Augmented Reality Indoor Environment Detection: Proof-of-Concept

RICARDO J.M. VEIGA[1], JOÃO A.R. PEREIRA[1], JOÃO D.P. SARDO[1], ROMAN BAJIREANU[1], PEDRO J.S. CARDOSO[1], AND JOÃO M.F. RODRIGUES[1],

[1] LARSyS (ISR-Lisbon) & ISE, University of the Algarve, 8005-139 Faro, Portugal
{rjveiga,japereira,jdsardo,rbajireanu,pcardoso,jrodrig}ualg.com

Abstract: - The conventional museum experience offers the visitors glimpses of the past with the narrative limited to the static art that garnishes it. Through technology we already can mix the past with the future, immersing the visitors in a true dynamic journey across the same walls that guard our history. One of these technologies is the Augmented Reality, which aims to enhance our surroundings into a new era of creativity and discovery. This paper presents the proof-of-concept of an indoor portable environment pose estimation module (PEPE) present inside M5SAR, a project that aims to develop a five senses augmented reality system for museums. The current state of development of this module shows that is already achievable real-world wall(s) detection and a new environment superimposition over the detection, i.e., it is now possible to have a dynamic museum experience with the ability of transforming rooms into historic live stages.

Keywords: - Augmented Reality, Superimposition, Indoor Localization.

1 Introduction

Augmented Reality (AR) [2] has benefited from the increased hardware capabilities of smartphones and novelty algorithms, resulting in a fast evolution over a short time, rapidly growing its number of users. It allows for a higher level of interaction between user and real-world objects, expanding this experience and creating a brand-new level of edutainment. The M5SAR: Mobile Five Senses Augmented Reality System for Museums project [29] aims for development of an AR system that acts as guide for cultural, historical and museum events. Most museums have their own mobile applications (App), see e.g. [17, 37], and some also have AR applications, see e.g. [14, 26, 33, 38]. The innovation in the M5SAR project is to extend the AR to the human five senses, see e.g. [29] for more details.

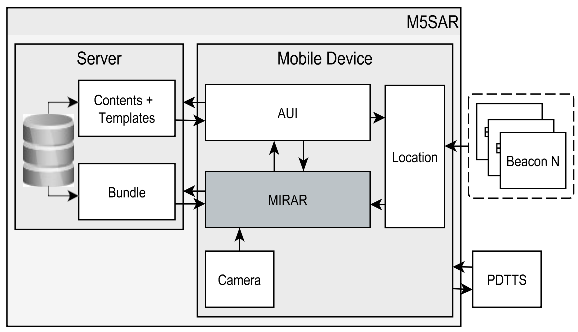
The Mobile Image Recognition based Augmented Reality Framework (MIRAR) framework is one of the modules of M5SAR project[25], aims to: (a) perform all computational processing in the client-side (mobile device); (b) use in real world with 2D and 3D objects as markers for the AR; (c) recognize environments, i.e., walls and its respective boundaries; (d) detect and segment human shapes; (e) project contents (e.g., text and media) onto different objects, walls and persons detected and displayed in the mobile device's screen. A framework that integrates these goals is completely different from the existing (SDK, frameworks, content management, etc.) AR systems [1, 6, 19, 20, 24].

This paper focus on one of the MIRAR sub-modules (sub-module c), the environment detection and overlapping of information. Considering a typical museum wall, there is usually artwork such as paintings and tapestry hanging on the walls, creating a unique rich environment full of visual information. Following the previous method introduced on the main object recognition module of MIRAR [25], we will use the features detection and description matching methods for the environment recognition. Considering the expect walls as planes, and taking into account the limited input information obtained from a monocular camera and smartphone performance, any methods of 3D matching, such as bundle adjustments, iterative closest point, among others, were discarded. Furthermore, with planes, it is possible not only to perform a faster recognition using the same methods used for object recognition, but also use the vanishing lines provided by the common geometric rules, for which we considered the existence of paintings' frames as a guarantee for the existence of vanishing lines.

In this paper the contribution is to fuse both approaches in order to achieve a better wall detection and also user's localization so that we can more accurately project content upon the walls through the use of AR superimposition.

The MIRAR sub-module for object recognition and environment detection presented in this paper is AR marker-based, often also called image-based [7]. AR image-based markers allow adding easily detectable preset signals in the environment, using computer vision techniques to sense them. There are many image-based commercial AR toolkits (SDK) such as Catchoom or Kudan [6, 19], and AR content management systems such as Catchoom or Layar [6, 20], including open source SDKs [1]. Some are expensive, others consume too much memory (and the present application will have at least one marker for each museum piece), while others load slowly on the mobile device. The increasing massification of AR applications brings new challenges, such as the demand for planar regions detection (walls), with the more popular being developed within the scope of Simultaneous Localization And Mapping (SLAM) [3, 9]. RGB-D devices or light detection and ranging (LIDAR) sensors [16, 27, 41, 34] usually used for image acquisition of 3D environments. Some advances within environment detection, localization or recognition include using Direct Sparse Odometry [11], or using descriptors, like ORB SLAM [22] or even Large-Scale Direct Monocular SLAM [12]. However, the MIRAR framework focuses on mobile devices with monocular cameras only. Following this, an initial study of an environment detection sub-module was presented in [25], using a geometric approach to the extracted edges of a frame. A frame is always captured from a perspective view of the surrounding environment, with the usual expected environment being characterized by the existence of numerous parallel lines which converge to the vanishing point [8, 32].

The paper is structured as follows: The MIRAR framework and architecture is introduced in Sec. 2, followed by the environment detection and AR overlapping at Sec. 3 and finally concluding with a final discussion and future work, Sec. 4.



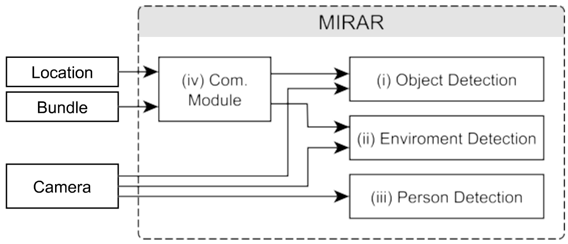


Fig. 1. Top: overall simplified system architecture. Bottom: MIRAR block diagram.

2 MIRAR Framework

The M5SAR system is shown on top of Fig. 1. From the left to the right we have the basic server-client communications, in our case a mobile device, where all the computer vision processing is computed, and communicated directly without Portable Device for Touch, Taste and Smell Sensations (PDTTSS) [31] used to enhance the five senses, and the displayed Beacons [13] are employed in the user's localization.

The M5SAR App architecture is divided into three main modules: (A) Adaptative User Interfaces (AUI), see [29]; (B) Location module, a detailed explanation is out of this paper's focus, and (C) MIRAR module (see Fig. 1 bottom).

The MIRAR divides in four features: (a) the detection and recognition of museum objects; (b) the detection, recognition and tracking of objects as the user moves along the museum, allowing to interact digitally the objects displayed, MIRAR sub-module (i); (c) detection, recognition, and superimposition of content over the museum walls, related with the recognized object's epoch, sub-module (ii); (d) detection of persons that are moving through the museum, replacing the everyday wear with clothes from the object's epoch, sub-module (iii).

Also, given that the sensor used to acquire the images from the environment is the smartphone's camera, to save battery, the camera is only activated when the AR option is selected in the UI. When the activation occurs, the user can see the environment in the mobile screen and effectuate the previously mentioned actions. As an additional effort to limit battery use, the device will enter a low-power state if the user turns the phone upside down, by dimming the phone's screen and interrupting the processing.

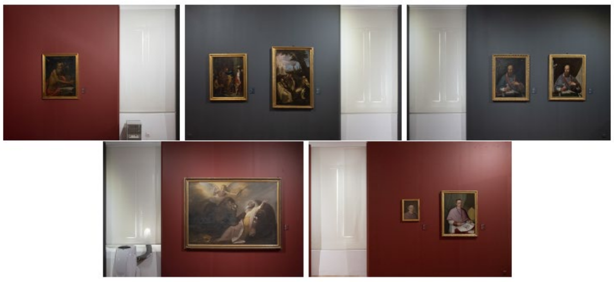


Fig. 2. From top to bottom: Example of five templates following each other.

3 Environment Detection

The conventional museum's environment is rich in details provided by the multiple artwork that embellishes it. This scattered information is always present along the visitor's navigation throughout the museum when in presence of artwork. In continuity of our previous work presented in former publications (see e.g. [28, 25, 30, 31]), the vast presence of unique features along the museum allows us not only to be able to superimpose content over the walls, but also to locate the visitor's position within the museum. The visitor's localization is also used within our main module of object recognition but using Bluetooth's beacons instead.

Also, in previous papers [30, 39] two distinct approaches were presented to solve the environment detection, one focusing on the geometry shape of a regular museum's division, assuming that all the vanishing lines are presented by the present walls within such division; while the other focuses on the recognition of already known parts of said walls. In this paper both methods are fused together in order to achieve a harmonious detection, recognition and localization of the environment, to dynamically superimpose different types of content over the walls, such as images, video, animations, or 3D objects.

Before continuing, it is important to remind that due to the necessity of regular cuboid rooms and image recognition, the method presented in this paper is intended to be only used on previously scanned and prepared environments. It is also relevant to remind that the purpose of this AR application is to be able to run seamlessly on any current monocular smartphones, from which only a RGB image is provided by the camera, without any additional depth information.

The current algorithm divides itself in four different stages: the bundle creation (a), the recognition and localization (b), tracking (c), and superimposition (d).

It is self-explanatory that the first stage is not performed inside the runtime, see below for a detail explanation, while the other two complement each other. All of the results presented were obtained running on an Intel i7-4820K CPU running on a single-thread. Beginning with the bundles' creation (a), for this task there are two distinct types of bundles created: a FLANN Index (FI) bundle, and a FLANN Based Matcher (FBM) bundle. The reason for this peculiar choice is based on performance evaluations made while testing the multiple alternatives available, being the Brute Forced Matcher out of the scope of this paper, due to its lack of "flexibility" present on a previous publication. Both methods used the same index parameters, with the chosen algorithm being the Locality-Sensitive Hashing (LSH), due to the choice of using non-patent binary descriptors, the number of tables used were only 1, with a key size of 12, and only 1 multiprobe level. The addition of a multiprobe to the LSH allowed for the reduction of the number of hash tables, which allow for a better performance while maintain the same obtained results. We observed an average reduction of 76.56% of processing time across different binary features detectors and descriptors (AKAZE,

Features	Number of Features	FLANN Based Matcher	FLANN Index	Performance Difference	Parameters
AKAZE	9500	10.62 ms	3.09 ms	-70.90 %	Default
BRISK	14521	22.64 ms	10.19 ms	-54.99 %	Default
BRISK	11299	17.01 ms	3.5 ms	-79.42 %	thresh=30, octaves=5, patternScale=float(2.0)
ORB	10891	11.72 ms	5.35 ms	-54.35 %	Default
ORB	35311	40.62 ms	22.90 ms	-43.62 %	nFeatures=2000

Table 1. Comparison of the performance between FLANN Based Matched and FLANN Index, presenting the results obtained from the matching of a real-world image to an index of 71 prepared images.

BRISK, ORB) [35], with the default and tweaked specifications, while using only 1 hash table versus the 6 hash tables originally recommended, with the corresponding images' indexes returned with equal accuracy.

Regarding the choice of having two bundles of similar matchers, although the FBM is built upon the FI, we performed search tests with the same query image on both and obtained a better result retrieving the matching image index by an average of 60.66% less processing time while using the FI, as can be seen at table 1. This justifies the creation of an FI bundle, although while matching using the FI only the original image index is retrieved, accompanied by the KNN's distances. This way, it is only possible to know what image was matched but it is impossible to find the homography of said image with the queried one, which prevents the possibility of user's localization. In order to contour this limitation, a second bundle was created. With the Flann Based Matcher, the matches obtained are correlated between the trained index and the queried image. Furthermore, in our tests the Flann Based Matcher, while using a single image, matched with an average processing time of 5.5 ms. This allows for an initial faster and broad user's localization within the museum environment, followed by a more specific approach once the localization is found. It is important to notice that, with our method, even when adding the FI and the posterior FBM processing time, it is still faster comparing to the only use of the FBM.

An additional method was also analyzed based on ASIFT. Due to the nature of the application, it is expected that the users explore the superimposed content not only frontal-facing to the walls, but also shifting the smartphone to the side, which creates an image perspective more difficult to match. With the ASIFT algorithm we expected to explore the additional affine matching while using the FLANN index matcher to maintain an acceptable performance with the new additional descriptors. Unfortunately, the obtained results, while successful, returned a large reduction of matched indexes, in some cases more than 13 times less. For this reason, further tests and analyses will be performed and presented on a future publication.

Regarding the templates used to train the FLANN indexes, it was observed while advancing the presented algorithm that the wooden frames of the museum's artworks represented a large part of the retrieved features from the images, as can be seen on Fig. 3. When implemented, it was verified that the wooden frames' descriptors matched vastly between themselves across different artworks, which introduced plenty of false positive matches. To prevent this result, we applied masks over the templates, as can also be seen on Fig. 3, allowing only the features present on the artworks to be computed as descriptors. This improved the performance and reduced the observed false positives. The nature of the shape and form of the template's images will be further explained at the finding homography step. In continuity of our previous work, the height of the templates was limited to 480 pixels, which is the obtained height of the smartphones' camera frame, and the detector and descriptor remains the BRISK, which allows for some image scaling in the recognition, which is expecting to occur while the users navigate the museum.

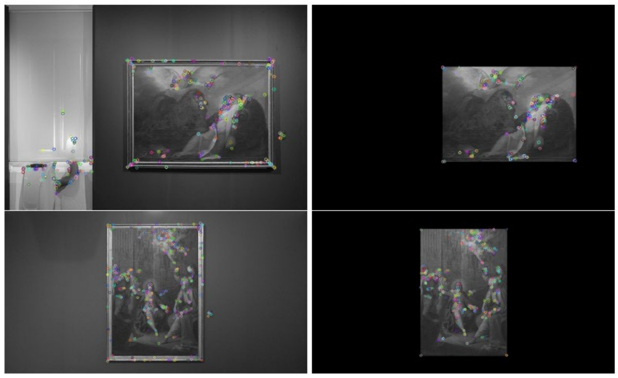


Fig. 3. Top and Bottom: Example of the different amount of keypoints once the painting's frame is removed through the use of masks.

Advancing to the recognition and localization stage (b) of the algorithm, while retrieving the frames from the user's smartphone camera, if there isn't a previous matched frame, the FLANN Index is used to find the corresponding image's index. As the FI usually returns the more approached image to the frame's descriptors, it is always necessary to perform at least the corresponding FLANN Based Matcher of the obtained image index to discard the insufficient matches. It was observed that the amount of returned matches from the FI is not correlated to the certainty of the retrieved match. Nevertheless, this method continues to be faster than a plain FBM use. All the matches obtained from the FBM are subjected to the Lowe's ratio test, where only the matches with distances inferior to a relation of 0.65 are considered good matches. When a match is found with at least 10 good matches, then we proceed to find the homography. Normally only 4 matches are needed for the homography calculation, but as mentioned on previous work, for this AR application it is mandatory the computation of a good homography, and therefore the number of minimal good matches was increased. The following homography refinement method can be found on our previous publication, with the addition of a symmetry test and also a verification if the return matrix isn't transposed, being this way possible to salvage some bad outputs [36]. Having in mind the necessity of a considerable amount of good matches but also a smooth performance, the amount of descriptors detected from the frame is directly associated with the previous frame processing time, with all being firstly sorted by their response parameter.

As referred before, the templates' shape form was made with a purpose. When calculating the homography we found the perspective relation between two different planes: an image in the 2D world, and an object in the 3D world. Due to the similar construction structure between different smartphones, we were able to observe a limited variation in the intrinsic camera matrix, which allow us to assume a acceptable outcome within error, if needed we might implement an auto-calibration method as future work [21]. In order to reduce additional computing calculations, when the template's images where obtained, the corresponding wall height was included, which allows for a direct relation between the artworks full of features and the plain walls that lack them. It is important to refer that a panoramic of all the walls of a museum room was also created, but giving the vastness of the information, the panoramic reduced the performance and increased the false positives matches, therefore increasing the amount of bad homographies computed. The current arrangement of templates covers completely the walls of the museum room with at least two artworks always present, with the exception of large artwork pieces. Using these limits, and with each template preceding the other and never overlaying, it is possible to retrieve the already known shape of the room without the need of advanced 3D calculations.

With the homography known, the next steps are the fusion of the previous two methods presented on former publications. A Gaussian Blur is applied to the obtained frame. A dynamically adjusted Canny edge detection [5] is applied to the camera's frame, using the Otsu threshold [23] to replace the high Canny's threshold while the lower varies with a direct proportion of 10% to the higher.

From there, the Probabilistic Hough Transform [15, 18] is applied in order to retrieve the presented lines in the environment, as can be seen in Fig. 4.

Fig. 4. Example of the lines found in the environment through different perspectives.

The Line Segment Detector was also considered, but it presented a performance 3 times worse for the same amount of lines retrieved. The obtained lines are then filtered with the vertical and horizontal lines being discarded relatively to the horizon line. Afterwards the similar lines are removed, remaining only the unique lines, expected to be the environment vanishing lines. The intersecting points between these lines are calculate following the Cramer's Rule. The obtained intersecting points are added to a k-means clustering, where the densest cluster is chosen, and its centroid is considered as vanishing point. The already found lines from the homography are then adjusted to the obtained lines corresponding to the walls' horizontal delimitations, improving the already refined homography.

Following the last steps comes the stage of tracking (c). As referred in previous publications, it is not expected the possibility of always achieving a valid match with the templates. In order to be able to continue tracking the user's navigation it is necessary to deploy different methods for confirming the user's actions. The direct method is to continue tracking the matching image and the ones surrounding it to the left and right. We also used the retrieved homography perspective to generate a mask which is used while the same image is matched, which allows discarding unnecessary descriptors from the frame. With the fusion of both methods previously presented we are able to use a novel approach, where we apply Kalman Filters to the vanishing point and corresponding points of the found and adjusted homography through vanishing lines. This method allows for a better perception of the user's movement and smooths the transitions of the superimposed content.

Finally, we reach the last stage, the superimposition (d). Although it was already possible after the second stage, we decided it was a higher priority to first start tracking so we could evaluate the initial tracking frames, and after a small amount of good tracked frames, initializing the projection of content over the walls. Considering the purpose of the users' visit being the museum's artwork, to be able to superimpose contents while maintaining the artwork visible, the templates' masks already generated and used while building the bundles, are used here. With all the templates following each other, an example can be seen of Fig. 2; we are able to find where the corresponding vanishing line end and another perpendicular wall commences, allowing us to generate a perspective matrix corresponding to the projected wall(s). With this information, we can project specific content on different walls throughout the museum's rooms. A desired result is presented on Fig. 5.

Fig. 5. Left to right: the desired segmentation of the environment's walls, two examples of superimposing results.

4 Conclusions

The current state of this module shows promising results, presenting the fusion of two different methods previously introduced that allow for a better filtering and recovery of bad homographies, while introducing an additional geometric tracking method. With the possibility of acquiring mainly good homographies, it is possible to consider the calculations of the user's camera pose on the real world [10, 4, 40], which in future work is being considered to be projected into a 2D map of the museum and the localization and direction of the users being computed using Kalman Filters to reject the remaining bad homographies.

The presented form of the templates are considered the final version, with the complete wall height and shape in the templates being used to retrieve the walls' horizontal limits and localization while also using masks to discard the unwanted features retrieved from the paintings' frames, and being continuous to each other in the real world, allowing the calculation of an accurate perspective matrix in order to superimpose content.

Regarding the search and matching between templates and the camera frame, this was also addressed with the introduction of a mixed FLANN indexes search engine, which has shown excellent time results and allows for a faster and broader localization while remaining with a more specific matching intended for the AR superimposition method.

For future work, the unexpected occurrence of few returned indexes while adding and training using the ASIFT method to the current algorithm could be further explored and evaluated. Although, with the environment's vanishing lines, it is also possible to continue the tracking into more obtuse view perspectives, as can be seen on Fig. 4, which could disprove the necessity of properly implementing the ASIFT method.

With the expected algorithm fully implemented, a battery of tests shall be produced to evaluate the performance and quality of this module in real-time and introducing additional rooms with different configurations.

Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT), project LARSyS (UID/EEA/50009/2013), CIAC, and project M5SAR I&DT nr. 3322 financed by CRESC ALGARVE2020, PORTUGAL2020 and FEDER. We also thank Faro Municipal Museum and the M5SAR project leader, SPIC - Creative Solutions [www.spic.pt].

References:

[1] Artoolkit. ARtoolKit, the world's most widely used tracking library for augmented reality. http://artoolkit.org/, 2017. Retrieved: Nov. 16, 2017.

[2] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. IEEE Computer Graphics and Applications, 21(6):34–47, 2001.

[3] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II. IEEE Robotics & Automation Magazine, 13(3):108–117, 2006.
[4] Adrien Bartoli and Peter Sturm. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. Computer vision and image understanding, 100(3):416–441, 2005.
[5] John Canny. A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6):679–698, 1986.
[6] Catchoom. Catchoom. http://catchoom.com/, 2017. Retrieved: Nov. 16, 2017.
[7] Kun-Hung Cheng and Chin-Chung Tsai. Affordances of augmented reality in science learning: Suggestions for future research. Journal of Science Education and Technology, 22(4):449–462, 2013.
[8] Wenting Duan. Vanishing points detection and camera calibration. PhD thesis, University of Sheffield, 2011.
[9] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part I. IEEE robotics & automation magazine, 13(2):99–110, 2006.
[10] Ali Elqursh and Ahmed Elgammal. Line-based relative pose estimation. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 3049–3056. IEEE, 2011.
[11] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. IEEE transactions on pattern analysis and machine intelligence, 2017.
[12] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-scale direct monocular slam. In European Conference on Computer Vision, pages 834–849. Springer, 2014.
[13] Estimote. Create magical experiences in the physical world. https://goo.gl/ OHW04y, 2017. Retrieved: April 04, 2017.
[14] HMS. Srbija 1914 / augmented reality exhibition at historical museum of Serbia, Belgrade. https://vimeo.com/126699550, 2017. Retrieved: April 04, 2017.
[15] Paul VC Hough. Method and means for recognizing complex patterns. Technical report, 1962.
[16] Rostislav Hulik, Michal Spanel, Pavel Smrz, and Zdenek Materna. Continuous plane detection in point-cloud data based on 3D Hough transform. Journal of Visual Communication and Image Representation, 25(1):86–97, 2014.
[17] InformationWeek. Informationweek: 10 fantastic iPhone, Android Apps for museum visits. https://goo.gl/XF3rj4, 2017. Retrieved: April 04, 2017.
[18] Nahum Kiryati, Yuval Eldar, and Alfred M Bruckstein. A probabilistic hough transform. Pattern recognition, 24(4):303–316, 1991.
[19] Kudan. Kudan computer vision. https://www.kudan.eu/, 2017. Retrieved: Nov. 16, 2017.
[20] Layar. Layar. https://www.layar.com/, 2017. Retrieved: Nov. 16, 2017.
[21] Paulo RS Mendon¸ca and Roberto Cipolla. A simple technique for self-calibration. In cvpr, page 1500. IEEE, 1999.
[22] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. IEEE Transactions on Robotics, 31(5):1147–1163, 2015.
[23] Nobuyuki Otsu. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1):62–66, 1979.
[24] Luís Pádua, Telmo Adão, David Narciso, António Cunha, Luís Magalhães, and Emanuel Peres. Towards modern cost-effective and lightweight augmented reality setups. International Journal of Web Portals (IJWP), 7(2):33–59, 2015.
[25] João AR Pereira, Ricardo JM Veiga, Marco AG de Freitas, JDP Sardo, Pedro JS Cardoso, and Joao MF Rodrigues. Mirar: Mobile image recognition based augmented reality framework. In International Congress on Engineering and Sustainability in the XXI Century, pages 321–337. Springer, 2017.
[26] Qualcomm. Invisible museum. https://goo.gl/aS0NKh, 2017. Retrieved: April 04, 2017.
[27] James Ring. The laser in astronomy. New Scientist, 18(344):672–673, 1963.
[28] J.M.F. Rodrigues, J. Lessa, M. Gregório, C. Ramos, and P.J.S. Cardoso. An initial framework for a museum application for senior citizens. In. Proc. 7th Int. Conf. on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 2016.
[29] J.M.F. Rodrigues, J.A.R. Pereira, J.D.P. Sardo, M.A.G. Freitas, P.J.S. Cardoso, M. Gomes, and P. Bica. Adaptive card design UI implementation for an augmented reality museum application. In Proc. 11th Int. Conf. on Universal Access in Human-Computer Interaction, integrated in the

19th Int. Conf. on Human-Computer Interaction, 2017.

[30] João MF Rodrigues, Ricardo JM Veiga, Roman Bajireanu, Roberto Lam, João AR Pereira, Joao DP Sardo, Pedro JS Cardoso, and Paulo Bica. Mobile augmented reality framework-mirar. In International Conference on Universal Access in Human-Computer Interaction, pages 102–121. Springer, 2018.

[31] J.D.P. Sardo, J. Semião, J.M. Monteiro, J.A.R. Pereira, M. Freitas, J.M.F. Rodrigues, and E. Esteves. Portable device for touch, taste and smell sensations in augmented reality experiences. Accepted for the INternational CongRess on Engineering and Sustainability in the XXI cEntury – INCREaSE 2017, 2017.

[32] M. Serrão, S. Shahrabadi, M. Moreno, J. T. José, J. I. Rodrigues, J. M. F. Rodrigues, and J. M. H. du Buf. Computer vision and GIS for the navigation of blind persons in buildings. Universal Access in the Information Society, 14(1):67–80, 2015.

[33] SM. Science museum - atmosphere gallery. https://vimeo.com/20789653, 2017. Retrieved: April 04, 2017.

[34] L. Sousa, JMF Rodrigues, J. Monteiro, PJS Cardoso, J. Semião, and R. ALVES. A 3D gesture recognition interface for energy monitoring and control applications. Proceedings of ACE'14, pages 62–71, 2014.

[35] Shaharyar Ahmed Khan Tareen and Zahra Saleem. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on, pages 1–10. IEEE, 2018.

[36] Giorgos Tolias and Yannis Avrithis. Speeded-up, relaxed spatial matching. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 1653– 1660. IEEE, 2011.

[37] TWSJ. The wall street journal: Best apps for visiting museums. https://goo.gl/ cPTyP9, 2017. Retrieved: April 04, 2017.

[38] Natalia Vainstein, Tsvi Kuflik, and Joel Lanir. Towards using mobile, head-worn displays in cultural heritage: User requirements and a research agenda. In Proc. 21st International Conference on Intelligent User Interfaces, pages 327–331. ACM, 2016.

[39] R.J.M. Veiga, R. Bajireanu, J.A.R. Pereira, J.D.P. Sardo, P. J.S. Cardoso, and João M.F. Rodrigues. Indoor environment and human shape detection for augmented reality: an initial study. In Procs23rd edition of the Portuguese Conference on Pattern Recognition, Amadora, Portugal, 28 Oct., pp. 21., 2017.

[40] Etienne Vincent and Robert Laganiére. Detecting planar homographies in an image pair. In Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on, pages 182–187. IEEE, 2001.

[41] Junhao Xiao, Jianhua Zhang, Benjamin Adler, Houxiang Zhang, and Jianwei Zhang. Three-dimensional point cloud plane segmentation in both structured and unstructured environments. Robotics and Autonomous Systems, 61(12):1641–1652, 2013.