Designing the Virtual Rosetta: A Tool for Exploring Historical Drawings in VR

Anessa Petteruti
Brown University

Cindy Nguyen
University of California, San Diego

David H. Laidlaw
Brown University

ABSTRACT

We present the results of a comparative analysis of various layout and design displays for historical drawings. This involved developing a new visualization environment, known as the "Virtual Rosetta" for a collection of early twentieth century Vietnamese drawings known as Technique du Peuple Annamite (Mechanics and Crafts of the People of Annam) compiled by French colonial administrator Henri J. Oger (1872-1929). In this paper, we discuss similar work that has been pursued by researchers in digital humanities, specifically working in virtual displays of artwork, the design of the Virtual Rosetta, including text readability, wall and text contrast, and organization of drawings, as well as the associated results. We used an analysis technique known as hierarchical clustering utilizing sentence embeddings, a technique in natural language processing in which sentences are mapped to numerical vectors, to organize a subset of the drawings so that drawings with similar text descriptions are located near each other. Performing this on an entire dataset would allow humanities researchers to most effectively and efficiently visualize their findings. We report on user experience and efficacy for a number of virtual environment layouts for displaying drawings and visual information for research analysis in virtual reality.

1 INTRODUCTION

Virtual reality provides more space to explore and analyze larger datasets that other visualizations would not ordinarily be able to. From various viewing systems to interactive components coupled with machine learning elements such as image and text detection, the medium has the power to visualize data that is ordinarily cumbersome to display as well as complement learning and education. This paper's primary goal is to demonstrate the application of virtual reality to visual art in digital humanities research, specifically the display of historic Vietnamese drawings with accompanying Vietnamese, French, and English translations. We have named our software the "Virtual Rosetta." The images are displayed in a room setting, as shown in Figure 1, with organizational techniques to facilitate easier reading of the collection of 4,200 drawings. Current humanities research by Nguyen examines indigenous knowledge production and aesthetic style within the text. We investigated how virtual reality enhances historical research by visualizing historical drawings and translation text in Unity and the Oculus Quest 2. Displaying a series of visual work in a compelling, organized way is challenging. We settled on creating a virtual room, similar to a museum supporting multiple languages (hence the "Rosetta") that includes a series of panels displaying annotated drawings that the viewer can walk through to examine work of the same theme. The drawings are categorized by the similarity of their English translation captions and arranged within a category according to their size and shape. These categories were created based on each drawing's similarity values defined by BERT (Bidirectional Encoder Representations from Transformers) embeddings stored in a similarity matrix [2]. We also used other approaches such as grouping drawings by pages and as well as a manual tagging of each drawing to a relevant category.

Figure 1: A broad view of the Virtual Rosetta displayed in Unity.

In this paper, we discuss similar work that has been pursued by researchers in digital humanities, specifically working in virtual displays of artwork, the design of the Virtual Rosetta, including text readability, wall and text contrast, and organization of drawings, as well as the associated results.

We built on Unity for several reasons: it is a convenient tool with a responsive UI and easy debugging and works with C#, C++, Python, and Java in the form of libraries, to leave the door open for creating scripts in these supported languages in the future. It is also the most widely used virtual reality development platform with over 91% of Hololens experiences made through the software [3].

In 1908 and 1909, Oger published Technique du peuple Annamite, a collection of 4,200 drawings (a single page of the original drawings is shown in Figure 2) with annotated captions written in Vietnamese demotic script. The work also contains an introductory essay titled Introduction Générale à L'étude de la Technique du peuple Annamite: Essai sur la vie materielle, les arts et industries du peuple Annamite (General Introduction to the study Mechanics and Crafts of the Vietnamese: Essay on the Material life, the arts and industries of the people of Vietnam). Each drawing includes its own descriptive story. Some drawings feature people and objects while others solely present objects. Vietnamese social life and cultural practices such as cooking, housework, and farming are common themes in the collection.

Specifically, we sought to present the drawings, annotated captions, and translations together in an immersive virtual reality setting.

2 RELATED WORK

Incorporating mixed reality into art is not a new phenomenon. Virtual reality has propelled different facets of art into new dimensions. In 1998, Marc Levoy, Computer Science Professor and Researcher at Stanford University, lead a team to scan Michelangelo's David in Florence, Italy. The dataset has been used by several papers presented in SIGGRAPH, but it was never viewable in real time until 2017 when Chris Evans and Adam Skutt of Epic Games modeled

978-1-6654-1370-1/21/$31.00 ©2021 IEEE
DOI 10.1109/VIS4DH53644.2021.00006

Figure 2: A single page of the original historic Vietnamese drawings.

the interior of the Tribune in the Galleria della Accademia in Florence where David is housed and created an application viewable from a headset in which individuals can ride a scaffold up and down while walking around the statue [7]. Similar to Evans and Skutt, we seek to provide a virtual reality visualization of historic Vietnamese drawings.

Text display is a significant aspect of the Virtual Rosetta. Several text locations and type were have been investigated [9]. Researchers found that Rapid Serial Visual Presentation (RSVP) text is a promising presentation type for reading a short portion of text. We evaluated several text sizes in order to settle on a display of caption text that the reader can easily and comfortably read.

There are existing examples of physical displays of drawings. We modeled the Virtual Rosetta on physical displays of drawings like those at the Kremer Museum launched in 2017. Designed by Johan van Lierop, Founder of Architales and Principal at Studio Libeskind, the museum includes 74 Dutch and Flemish Old Master paintings accessible exclusively through virtual reality technology [4]. Projects such as VIVE Arts aim at bringing museum experiences to people who would ordinarily not be able to participate. The primary mission of the VIVE Arts' series of projects is to preserve cultural heritage for the world and democratize creation with digital innovation in the arts [1]. This implementation is an example of a technological exhibition of physical artwork. We also present a virtual exhibition for drawings - however, the Virtual Rosetta includes direct translations from Oger's Technique du peuple Annamite as well as descriptions of each drawing, making it both a "Rosetta" and museum, rather than solely a museum. Oger's encyclopedia includes visual sketches of Vietnamese crafts and practices as well as annotations in both French and Vietnamese (in ch nôm, a logographic Chinese writing system of Vietnamese language).

3 Design

We presented several design iterations to a half dozen colleagues and informally solicited feedback. We discuss our design decisions, including several room layouts that we evaluated, drawing and text sizes as well as layouts, and organizational techniques. We hypothesize that a museum-like layout of L-shaped walls within a rectangular room, displaying panels of drawings on each wall is a spatially-economical and legible reading and viewing environment for drawings.

There are several metrics we experimented with in order to determine the best layout and orientation of the drawings and their accompanying labels. The benchmark for our experimentation primarily involved the first iteration of the Virtual Rosetta as well as

Figure 3: A visual display of the result of a sample of clustering.

prior research performed by various research teams. Questions involved: what are the best and worst text parameters users generally choose for reading in virtual reality? Are selected text parameters significantly different depending on text length? What is the influence of text length and device type on user experience while reading in virtual reality? In order to answer these questions, we evaluated different designs and layouts of the Virtual Rosetta and recorded the results.

3.1 Room Layout and Organization

Initially, we had to determine a good room layout to display all 4,200 drawings. We considered three designs for the Rosetta: a large room with no inner walls, a large room with four rows of inner walls, and a large room with two inner L-shaped walls. In the L-shaped wall layout approach, users are able to navigate in and out of the center walls in order to view each drawing or group of drawings. This keeps all the drawings in one space (i.e. a single room) while keeping them organized. This also decreases the likelihood a user will feel dizzy or uncomfortable while navigating the Virtual Rosetta since all the drawings and information are in one place without being in several separate areas where the user would have to be transported. If users are required to visit multiple rooms in VR in order to view all the artwork, they could become confused regarding the order of the rooms and organization of each room. Therefore, displaying the art in one room would reduce this possible confusion and dizziness since all the artwork is displayed in a single area.

3.2 Individual Wall Layout

We planned the overall design of the Rosetta's walls to be clean and minimalist to result in a high color contrast and legible text. Therefore, the background of the entire Rosetta is a light gray stucco texture to resemble stone. The stucco texture also helps with the stereo vision in VR. As far as organization for the manual categorical approach, we decided to include text signs displaying the category names ('Cooking,' 'Clothing,' 'Maintenance,' 'Items,' and 'Entertainment') on the walls in order to help the user locate drawings under categories they would like to view. This way, users will always have an idea of where they are located in the library.

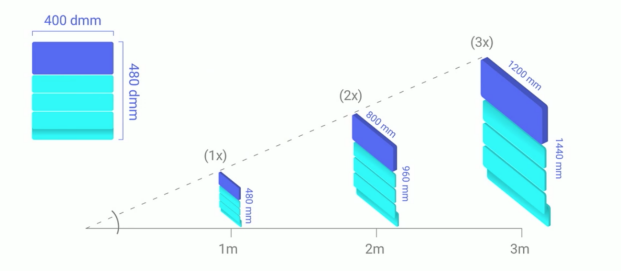

Figure 4: Various DMM values. [5]

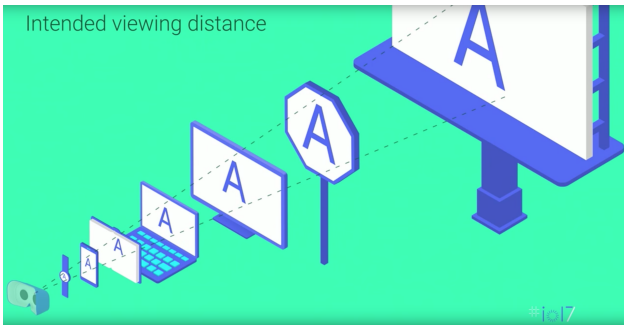

Figure 5: Viewing distances from a headset perspective. [5]

3.3 Drawing Size and Framing

In order to determine the canvas size and general layout of the Rosetta, we took into account various spacing between text and images to settle on a design so the user could navigate from wall to wall and study each drawing, reading each caption with ease. The full width of the equirectangular projection represents 360 degrees horizontally and 180 degrees vertically.

We presented participants with two different spacing layouts of the drawings. The drawings in one version were directly adjacent to each other. We subsequently asked each participant which version they preferred and recorded the results to settle on a final layout design. This design presents the drawings 2 inches away from each other. A panel of the clustered drawings (five clusters, specifically) is shown in Figure 3.

3.4 Text Size and Format

In an experiment conducted by the Quality and Usability Lab of TU Berlin, Germany and the German Research Center for Artificial Intelligence, the mean values for angular size (the combination of distance and font size) and color contrast depending on varying text length were reported [8]. Considering the results of their experimentation, we additionally concluded that a contrast ratio of at least 4.5:1 for standard text and 3:1 for large (heading) text is ideal with an initial distance from view of 2 to 3 feet. Similarly, we also had to consider text color, which goes hand in hand with the contrast. To produce the most contrast, we concluded there should be a gradient of color text in order to also distinguish between the three separate languages, Vietnamese, French, and English. The Vietnamese translation is in black, the French in a charcoal, and the English in black italics. If a drawing included character text (defined as supplementary captions written in Vietnamese), we added this in bold red text underneath the other Vietnamese, French, and English translations. We found that this color scheme keeps the design of the Rosetta consistent and simple while simultaneously making it clear to the user that three different languages are displayed and correspond to each drawing. In addition, this black/grey color scheme contributes to a higher contrast ratio against the stucco background.

Google proposed the distance-independent millimeter (dmm) in 2017 since there was no standardized unit incorporating text size and viewing distance within the virtual reality sphere [5]. Figures 4 and 5 show various DMM values from a VR headset perspective. DMM is described as one millimeter at a meter away. DMMs allow for initial design and then scaling later. Therefore, the unit was quickly adopted and is used in industry and research. We used it to determine the optimal text size and viewing distance for the Vietnamese drawing captions displayed in three languages.

In Unity, the canvas dimensions are in meters so DMMs have to be scaled down by 1000%. Depending on its intended viewing distance, the text can be scaled so that it is legible. We displayed the drawing captions in different angular sizes and determined the angular size at which the text is legible. Based on the results of our experiment, we found the optimal angular size of drawing captions to be 38dmm +/- 16. This angular size is optimal because 38dmm +/- 16 is legible when read at several distances and heights, 4 feet above the ground, 6 feet above the ground, and 4 feet away from the wall. We calculated the amount of time the user took to read a single caption at different text and angular sizes from start to finish. We gathered feedback that the 35pt text size was the most legible out of all the text sizes we presented to our users. This is approximately 17% larger than the 30pt text size of the captions in the first Virtual Rosetta iteration.

After showing users the drawing captions in three different text fonts and sizes, Arial at 30pt size, Palatino at 30pt size, and Helvetica at 30pt size, we settled on Arial since it resulted in the lowest reading time for a single drawing caption of the Rosetta. In addition, Arial is a sans serif font and provided the most contrast against the Rosetta's white background. While serif fonts might be comfortably legible on desktop, we discovered that sans serif fonts were ideal for virtual reality. Excessive strokes and lines present in serif fonts would distract from the viewing experience.

Text line length was also something we had to consider - 50-75 characters per line is preferable length, which we found to be more ideal than the shorter line length of 30-40 characters per line that we found in several research papers. Approximately 60 characters per line of text in the drawing descriptions allow enough text to be seen while not overwhelming the user with too much information in one view.

For both desktop views and virtual reality, it is convenient to keep line spacing between 1.2 and 1.5. Since in VR especially it is more complicated to keep focus, it is beneficial to avoid minimizing the line spacing to less than 1. Regarding alignment and text placement, researchers at the Center for Interaction, Visualization and Usability (CIVU) in Germany found that the readability of text in VR can be improved by constantly magnifying a part of the screen, augmenting the scene by adding floating texts, and magnifying scene elements such as signs depending on the gaze direction [6]. However, we found that floating texts were not suitable for a Rosetta/museum/library type interface since they distract from the intended central focus of the scene: the images (in this case, the drawings). Therefore, we sought to improve the appearance of imagery in virtual reality by keeping the text left aligned either beside or underneath the drawings as secondary information that is not the user's primary focus.

The reading speed for each drawing caption as well as the category headings was evaluated. Users can read through text of the Virtual Rosetta at their own pace. However, we wanted to settle on a font type and size that users would not have trouble reading. Specifically, the reading speed of several Radner sentence optotypes [6] revealed certain text sizes are more legible in a virtual museum or

Rosetta setting than others. The mean reading speed obtained with 10 selected short sentences in 30pt text size in the three displayed languages, Vietnamese, French, and English, was 190.36 ± 33.76 words per minute (wpm) as compared to 195.64 ± 34.83 words per minute (wpm) for longer paragraphs of text (i.e. longer drawing captions). The mean reading speed obtained with 10 selected short sentences in 35pt text size in the three displayed languages, Vietnamese, French, and English, was 190.36 ± 33.76 words per minute (wpm) as compared to 195.64 ± 34.83 words per minute (wpm) for longer paragraphs of text (i.e. longer drawing captions). We determined that caption sizes of text could be comfortably read at 35pt based on these reading speeds.

3.5 Organizational Analysis Approach: Hierarchical Clustering

One of the major points we had to consider was the organization of the individual drawings in a virtual reality setting. We considered several organization methods, such as organization by pages (the original groups in the dataset that the drawings were extracted from), before settling on hierarchical clustering to organize the drawings. In the first iteration of the wall design, each individual wall of the Rosetta included an entire page of 14 drawings. However, we decided to supplement this by giving the user the option of organizing the individual drawings by their corresponding tags, i.e. descriptions or labels. We also considered manually categorizing the drawings based on five chosen categories: 'Cooking,' 'Clothing,' 'Maintenance,' 'Items,' and 'Entertainment.' We settled on these categories after assessing the visual similarities among all the drawings. We decided to implement a programmatic method with hierarchical clustering on a subset of the original dataset of drawings. SciKit Learn's Agglomerative Clustering uses a bottom-up approach in which each observation starts in its cluster. Clusters were then gradually merged together.

If descriptions of each drawing are provided in a visual dataset a researcher would like to analyze, textual hierarchical clustering can be performed to analyze and organize the data. In order to replicate the Rosetta's design and functionality, a developer or digital humanities researcher can find BERT embeddings and create a similarity matrix based on each drawing's description's ('tag') contextual closeness to another drawing's description in the dataset. Why are BERT embeddings useful? In this case, we wanted to take the context of a word in a phrase into account, and BERT has an advantage over other models, such as Word2Vec, because it outputs a representation for a word that is informed by the words around it (i.e. the word's context). We used NLTK's corpus to remove stop words from the drawing tags as well as the built-in tokenizer to splice phrases into smaller units, typically individual words. Sklearn's Metrics and Feature Extraction packages were useful in determining the cosine similarities of phrases our dataset as well as converting strings to numerical vectors through the process of vectorization. Python's SentenceTransformers framework provided us with text embedding accessibility. We visualized the produced clusterings in dendrograms. Individual drawings were grouped and organized according to their similarity values (drawings that were more similar to each other were grouped together).

We evaluated two similarity metrics for the text tags, Jaccard similarity and cosine similarity. In Jaccard similarity, lemmatization is first performed to reduce variations of the same root word to their corresponding root word. The ratio of the size of the symmetric difference to the union is then calculated to arrive at the Jaccard index. In cosine similarity, the similarity is measured by the cosine of the angle between the two vectors. Sentences are converted into vectors using the bag of words approach, the term frequencies are normalized with their corresponding magnitudes, and the dot product of these normalizations is taken. Because we did not want to take solely unique sets of words for each sentence (as the Jaccard similarity does) in our dataset, we ultimately decided to utilize the cosine similarity which takes into account the total length of the vectors.

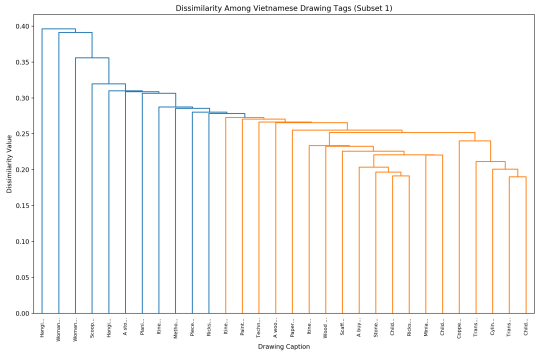

Figure 6: Clustering diagram displaying the dissimilarity of drawing tags in Vietnamese in which lower values represent more similar tags and higher values represent more dissimilar tags. This diagram displays the clustering for a subset consisting of 20 drawings of the larger dataset of 700 pages of Vietnamese drawings. The complete phrases for the above drawing tags from left to right are: 'Hanging bar for sling to carry paper.', 'Woman gluing ritual papers that represent sheets of gold or silver.', 'Woman preparing leaf to be made into a hat.', 'Scoop and scraper to work dough for keo,' 'Hanging hook,' 'A stone dog to guard a village entrance,' 'Planing a cylindrical pole,' 'Itinerant charcoal seller,' 'Method for cutting a small piece of wood,' 'Piece from old time gun,' 'Rickshaw coolie carrying lantern,' 'Itinerant sales woman sitting on her balancing pole,' 'Painted wooden sign representing a squash,' 'Technique to shorten a balancing pole,' 'A wood sculptor's square and chisels,' 'Paper fan,' 'Itinerant fire wood seller,' 'Wood or iron measures,' 'Scaffolding of house under construction,' 'A buyer of remnants from rice husking,' 'Stone mortar,' 'Child's bib,' 'Rickshaw coolie relaxing,' 'Miniature landscape in a basin,' 'Child's pants,' 'Copper object of veneration with Sanskrit phrases,' 'Transportation on bearer's head,' 'Cylindrical-headed dais carried in processions,' 'Transportation of skins,' 'Child's smock'. There are more useful clusters displayed in the orange than in the blue with low dissimilarity values around 0.35.

The clustered tags of the drawings can be organized in a dendrogram as shown in Figure 6 in which the nodes of the dendrogram represent the individual words clustered by and the drawings containing that word as a tag. This is useful since the user is able to recognize common themes among thousands of drawings. In fact, such a format is useful for several cultural or humanities applications in which a series of works containing common themes is displayed. This hierarchical clustering can be used for organizing a display of the drawings in which the tags representing each cluster is shown alongside the drawings. Users would be able to visualize drawings in their corresponding clusters in order to view the similarities of drawings on a larger scale.

4 Discussion

The Virtual Rosetta provides a display of drawings, which does not have to be limited to historic Vietnamese drawings. For example, work of ancient Rome, Greece, or the Middle East can be visualized in similar ways. Hierarchical clustering is application-agnostic and provides a method of organizing a photo or video dataset. The Rosetta provides humanities researchers with a comprehensive platform to visualize images in depth. However, the platform does have its limitations - certain features and functionalities could be implemented in order to make the Rosetta a more comprehensive medium

for viewing and analyzing imagery. Overall, certain design decisions were made in order to provide a satisfying, informative viewing experience for users, primarily researchers. As far as visualization, the format of the Rosetta's walls, text size, incorporating angular size, drawing size, and overall room size were all essential metrics. Regarding analysis techniques, hierarchical clustering proved useful in determining how effectively to organize the individual drawings on a single page of drawings.

Some new questions arise as a result of our research: are there other analytical methods, in addition to hierarchical clustering, that could be used in order to gather more information on our dataset and how to best represent the data, especially the visual data? Would the Virtual Rosetta be a useful exhibit and popular attraction at a museum? Could the Virtual Rosetta present a larger dataset? Or can the Rosetta only display a certain amount of data? How much data is too much?

The first design of the Rosetta, a large room with no inner walls in which each room wall displayed drawings organized by page, allowed the user to view all drawings. However, it was difficult to navigate since the drawings were too small to view and oftentimes drawings on the same page had no relation to one another. This inspired us to iterate on this design, producing the second version of the Rosetta in which the large room included four rows of inner walls. This presented more wall area for the drawings to be displayed - however, the straight rows closed off the room too much so that the user could not navigate and view most of the drawings at once. We designed the third version of the Rosetta in which the four rows of inner walls turned into two L-shaped walls to preserve the open space of the room while still presenting more wall surface area for the drawings.

We performed the hierarchical clustering on subsets of the dataset of approximately 4,200 drawings, and useful clusters were produced, potentially indicating that performing the clustering on the entire dataset could yield even more useful results (since there could be more similarities found in a larger dataset of drawing descriptions than a smaller dataset of drawing descriptions.

5 Conclusion

We found support for our initial hypothesis that a museum-like layout of L-shaped walls within a rectangular room, displaying panels of drawings on each wall is a spacially-economical and legible reading and viewing environment for drawings based on the collected results. We believe the Virtual Rosetta will allow for finding important insights in Oger's dataset of Vietnamese drawings. The Rosetta can also be applied to other datasets with visual examples. We also found support for a format to organize the drawings by their corresponding tags with hierarchical clustering since this is both an intuitive and visual approach to organizing thousands of individual drawings.

Our results show that specific display layouts and features for Rosetta and museum environments can be effective for immersive VR visualizations. In particular, light-colored, textured walls and dark, crisp text for high contrast improves text readability. Overall, we provide drawing and text design recommendations for Rosetta and museum settings.

References

[1] Fuhrman. Our Projects. https://arts.vive.com/us/our-projects/, 2016.

[2] D. J. Bert: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT 2019*, 1(1):4171–4186, Oct. 2018.

[3] Jagneaux. HoloLens 2 Development Edition Includes Three Free Months Of Unity Pro. https://uploadvr.com/hololens-2-development-edition-unity-pro/, 2019.

[4] Kremer. The Kremer Museum - The Kremer Collection. https://www.thekremercollection.com/the-kremer-museum/, 2017.

[5] D. L. Vr reading uis: Assessing text parameters for reading in vr. *Extended Abstracts of the 2018 CHI Conference*, 18(3):13–15, Apr. 2018.

[6] K. L. Improving readability of text in realistic virtual reality scenarios: Visual magnification without restricting user interactions. *MuC'19: Proceedings of Mensch und Computer 2019*, 18(3):749–753, Aug. 2019.

[7] L. M. The digital michelangelo project. *Computer Graphics Forum*, 18(3):xiii–xvi, Aug. 1987.

[8] G. R. User experience of reading in virtual reality – finding values for text distance, size and contrast. *2020 Twelfth International Conference on Quality of Multimedia Experience*, 18(3):6–9, Aug. 2020.

[9] R. R. Reading in vr: The effect of text presentation type and location. *CHI Conference on Human Factors in Computing Systems*, 1(531):1–10, May 2021.