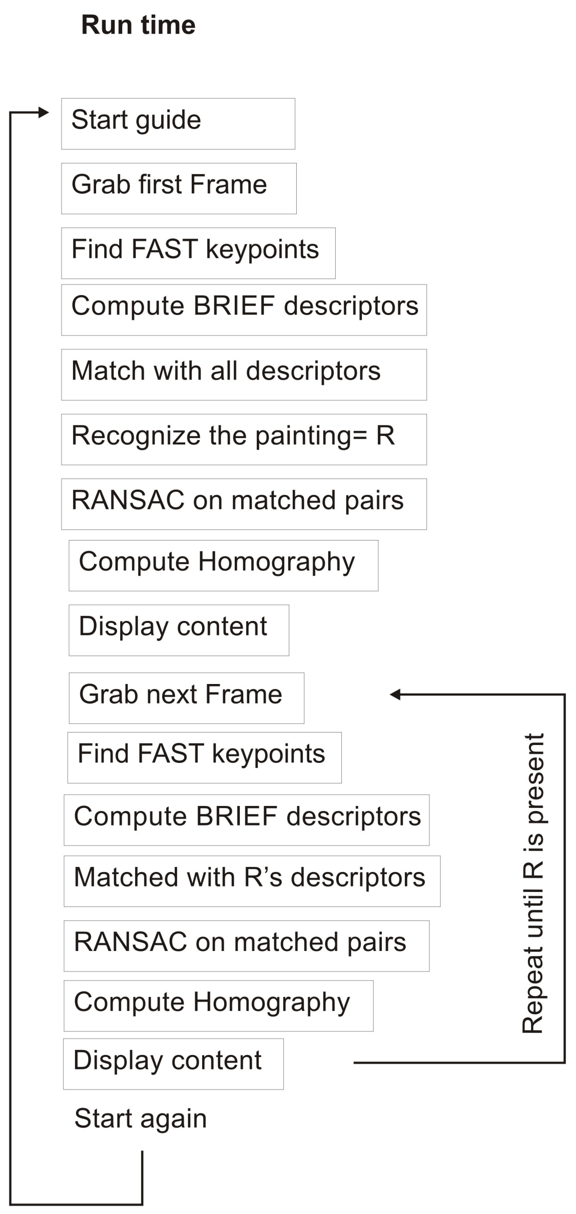

Figure 2: The pipeline of the run-time of the Augmented Museum guide system.

However this approach has no advantage compared to visual guide solution as it does not take advantage of the augmentation. On the other hand if we eliminate all the text information we can lose the role of the actual guide. To preserve both the augmentation and the guide at the same time and to utilize all the possibilities of the augmented reality guide we have decided to add the audio information. The scheme of our proposed system can be seen on the Figure 1.

Based on our tests of the application, we have also decided to create two different modes in our application. The first one is the proposed Augmented reality solution and the second one is the Virtual reality mode which can be displayed after the painting is recognized in the previous mode. This mode was created because it is not very comfortable to point your smartphone's camera to the painting for several minutes (to see displayed visual content while listening to the audio track). The virtual reality mode enable user to watch the paintings "augmentation" in the lastly processed frame.

5. CONCLUSION

In this paper we have proposed the Augmented Gallery guide system based on comparison of different systems and conclusions we have made. Proposed system utilize the recognition of the exponates (paintings) using the matching of the local feature vectors of paintings in the database and in the camera frame. We have implemented our prototype for the Android platform using OpenCV library. Our current prototype is running on 5 fps on the common smartphone.

We have decided to design the first solution for the small gallery as this eliminate two shortcomings. The first one is the fact that the preparation of the material (audio tracks, images...) for each painting is manual and though time consuming. The second shortcoming is connected with the matching of the feature vectors. To search the database of hundreds paintings can apparently slow down the application.

6. FUTURE WORK

In the next phase we want to complete the proposed Augmented Gallery Guide and provide the complex user study in which we want to focus on the several aspect of user gallery visit. Firstly we want to measure the time spend in the gallery by the user with Augmented Gallery Guide, book guide or no guide at all. In the second phase we will provide the visitors with the questionnaire containing questions about user experience with the guide, but also test questions. The goal is to investigate if the user can acquire more interesting information using a guide.

ACKNOWLEDGEMENTS

This work is partially funded from project KEGA 068UK-4/2011.

REFERENCES

Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R., and Pinkerton, M. 1997. Cyberguide: A mobile context aware tour guide. *Wireless Networks 3*, 421–433. 10.1023/A:1019194325861.

Azuma, R. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6, 4, 355–385.

Bay, H., Fasel, B., and Van Gool, L. 2006. Gool. Interactive museum guide: Fast and robust recognition of museum objects. *In Proc. Int. Workshop on Mobile Vision*.

Bimber, O., and Raskar, R. 2005. Spatial Augmented Reality: Merging Real and Virtual Worlds. *A K Peters/CRC Press*, July.

Bimber, O., Frohlich, B., Schmalstieg, D., and Encarnac ̧ ̃A O, L. M. 2006. The virtual showcase. *In ACM SIGGRAPH 2006 Courses*, ACM, New York, NY, USA, SIGGRAPH'06.

Bruns, E., Brombach, B., Zeidler, T., and Bimber, O. 2007. Enabling mobile phones to support large-scale museum guidance. *Multimedia, IEEE* 14, 2 (april-june), 16 –25.

Calonder, M., Lepetit, V., Strecha, C., and Fua, P. 2010. Brief: Binary robust independent elementary features. *In Computer Vision ? ECCV* 2010, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6314 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 778–792.

Flavia, S. 2002. The museum wearable: real-time sensor-driven understanding of visitors' interests for personalized visually augmented museum experiences. *In In: Proceedings of Museums and the Web* (MW2002, 17–20.

Fockler, P., and et al. 2005. PhoneGuide: museum guidance supported by on-device object recogn. on mob. phones. *In MUM '05: Proc. of the 4th intern. conf.*, ACM, USA, 3–10.

Harris, C. and Stephens, M. A combined corner and edge detection. *In Proceedings of The Fourth Alvey Vision Conference*, pages 147-151, 1988.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME- Journal of Basic Engineering* 82, Series D, 35–45.

Kusunoki, F., Sugimoto, M., and Hashizume, H. 2002. Toward an interactive museum guide system with sensing and wireless network technologies. *In Wireless and Mobile Technologies in Education*, 2002. Proceedings. IEEE International Workshop on, 99 – 102.

Kourogi, M., and Kurata, T. 2003. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. *In Proc. of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, 103–112.

Lowe, D. 1999. Object recognition from local scale-invariant features. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, 1150 –1157 vol.2.

Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., Gerl, O., Arnaudov, S., and Lieberknecht, S. 2008. An augmented reality museum guide. *In Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE Computer Society, Washington, DC, USA, ISMAR '08, 103–106.

Rosten, E., and Drummond, T. 2006. Machine learning for high-speed corner detection. *In Computer Vision ? ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds., vol. 3951 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 430–443.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. 2011. Orb: An efficient alternative to sift or surf. *In Computer Vision (ICCV)*, 2011 IEEE International Conference on, 2564–2571.

Wagner, D., and Schmalstieg, D. 2003. First steps towards handheld augmented reality. *In Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on*, 127 – 135.

Zimmermann, A., and Lorenz, A. 2008. Listen: a useradaptive audio-augmented museum guide. *User Modeling and User-Adapted Interaction* 18, 389–416. 10.1007/s11257-008-9049-x.

AUTHORS BIOGRAPHY

Zuzana Haladova was born in 1987 in Trnava, Slovakia. She received Msc. Degree in Applied Informatics- Computer graphics from Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava in 2010. She is a PhD. student on Informatics at the department of Applied Informatics at the same university. She is the member of IVG (Image and Vision group). Her research interests include Augmented reality and Object recognition. Currently she is working on her PhD. thesis called: Augmented reality in the cultural heritage applications.

Csaba Boylos was born in 1989 in Lučenec, Slovakia. He received Bc. Degree in Applied Informatics from Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava in 2010. He is a Msc. student of Computer graphics and Geometry at the same university. This work is connected with his diploma thesis. His research interests include Augmented reality on smartphones and Local feature based recognition.