Automatic Generation of Coherent Image Galleries in Virtual Reality

Simon Peterhans[0000−0002−6876−7051], Loris Sauter[0000−0001−8046−0362],
Florian Spiess[0000−0002−3396−1516], and Heiko Schuldt[0000−0001−9865−6371]

University of Basel, Switzerland
{firstname.lastname}@unibas.ch

Abstract. With the rapidly increasing size of digitized and born-digital multimedia collections in archives, museums and private collections, manually curating collections becomes a nearly impossible task without disregarding large parts of the collection. In this paper, we propose the use of Self-Organizing Maps (SOMs) to automatically generate coherent image galleries that allow intuitive, user-driven exploration of large multimedia collections in virtual reality (VR). We extend the open-source VR museum VIRTUE to support such exhibitions and apply it on different collections using various image features. A successful pilot test took place at the Basel Historical Museum with more than 300 participants.

Keywords: Virtual Reality · Self-Organizing Maps · Automatic Collection Curation.

1 Introduction

Digital multimedia collections have been growing rapidly over the past decades, both due to the increasing availability and affordability of devices for digital capture, and the growing efforts to digitize existing multimedia collections, especially in the cultural heritage domain. While the size and number of such digital collections is increasing at a rapid pace, methods to organize, manage and explore such collections struggle to keep up with the size and diversity of these data. Especially in the context of museums and archives, manually curating such collections for exhibitions or presentations becomes infeasible very quickly with growing size. Moreover, such digital collections are often far too large to be viewed by a person within reasonable time, much less to be displayed within the physical space available to a museum or archive, leading to only a small subset of the collection being shown and some artifacts never being included at all.

To allow large multimedia collections to be displayed even in small physical spaces, a number of digital-only exhibition solutions have been developed, some, such as the VIRTUE [4] virtual museum, in virtual reality (VR). In this work, we extend the VIRTUE project with the functionality to automatically and interactively generate galleries of coherent images using Self-Organizing Maps (SOMs) [7] for exploration in VR. In May 2022, the VIRTUE system presented here was successfully trialed by the general public in a large-scale deployment at the Basel Historical Museum[1] at the Basel Museum Night[2]. Furthermore, it was featured at Fantasy Basel 2021[3], one of Europe's largest conventions for popular culture. The contribution of this paper is twofold: first, we show how SOMs can be leveraged to automatically create thematic exhibition rooms and second, we report on a practical deployment of the system in a real museum.

2 Related Work

Virtual museums and even museums in VR, such as [5] and the open-source VR museum VIRTUE [4] on which our approach is based, have been developed and investigated for many years. Recent advances in machine learning have allowed for the extraction of content-based semantic feature representations. While there is no existing work on the application of such methods in the realm of virtual museums, such methods have already been used in the context of archives and museums. One such application is described in [1]. This work uses a semantic feature extraction to cluster images and visualize them through a scatterplot and an image path through the feature space.

3 System Architecture

The open-source system VIRTUE[4] [4], consists of a VR-frontend, Virtual Reality Exhibition Presenter (VREP), an admin frontend, Virtual Reality Exhibition Manager User Interface (VREM-UI) and the backend, Virtual Reality Exhibition Manager (VREM). In order to support content-based automatic exhibition generation, we rely on the open-source content-based multimedia retrieval system *vitrivr* [8]. More specifically, we employ its retrieval engine Cineast [3] and the storage layer Cottontail DB [2].

As illustrated in Figure 1, Cineast and Cottontail DB expand the existing backend. VIRTUE's frontend VREP and its backend VREM communicate over a REST API provided by VREM using Javalin[5]. The REST API of VREM, likewise the one of Cineast, provide OpenAPI specifications in order to easily generate corresponding clients. Connections to the storage layers of VREM and Cineast are provided over a dedicated MongoDB Java Driver and gRPC respectively. VREP receives exhibition information from VREM and builds the VR experience on the fly. Previously, VIRTUE only supported static exhibitions, which had to be manually created by users using VREM-UI. In this work, we expand the capabilities of VIRTUE to support dynamically generated exhibitions as well. Storage and management of exhibitions is performed in VREM, as well as the interface to Cineast for dynamically generated exhibitions using the means outlined in the following sections.

[1] https://www.hmb.ch/en/

[2] https://museumsnacht.ch/en/

[3] https://www.fantasybasel.ch/en

[4] https://github.com/VIRTUE-DBIS

[5] https://github.com/tipsy/javalin

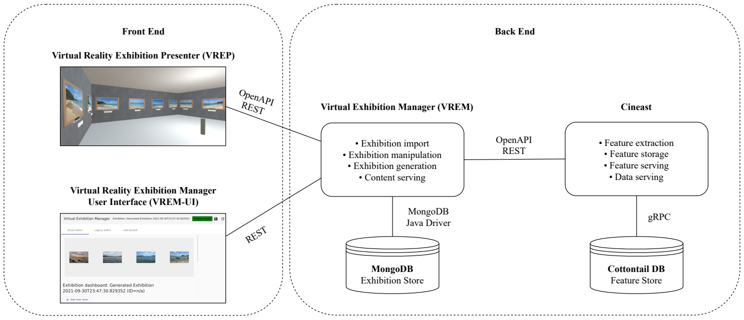

Fig. 1. The architecture of VIRTUE including the extension allowing automatic image gallery generation. Notable interfaces are the two OpenAPI powered REST APIs provided by Cineast and VREM. While the former enables automatic exhibition generation the latter is required in order to display the generated exhibition in VR.

4 Automatic Collection Generation

VIRTUE allows the automatic generation of subsets from an image collection in the form of virtual museum rooms with images being represented by framed exhibits. Much like in an actual museum, the images are not randomly chosen and arranged, but placed in a coherent fashion based on the parameters and features chosen by the user. In order to dynamically generate exhibitions and exhibition rooms, VIRTUE relies on features previously extracted by Cineast and a simple form of the Self-Organizing Map (SOM) algorithm [7].

4.1 Self-Organizing Maps

The SOM, as described by [7], is a type of unsupervised artificial neural network used for dimensionality reduction while maintaining the topological structure of the training data. When trained, it can be used for clustering and data visualization, with samples assigned to nodes in close proximity on the grid being closer in their original vector space than samples assigned to distant nodes. Every node of a SOM grid contains a weight vector of the same dimensionality as the input data, in our case a feature vector of an image. We denote these weights for the i-th node n_i in the grid with $\boldsymbol{m}_i$. The grid is presented with one sample $\boldsymbol{x}$ at each distinct time step t during the training phase and nodes compete for each sample. The winning node holds the weight vector with the smallest Euclidean distance to the sample. Its index c in the grid can thus be determined as follows:

$$c = \underset{i}{\operatorname{argmin}} \|\boldsymbol{x}(t) - \boldsymbol{m}_i(t)\| \tag{1}$$

Once the winning node has been declared, the weights of all nodes are updated prior to the next time step $t+1$ by adjusting a node's weight by a fraction of its difference to the sample. The respective fraction is determined by a neighborhood function h_{ci}, which assigns a scalar to node n_i based on its distance to the winning node n_c on the grid. The further away a node n_i is from the winning node n_c, the smaller the value returned by h_{ci} is. Nodes with larger distances that are not in the neighborhood of the winning node are thus affected less by the current sample and may not have their weight updated at all. Using the neighborhood function, the weights of each node can the be updated prior to the next iteration $t+1$:

$$\boldsymbol{m}_i(t+1) = \boldsymbol{m}_i(t) + h_{ci}(t)\left[\boldsymbol{x}(t) - \boldsymbol{m}_i(t)\right] \tag{2}$$

The number of iterations required for convergence depends on the feature vectors, the grid size, and the neighborhood function at hand. When trained, the map can be used for classification and clustering by letting the nodes compete and assigning the sample to the winning node, similar to the process in the training phase. The grid topology of the SOM is chosen based on the desired dimensionality reduction of the input data.

4.2 From SOM to Exhibition

VIRTUE currently generates quadratic and rectangular rooms, consisting of four walls, a floor, and a ceiling. The pictures featured in each room are, by default, arranged in a single row on each wall. To obtain a clustering that is coherent within this room architecture, we arrive at an evenly spaced, circular, one dimensional grid projected onto a cylinder, which defines the topology of the SOM to be trained to cluster and arrange the image collection. To respect these constraints, a two-dimensional toroidal grid that wraps on the vertical axis, essentially representing a bottom- and top-less cylinder, is used by choosing an appropriate neighborhood function. When trained, all images in the collection are clustered on the grid, such that each image is assigned to the closest grid node. For each node the closest image is chosen as representative image for the respective cluster and is displayed in the generated exhibition room. Ultimately, in doing so, the coherent representation of the collection is established, as close images are topically related while images across the exhibition room are only related in so far, as they belong to the same collection.

To illustrate this approach, Figure 2 shows a SOM trained on a simple color feature from a collection of beach images[6] generated using the method presented in [6]. This 1×16 grid was trained with a wrapping neighborhood function and shows the layout of the 16 nodes and their representative images on the walls in a single exhibition room. As all images that were used in the generation process have been assigned to one of the nodes and are thus represented by a node's representative image, additional SOMs can be recursively generated for each node. This results in users being able to select one of the images in the room and generate a sub-room with images that were also assigned to this node to further explore the collection.

[6] https://thisbeachdoesnotexist.com/

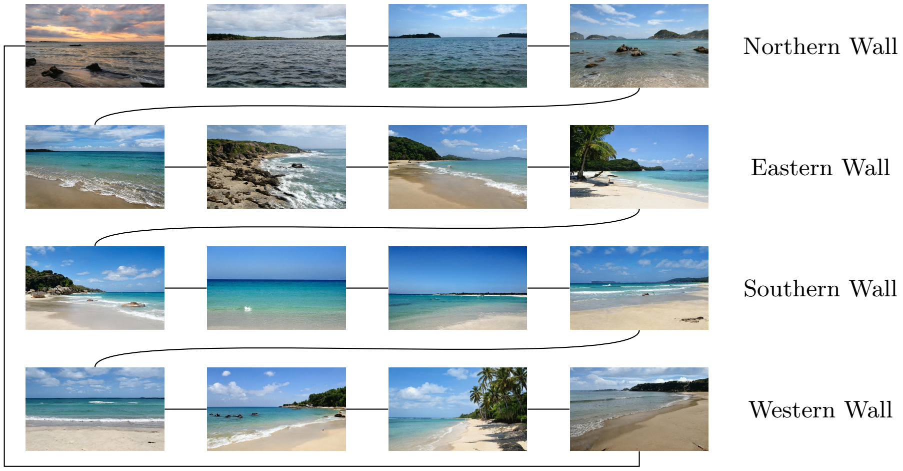

Fig. 2. An example of a 1 × 16 SOM grid trained on color features of beach images and mapped to the exhibition room with four images per wall. Images close to each other exhibit similar color regions and a smooth transition around the room occurs, e.g., with the beach and forest being on the left side on the eastern wall and slowly shifting over to the right side on the Southern and Western wall.

4.3 Features

We have tested our method on a number of different combinations of image collection and feature vector. The features we have found to work best for the purpose of automatic coherent collection generation are an average color grid feature and a deep learning based semantic image content feature.

The average color grid feature consists of a vector containing the average color values sampled from an image in an 8×8 grid. Despite being rather simple, it results in visually coherent exhibitions (cf. Figure 2), especially for collections of images containing diverse and saturated colors. This feature is not suitable to generate coherent exhibitions for all collections, in particular collections of cultural heritage collections (e.g., the ones from the Swiss Society for Folk Studies[7] used in our evaluations), where images only have muted color or no color at all.

As a deep learning based semantic image feature, we use the visual-text co-embedding originally developed for multimedia retrieval, which is described in more detail in [9]. Trained to embed images and associated text descriptions into a common vector space such that semantically similar inputs are close together, this feature extracts a semantic representation from the content of an image. Using this feature, the generated exhibitions exhibit a coherence based on semantic image content, which lends itself well also to gray scale image collections and those with muted colors. As seen in Figure 3, this feature allows for a clustering more focused on the semantic content of the images and, as a result, allows users to explore the collection based on thematic rather than purely visual similarity.

[7] https://archiv.sgv-sstp.ch

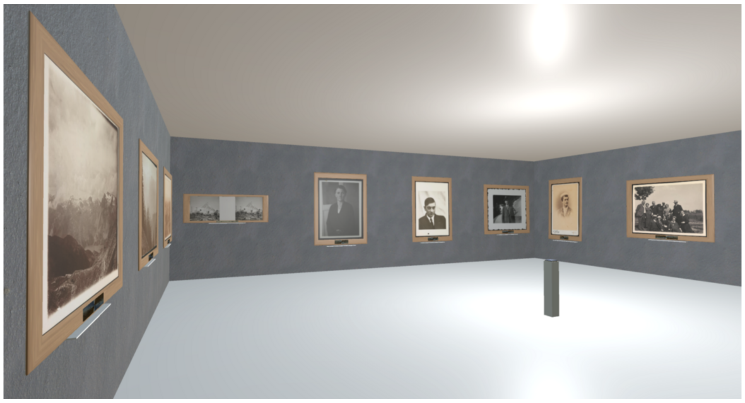

Fig. 3. A generated room for the SGV_10 collection of digitized cultural heritage images provided by the Swiss Society for Folk Studies. Portraits, pictures of multiple people and images featuring landscapes are coherently grouped together.

5 Discussion and Conclusion

In this paper, we proposed the use of SOMs to automatically generate coherent image galleries in VR, that allow for hierarchical exploration of an image collection, and extended the open-source VR museum VIRTUE to support dynamically generated coherent image galleries. Using a simple color feature and a semantic image feature based on deep learning, the approach has been applied to two converse image collections featuring thousands of images. With the topology-preserving nature of the SOM, visually and semantically coherent exhibition rooms were generated in both cases.

The implementation was tested with two image collections, one consisting of 10'000 computer generated images and the other consisting of 6'461 digitized cultural heritage photographs. Our experiments show that while exhibitions can be pre-generated using this method, our implementation is powerful enough to be used for real-time generation, even for large collections. At the Basel Museum Night 2022, we successfully presented the virtual museum to the general public in the premises of a real museum, the Basel Historical Museum, with digitized artworks from this museum. During the event, more than 300 participants had the chance to explore the SOM-generated thematic exhibition rooms.

While our findings are very promising, further research is necessary to develop better methods to generate dynamic and interactive exhibitions from large image corpora and to make them appropriately accessible in VR. Further work is needed especially for personalizing generated exhibitions based on user preferences.

Acknowledgements This work was partly supported by the Swiss National Science Foundation (project “Participatory Knowledge Practices in Analogue and Digital Image Archives”, contract no. CRSII5_193788).

References

1. Bönisch, D.: The curator's machine: Clustering of museum collection data through annotation of hidden connection patterns between artworks. International Journal for Digital Art History **5**, 5–20 (2020)
2. Gasser, R., Rossetto, L., Heller, S., Schuldt, H.: Cottontail DB: an open source database system for multimedia retrieval and analysis. In: Chen, C.W., Cucchiara, R., Hua, X., Qi, G., Ricci, E., Zhang, Z., Zimmermann, R. (eds.) MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020. pp. 4465–4468. ACM (2020). https://doi.org/10.1145/3394171.3414538, https://doi.org/10.1145/3394171.3414538
3. Gasser, R., Rossetto, L., Schuldt, H.: Multimodal multimedia retrieval with vitrivr. In: El-Saddik, A., Bimbo, A.D., Zhang, Z., Hauptmann, A.G., Candan, K.S., Bertini, M., Xie, L., Wei, X. (eds.) Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR 2019, Ottawa, ON, Canada, June 10-13, 2019. pp. 391–394. ACM (2019). https://doi.org/10.1145/3323873.3326921, https://doi.org/10.1145/3323873.3326921
4. Giangreco, I., Sauter, L., Parian, M.A., Gasser, R., Heller, S., Rossetto, L., Schuldt, H.: Virtue: a virtual reality museum experience. In: Proceedings of the 24th international conference on intelligent user interfaces: companion. pp. 119–120 (2019)
5. Hirose, M.: Virtual reality technology and museum exhibit. In: Subsol, G. (ed.) Virtual Storytelling, Using Virtual Reality Technologies for Storytelling, Third International Conference, ICVS 2005, Strasbourg, France, November 30 - December 2, 2005, Proceedings. Lecture Notes in Computer Science, vol. 3805, pp. 3–11. Springer (2005). https://doi.org/10.1007/11590361_1, https://doi.org/10.1007/11590361_1
6. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 12104–12114. Curran Associates, Inc. (2020)
7. Kohonen, T.: Automatic formation of topological maps of patterns in a self-organizing system. In: Proceedings of 2SCIA, Scand. Conference on Image Analysis. pp. 214–220 (1981)
8. Rossetto, L., Giangreco, I., Tanase, C., Schuldt, H.: vitrivr: A flexible retrieval stack supporting multiple query modes for searching in multimedia collections. In: Hanjalic, A., Snoek, C., Worring, M., Bulterman, D.C.A., Huet, B., Kelliher, A., Kompatsiaris, Y., Li, J. (eds.) Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, October 15-19, 2016. pp. 1183–1186. ACM (2016). https://doi.org/10.1145/2964284.2973797, https://doi.org/10.1145/2964284.2973797
9. Spiess, F., Gasser, R., Heller, S., Parian-Scherb, M., Rossetto, L., Sauter, L., Schuldt, H.: Multi-modal video retrieval in virtual reality with vitrivr-vr. In: Jónsson, B.Þ., Gurrin, C., Tran, M., Dang-Nguyen, D., Hu, A.M., Binh, H.T.T., Huet, B. (eds.) MultiMedia Modeling - 28th International Conference, MMM 2022, Phu Quoc, Vietnam, June 6-10, 2022, Proceedings, Part II. Lecture Notes in Computer Science, vol. 13142, pp. 499–504. Springer (2022). https://doi.org/10.1007/978-3-030-98355-0_45, https://doi.org/10.1007/978-3-030-98355-0_45