Artwork Identification in a Museum Environment: A Quantitative Evaluation of Factors Affecting Identification Accuracy

A.Lanitis[1,2][0000−0001−6841−8065], Z. Theodosiou[2][0000−0003−3168−2350], and H. Partaourides[2][0000−0002−8555−260X]

[1] Visual Media Computing Lab, Dept. of Multimedia and Graphic Arts Cyprus University of Technology Limassol, Cyprus
[2] Research Centre on Interactive Media, Smart Systems and Emerging Technologies, Nicosia, Cyprus
andreas.lanitis@cut.ac.cy, Z.Theodosiou@rise.org.cy, H.Partaourides@rise.org.cy

Abstract. The ability to identify the artworks that a museum visitor is looking at, using first-person images seamlessly captured by wearable cameras can be used as a means for invoking applications that provide information about the exhibits, and provide information about visitors' activities. As part of our efforts to optimize the artwork recognition accuracy of an artwork identification system under development, an investigation aiming to determine the effect of different conditions on the artwork recognition accuracy in a gallery/exhibition environment is presented. Through the controlled introduction of different distractors in a virtual museum environment, it is feasible to assess the effect on the recognition performance of different conditions. The results of the experiment are important for improving the robustness of artwork recognition systems, and at the same time the conclusions of this work can provide specific guidelines to curators, museum professionals and visitors, that will enable the efficient identification of artworks, using images captured with wearable cameras in a museum environment.

Keywords: Paintings Recognition · Computer Vision · Deep Networks.

1 Introduction

The study of museum/gallery visitors has been a rapidly evolving topic within the museum research community which is interested in optimising the overall visitor experience, while analysing visitors' activities, behaviours and experiences. In this paper we describe work related to the development of an in-museum application for tracking the artworks that a visitor is looking at, using first-person images seamlessly captured either by a wearable camera or a smartphone camera. To accomplish the artwork identification task, we utilize deep learning-based object identification algorithms tuned to recognize different artworks in a

museum/gallery environment. The key issue considered in the work is the identification of artworks using object recognition methodologies [12], rather than dealing with the problem of art style identification or visual interpretation problem. Once an artwork is identified, information about the style and interpretation of the artwork can be retrieved from a database that stores such information.

As part of our efforts to produce a system that works with high accuracy in different conditions, we present an investigation aiming to determine the effect of different conditions on the artwork recognition accuracy. To assess the effect of different conditions on the recognition performance, we stage experiments in a virtual environment that allows the controlled introduction of different distractors. The results of the experiment are important for improving the robustness of artwork recognition systems, and at the same time the conclusions of this work can provide specific guidelines to curators, museum professionals and visitors, that will enable the use of this technology in a highly efficient manner. While there are other artwork recognition attempts recorded in the literature [9] [11], to the best of our knowledge, this is the first time that a virtual space is used for simulating different conditions in a controlled way, allowing in that way the derivation of conclusions related to the performance and limitations of artwork recognition using first-person images.

The ultimate aim of our work is to develop a dedicated application capable of identifying the artworks that a visitor is looking at, enabling in that way: a) the provision of additional information about the artwork to the visitor, and b) to register the artworks that the visitor is paying more attention as a means of automating the process of visitor experience evaluation studies [6]. The work described in this paper constitutes our first step in the application development process, where we aim to understand the effect on the identification accuracy due to the introduction of different sources of variation in images captured by a first person camera.

2 Literature Review

The new domain of computer vision focusing on the analysis of images resembling the point of view of a user, collected through wearable cameras or other smart devices, is known as egocentric or first-person vision [2]. Latest developments in image interpretation algorithms, mainly in the form of deep learning, along with the increased image capture abilities and computational power of mobile devices, facilitated the rabid development of novel egocentric applications.

In the case of cultural heritage, wearable camera technologies are often used for artwork interpretation in an attempt to enhance the interaction and experience of visitors of cultural heritage sites. Within this context Taverriti et al. [11] use the YOLO convolutional deep network for identifying eight different artworks within a museum area. Skoryukina et al [10] also consider the problem of recognizing 2D artworks from images captured with mobile devices using a Bag-of-features approach and point geometry. Banerji at al. [1] investigate the use of convolutional neural networks as a means for extracting features from paintings

that can support the tasks of painting, artist, and style recognition. Along these lines they perform experiments using different network architectures and layers.

Instead of focusing on artwork recognition, few researchers considered the problem of identifying the exact location of museum visitors based on images captured with wearable devices. Ragusa et al. [9] use first-person videos captured in the Monastero dei Benedettini, Italy, to identify the location of the visitor. Baseline location recognition performance was presented using a method proposed by Furnari et al. [4] who performs the localization task in three steps that include the steps of frame classification, negative rejection and temporal modelling of the recognized location in previous frames.

3 Paintings Identification Methodology

The process of paintings identification is carried out using a deep network tuned to classify paintings available in an exhibition area such as a museum or a gallery. To train a paintings identification network it is required to have at least one image for each painting. However, to be able to train a classification network, multiple images showing different instances of a painting are needed. To create the desirable training set, data augmentation techniques are adopted [8], that involve the transformation of a given image by changing the image scale, and by rotating the image. As a result of the data augmentation process for each painting we get multiple images showing different instances of a painting.

Recently, many object recognition tasks are performed with high accuracy using deep neural networks [5]. For this reason we have opted to utilize a deep neural network for performing the task of artwork recognition. The convolutional neural network "Squeeze Net" [5] was used due to the limited memory requirements that make it more appropriate for use on mobile devices. A pretrained version of the network trained on more than a million images from the ImageNet database is tuned for classifying paintings by initializing the convolutional layers with the pretrained weights and replacing the output classification layer with the appropriate layer size and fine-tuning the whole network architecture using the training set with paintings. During the classification phase, first-person images captured by visitors are normalized to the standard resolution and input to the tuned "Squeeze Net" allowing in that way the classification of the paintings shown, into the most similar class of paintings in the training set.

4 Experimental Evaluation

As part of our efforts for implementing an application that allows the efficient identification of paintings in a gallery, an experimental evaluation was conducted. The aim of the evaluation was to assess how different factors, frequently encountered in museum environments, affect the identification accuracy. As part of the preliminary investigation we have considered the identification of 10 paintings of renowned Cypriot artists. The main steps of the experimental set up are described below.

4.1 Creating a Virtual Gallery Environment

As a means of simulating accurately different in-museum conditions, we perform experiments in a virtual environment exhibiting the 10 selected paintings placed on a plain wall of a virtual 3D gallery (see figure 1). To simulate the views of a visitor, a virtual camera was placed looking towards the leftmost painting, at a height of 1.6 meters and 1.5 meters away from the wall with the paintings. The initial camera orientation allows the capture of images, as seen by a typical visitor. To simulate the movement of the visitor the camera is gradually moved to the rightmost painting and gradually returns to the starting position. A video showing the views of the virtual camera was recorded and all video frames were annotated to indicate frames where the virtual visitor focuses on a specific painting.

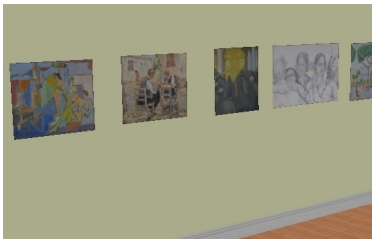

Fig. 1. The setup of the virtual gallery.

4.2 Introducing Distractors

During on-sight visits to museum/galleries exhibiting artworks, possible distractors that may affect the accuracy of automated artwork identification were defined. Identified distractors include changes in the gallery wall texture, changes in ambient light intensity, changes in camera position and orientation with respect to the observed paintings, speed of camera movement and the introduction of occluding structures in the form of additional visitors that inhibit the camera point of view. To assess the effect of the identified distractors, the appearance and overall setting of the virtual gallery and/or the settings of the camera movement were modified in a controlled way so that different distractors are simulated in a controlled way as shown in figure 2.

4.3 Classification

During the process of training the classifier the data augmentation techniques described in section 3 were used for creating, a training set with 100 images (10 images per painting) used for tuning a "Squeeze Net" [5] for classifying paintings. During the performance evaluation stage, all frames in a video were

Fig. 2. Samples showing examples of different conditions simulated in the virtual environment. Row 1 (left) shows the simulation of different wall textures, row 1 (right) the introduction of virtual visitors, row 2 (left) changes in camera orientation and row 2 (right) changes in ambient lighting.

classified using the "Squeeze Net" [5] trained for this purpose, and the correct identification performance was recorded. The performance evaluation indicator was the percentage of correct classification for frames where it is possible to determine the painting that the virtual visitor focuses on. Since at this stage of the experiments, the aim is to assess the effect of different conditions rather than producing the final classification system, for this set of experiments we don't consider the frames where the view is split among two paintings to an extend that it is not possible to indicated the painting on which the visitor focuses on.

4.4 Experimental Results

When the system was tested using the original settings for 99% of all frames considered (about 250 frames) the correct painting was identified. Table 1 shows how the recognition performance is affected by different simulated conditions, while Table 2 summarizes the conclusions regarding the effect of each condition on the classification accuracy.

5 Discussion

A preliminary investigation of the factors affecting the identification accuracy of paintings identification in a museum environment was presented. As part of the investigation, a deep network was trained to classify images of paintings placed in a virtual gallery. The use of a virtual gallery allows the controlled introduction of various conditions that may affect classification accuracy, enabling in that way the extraction of discrete conclusions in relation to the factors that impact the classification performance. To the best of our knowledge the results reported are unique because, unlike the ones reported by other researchers [9], they refer to the recognition of artworks in a virtual environment in the presences of simulated distractors. The results of the investigation provide a useful insight of issues that curators, museum professionals and visitors should consider to maximize the efficiency of this technology. Indicative recommendations include:

- Recommendations for gallery curators and museum professionals
  - Prefer the use of plain colors on the walls rather than textured surfaces
  - Make sure that there is strong ambient lighting in the exhibition area.
  - Limit the number of visitors allowed in a room
- Recommendations for visitors:
  - Keep a steady distance of about 1.5 meters from the paintings
  - Aim to observe a painting while looking straight forward rather than looking at a painting from an angle.
  - Avoid crowded rooms

Table 1. Recognition rates for different simulated conditions

Condition	Parameters	Correct Classification Rate
Wall Texture	No Texture (Baseline)	99%
	Green Stribe	95%
	Blue Stribe	94%
	Gold Fan	88%
Number of Visitors	0 (Baseline)	99%
	5	92%
	10	71%
	15	83%
Distance	150 cm (Baseline)	99%
	200 cm	88%
	250 cm	78%
	300 cm	76%
Camera point of view	Forward (Baseline)	99%
	Upwards	98%
	Downwards	77%
	Left	87%
	Right	86%
Ambient Light Intensity	255 (Baseline)	99%
	200	96%
	150	63%
	100	27%
Camera speed	10 sec	100%
	20 sec (Baseline)	99%
	30 sec	99%
	40 sec	99%

While the initial results obtained prove the feasibility of this approach, there is need for further work to capitalize on the early results. Areas that need further work is the training of deep networks with enhanced training set both in terms of the number of artworks to be recognized and in terms of the variations introduced in the training set. Future work plans also involve the investigation of the effect of additional distractors both in isolation and in combination. In

Table 2. Conclusions regarding the effect of each condition on the artwork classification accuracy.

Condition	**Comments**
Wall Texture	Gallery walls with textures may affect classification accuracy.
Number of Visitors	Increased number of visitors can cause a reduction in classification accuracy, as the point of view of the camera is occluded.
Distance from Paintings	As visitors move away from paintings, classification accuracy is affected. This effect may be attributed to the inability to capture details of a painting from a distance. Furthermore, as the distance between a painting and the camera increased, the proportion of the actual painting included in the point of view is reduced.
Camera Point of View	Changes in the camera orientation affect the classification performance, hence for optimum performance, the camera should be looking forward.
Ambient Light	Reduced ambient lighting has a severe effect on the classification performance.
Camera Speed	The speed of movement does not affect in a significant way the classification performance

parallel with experiments in virtual environments, we are in the process of running experiments in a real environment in the State Gallery of Contemporary Cypriot Art, so that the results of our initial investigation will be utilized for optimizing the performance of artwork identification in a real environment.

The development of robust painting identification technology can form the basis of implementing numerous applications that aim to enhance the experience of the visitor and at the same time provide important information to curators. As far as the visitors are cornered, the identification of a painting will allow the use of mark-less augmented reality systems [3] for obtaining information about exhibits, or the activation of dedicated multimedia applications related to a painting, such as 3d visualizations [7]. At the same time useful information can be derived that includes the time spend in front of each artifact and the level of concentration of the visitor while observing an artifact. We are currently in the process of developing an application that will utilize the proposed framework to introduce the functionalities stated above.

6 Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 739578 complemented by the Government of the Republic of Cyprus through the Directorate

General for European Programmes, Coordination and Development. We also like to thank the personnel of the State Gallery of Contemporary Cypriot Art for their support.

References

1. Banerji, S., Sinha, A.: Painting classification using a pre-trained convolutional neural network. In: International Conference on Computer Vision, Graphics, and Image processing. pp. 168–179. Springer (2016). https://doi.org/10.1007/978-3-319-68124-5_15
2. Bolanos, M., Dimiccoli, M., Radeva, P.: Toward storytelling from visual lifelogging: An overview. IEEE Transactions on Human-Machine Systems **47**(1), 77–90 (2016). https://doi.org/10.1109/THMS.2016.2616296
3. Brancati, N., Caggianese, G., Frucci, M., Gallo, L., Neroni, P.: Experiencing touchless interaction with augmented content on wearable head-mounted displays in cultural heritage applications. Personal and Ubiquitous Computing **21**(2), 203–217 (2017). https://doi.org/10.1007/s00779-016-0987-8
4. Furnari, A., Battiato, S., Farinella, G.M.: Personal-location-based temporal segmentation of egocentric videos for lifelogging applications. Journal of Visual Communication and Image Representation **52**, 1–12 (2018). https://doi.org/10.1016/j.jvcir.2018.01.019
5. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360 (2016)
6. Loizides, F., El Kater, A., Terlikas, C., Lanitis, A., Michael, D.: Presenting cypriot cultural heritage in virtual reality: A user evaluation. In: Euro-Mediterranean Conference. pp. 572–579. Springer (2014). https://doi.org/10.1007/978-3-319-13695-0_57
7. Panayiotou, S., Lanitis, A.: Paintings alive: A virtual reality-based approach for enhancing the user experience of art gallery visitors. In: Euro-Mediterranean Conference. pp. 240–247. Springer (2016). https://doi.org/10.1007/978-3-319-48974-2_27
8. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 (2017)
9. Ragusa, F., Furnari, A., Battiato, S., Signorello, G., Farinella, G.M.: Egocentric visitors localization in cultural sites. Journal on Computing and Cultural Heritage (JOCCH) **12**(2), 1–19 (2019). https://doi.org/10.1145/3276772
10. Skoryukina, N.S., Nikolaev, D.P., Arlazarov, V.V.: 2d art recognition in uncontrolled conditions using one-shot learning. In: Eleventh International Conference on Machine Vision (ICMV 2018). vol. 11041, p. 110412E. International Society for Optics and Photonics (2019). https://doi.org/10.1117/12.2523017
11. Taverriti, G., Lombini, S., Seidenari, L., Bertini, M., Del Bimbo, A.: Real-time wearable computer vision system for improved museum experience. In: Proceedings of the 24th ACM international conference on Multimedia. pp. 703–704 (2016). https://doi.org/10.1145/2964284.2973813
12. Zhao, Z.Q., Zheng, P., Xu, S.t., Wu, X.: Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems **30**(11), 3212–3232 (2019). https://doi.org/10.1109/TNNLS.2018.2876865