Texture-based clustering of archaeological textile images

Davit Gigilashvili; Dept. of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Ha Thu Nguyen; Dept. of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Casper Fabian Gulbrandsen; Dept. of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
Margrethe Havgar; Museum of Cultural History, University of Oslo; Oslo, Norway
Marianne Vedeler; Museum of Cultural History, University of Oslo; Oslo, Norway
Jon Yngve Hardeberg; Dept. of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway

Abstract

Archaeological textiles are often highly fragmented, and solving a puzzle is needed to recover the original composition and respective motifs. The lack of ground truth and unknown number of the original artworks that the fragments come from complicate this process. We clustered the RGB images of the Viking Age Oseberg Tapestry based on their texture features. Classical texture descriptors as well as modern deep learning were used to construct a texture feature vector that was subsequently fed to the clustering algorithm. We anticipated that the clustering outcome would give indications to the number of original artworks. While the two clusters of different textures emerged, this finding needs to be taken with care due to a broad range of limitations and lessons learned.

Motivation

Historical tapestries tell interesting stories and provide invaluable insight into the era they belong to. However, archaeological textiles are oftentimes highly fragmented that makes reading their motifs challenging if not impossible. A vivid illustration of the latter case is the Viking Age tapestry collection from the Oseberg burial, Norway [1]. Therefore, solving the puzzle is needed to re-assemble the fragments and recover the original motifs. The recent advances in computer science may serve this goal in two ways: first of all, machine learning may assist human experts in the process of puzzle solving and provide valuable suggestions; secondly, experts can freely interact with the images of the fragments, while this is impossible with their physical counterparts due to their highly fragile nature. What complicates things further is oftentimes the lack of information on how many original pieces of artworks need to be recovered, and whether a given pair of fragments belong to the same piece at all. To facilitate the puzzle solving process, clustering techniques may divide fragments into meaningful groups that separate pieces that belonged to the different original artworks and put the fragments from the same original together.

Problem

In this work we attempt to answer two research questions:

1. How many original artworks do the surviving Oseberg tapestry fragments come from?
2. Which fragments belong together for puzzle solving purposes?

While a broad range of machine learning literature has addressed puzzle solving problem in the artwork imagery [2-4], as well as for 3D archaeological artefacts [5-7], to the best of our knowledge none of them proposed a methodology that is robust enough for solving the puzzle of highly degraded, irregularly shaped artefacts that come from unknown number of originals. We believe that identification of the number of originals and grouping the fragments potentially from the same piece of artwork is the first and vital step toward the puzzle solving goal. From the computational perspective, the research problem can be formulated as follows: we need to extract information from the RGB photographs of the fragments that can be used to reliably measure visual and stylistic similarities and differences among them. There are several factors that make the task challenging: first, the fragments are irregularly shaped, and often substrate material is visible instead of the tapestry layer; secondly, many pieces from the original artwork are missing, and the surviving fragments are highly degraded; and finally, there is no ground truth information available to evaluate the accuracy of the computational solutions. Few works have addressed textile materials specifically. Texture features of a textile are often analyzed using X-ray images to count threads of a canvas for art forensics purposes [8]. As for reconstructing deteriorated textile heritage artefacts, Stoean *et al.* [9] proposed a deep learning solution for inpainting small parts missing from the costumes. To the best of our knowledge, no work has addressed a puzzle of fragmented archaeological textiles from the computer science perspective.

Approach

In this work we use texture analysis to cluster similar fragments together. In image processing, an image texture is defined as the spatial variation of the color or the brightness intensity of the pixel [10]. Texture analysis has been used in a wide range of applications, from texture classification (for example, in remote sensing), to segmentation (for example, in biomedical imaging), or pattern recognition [11]. In our approach, the procedure involves three stages: pre-processing, texture feature extraction, and clustering based on these features. The overall pipeline is shown in Figure 1.

Pre-processing

First of all, the ultra-high-definition RGB photographs are downsampled to lower pixel resolution due to memory limitations. Afterward, the fragments are segmented from the background. In the subsequent step, each fragment is split into smaller patches of 200×200 pixels (see Figure 2). There are two reasons for that: first of all, the number of fragments is low, which makes machine learning unreliable; secondly, as no ground truth is available, we create the ground truth for validation purposes (we know which 200×200 patches belong to the same fragment). The fragments with high degree of noise were discarded. If a fragment contained less than 60 patches, we applied data augmentation by rotating the patches by 90°, 180°, and 270°. Finally, if the image was blurry, image enhancement techniques were used to make them sharper. In total, we have 6650 patches from 77 fragments.

Feature Extraction

We tested three different methods for feature extraction – two classical and one deep-learning based. Firstly, we used Opponent Color Local Binary Patterns (OCLBP) [12]. Local Binary Patterns (LBP) measure structural information and statistical co-occurrences of pixel intensities in the image [13]. Unlike LBP, which is applicable to grayscale images only, OCLBP also incorporates color information in addition to texture, and thus, is more suitable for our images, where color can encapsulate important visual cues. Secondly, following the proposal in [10], we combined OCLBP features with those extracted from Co-occurrence Matrices (CoM) [14]. CoM captures the relative positions of the pixels and represents them as a matrix of probabilities of co-occurrence of certain pixels within a certain distance. Finally, we used pre-trained AlexNet convolutional neural networks (CNN) [15]. Some works [16] propose that this model could be taken as a simple replication of how primate visual system works.

Clustering

After extracting the texture features, we feed them into clustering algorithms to group the patches with similar textures together. Clustering is an unsupervised machine learning method, whose objective is to group data into clusters – minimizing intra-cluster and maximizing inter-cluster differences. We tested three clustering algorithms: K-means [17], Mean-Shift [18], and Agglomerative Hierarchical clustering [19]. While K-means requires the number of clusters to be specified by the user in advance, the other two methods can determine the optimal number of clusters automatically. In K-means, the points are grouped into *k* clusters, where *k* is pre-defined by the user. *K* centroids, i.e. the centers of each cluster are initiated randomly, and then an iterative process runs. On each iteration, the distance between the points and the centroid is found, and each point gets assigned to the cluster of the closest centroid. Afterward, the mean of all points in each cluster is found, which becomes a new centroid. This iterative process is repeated until none of the centroid positions change. Mean-Shift is also a centroid-based method, but it finds the optimal number of clusters automatically. The user, however, defines the radius (called "bandwidth") parameter. For each point, a local mean is found within this pre-defined radius. Given point is shifted to this local mean, and then the new mean to shift a point to is found again from this new position, and the process continues iteratively. If the clusters overlap, the one with more points within a radius is kept. The points shifted toward the same final centroid are concluded to be in the same cluster. Finally, hierarchical clustering is based on a dendrogram – a system similar to a tree, where each individual point is a leaf or a cluster at the finest level of granularity, and the root is a big super-cluster containing all points. Based on a similarity metric, the algorithm gradually groups similar points into the clusters creating the branches, and at the next level, the similar branches are grouped together and connected to the same higher-level branch. Granularity decreases as the cluster size increases and the other way round. The desired level of granularity can be selected by analyzing the dendrogram.

A Baseline Case Study

The approach somewhat similar to the one used in this work has been tested by Gulbrandsen [20]. The author manually cut well-preserved household textiles into irregularly shaped fragments and captured high resolution photographs (the example of the fragments can be found in Figure 3). At the feature extraction stage, the author used color and texture features (color histograms, color moments, color coherence vectors, LBP), as well as VGG16 deep learning model (ImageNet weights), and conducted K-means and hierarchical clustering. The fragments were usually clustered with high accuracy. The hierarchical clustering with VGG16 features performed the best. This demonstrates that the approach is effective in a relatively simple scenario, which can be taken as a baseline case. The objective of our study is to assess the robustness of the approach in a substantially more complex real-life scenario, which comes with a broad range of challenges already discussed above.

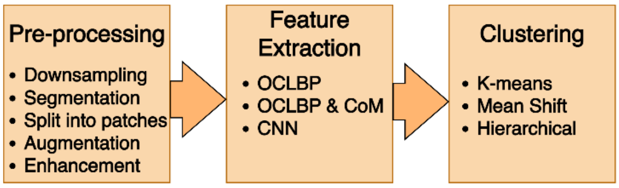

Figure 1. *The workflow of the proposed method.*

Figure 2. *The examples of the patches extracted from the fragments.*

Figure 3. *The example of the well-preserved household textile fragments used by Gulbrandsen [20].*

Results

All three clustering algorithms demonstrated similar accuracy, as well as high correlation between composition of each cluster. Mean-Shift and Hierarchical clustering determined the optimal number of clusters to be two. For K-means, we ran clustering with 2 to 77 pre-defined number of clusters, and the accuracy also turned out to be the highest for two clusters. The results for OCLBP and its combination with CoM are illustrated in Figure 4. The accuracy of the clustering result was measured as follows: we changed the number of clusters k, from 2 to 77 for patches that come from 77 fragment images, which are considered the pseudo-classes. Because this is an unsupervised problem, we used all 6650 patches to fit the clusters. At each class j, the number of patches that are assigned to each cluster (denoted as c) is calculated, which is represented by $p_{c,j}$. The accuracy for each class j is then calculated as:

$$acc_j = \frac{\max\left(p_{c,j}\right)}{N_j} \tag{1}$$

where $c \in [1; k]$, and N_j is the number of patches in a given class (from the same fragment). Finally, we found mean accuracy among all 77 classes and its standard deviation. We want to highlight that this method captures false negatives (patches that we know are from the same fragment and end up in different clusters), and it does not penalize for false positives (patches from different fragments end up in the same cluster). This decision is intentional, because our objective is not clustering patches to 77 original classes. Any false positive may, in fact, indicate that the fragments that those patches come from belonged to the same original. This, however, comes at the cost of risking that the accuracy measure is biased toward lower number of clusters. Therefore, the finding that there can be two original artworks need to be taken with great care.

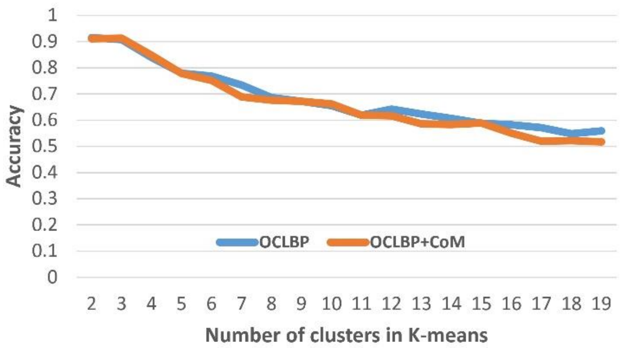

Figure 4. *The accuracy as a function of number of clusters in K-means. The blue curve corresponds to OCLBP features, and the orange curve corresponds to a combination of OCLBP+CoM.*

Let's have a closer look at the clustering results (Figure 5). The fragments in one cluster seem to have thicker threads and lower spatial frequency texture, while in the second cluster, the patches with smaller thread size and higher spatial frequency variation are grouped. This could be an indication that thread thickness is an useful cue to fragment similarity, but as the small patches do not capture global motifs, it can also be misleading if two different artworks were weaved with a similar technique.

Figure 5. *The left and right pair of images were grouped in different clusters, respectively; while the left two patches have thicker threads and relatively low spatial frequency texture, the opposite is true for the right ones.*

As for AlexNet, we extracted feature vector with 507 features. Since there are only 6650 samples in our data, it can be considered as a high dimensional space. Therefore, K-means was not able to perform the prediction on this data because of the curse of dimensionality. The Hierarchical clustering detected two clusters among the patches, with accuracy of 0.954 and the standard deviation of 0.09. However, the groups can be named as textiles and outliers. In the outliers' group, the number of patches is relatively small in comparison with another (200 vs 6450), and it includes mostly extremely degraded homogeneous patches with barely visible variation.

Evaluation by the Archaeologists

The clustering results were evaluated by the archaeologists who work with the Oseberg Tapestry. Although the ground truth is not known, the archaeologists hypothesize which fragments may belong together based on the analysis of the motifs depicted on the fragments and the weaving techniques that they are woven with. The archaeologists pointed out many potential false positives. In other words, the same cluster included the patches from the fragments that are highly unlikely to be part of the same whole due to large differences in the way they are woven as well as their motifs. This makes us conclude that rigorous further work is needed to improve the accuracy and robustness of the algorithm.

Conclusions

There are several conclusions that can be drawn, and several lessons learned from this work: the classical texture classification separates textures in two groups of fine and coarse textures, while deep learning separates textured patches from rather homogeneous ones. This grouping can be far from the original composition of the artworks. One significant limitation is the insufficient number of samples, which can be to some extent mitigated with data augmentation techniques. However, the most fundamental problem is the lack of ground truth information that makes the reliable assessment of the performance impossible (e.g. assessment of false positives). An additional factor that this work has not taken into account is the degradation process of the textiles. Variation in micro contexts in the ground can cause threads and fibers to degrade at different rates, even if they are from the same grave. The experts have several hypotheses on which fragments could belong together – based on the manual analysis of the weaving techniques and interpretation of the high-level motifs [1]. This information should be used in future works. Finally, we cannot rule out that the problem is ill-posed, and RGB photographs simply do not have enough information for telling the fragments apart. For this purpose, future works should additionally rely on more sophisticated imaging techniques, such as hyperspectral or reflectance transformation imaging. We believe that the approach proposed in this work is not limited to archaeological textiles, and it can be extended to a puzzle problem of any type of fragmented cultural heritage artefacts, such as papyrus, inscriptions, engraving, paintings, and ceramics, where the number of original items is not known.

References

[1] Marianne Vedeler, The Oseberg Tapestries (Scandinavian Academic Press, Oslo, Norway, 2019).

[2] Marie-Morgane Paumard, David Picard, and Hedi Tabia, Image reassembly combining deep learning and shortest path problem, In Proceedings of the European Conference on Computer Vision (ECCV), pp. 153–167. (2018).

[3] Marie-Morgane Paumard, David Picard, and Hedi Tabia, “Deepzzle: Solving visual jigsaw puzzles with deep learning and shortest path optimization,” IEEE Transactions on Image Processing, 29, pp. 3569–3581 (2020).

[4] Nicolas Lermé, Sylvie Le Hégarat-Mascle, Boyang Zhang, and Emanuel Aldea, “Fast and efficient reconstruction of digitized frescoes,” Pattern Recognition Letters, 138, pp. 417-423 (2020).

[5] Roberto de Lima-Hernandez, Stan Vincke, Maarten Bassier, Lukas Mattheuwsen, Jens Derdeale, and Maarten Vergauwen, “Puzzling Engine: A Digital Platform to Aid the Reassembling of Fractured Fragments,” International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, (2019).

[6] Roberto de Lima-Hernandez and Maarten Vergauwen, “A Hybrid Approach to Reassemble Ancient Decorated Block Fragments through a 3D Puzzling Engine,” Remote Sensing, 12, 16, 2526, pp. 1-19 (2020).

[7] Roberto de Lima-Hernandez and Maarten Vergauwen, “A Generative and Entropy-Based Registration Approach for the Reassembly of Ancient Inscriptions,” Remote Sensing, 14, 1, pp. 1-27 (2021).

[8] Don H. Johnson, C. Richard Johnson Jr., Andrew G. Klein, William A. Sethares, H. Lee, and Ella Hendriks, A thread counting algorithm for art forensics. In Proceedings of the IEEE 13th Digital Signal Processing Workshop, pp. 679-684. (2009)

[9] Catalin Stoean, Nebojsa Bacanin, Ruxandra Stoean, Leonard Ionescu, Cristian Alesca, Mircea Hotoleanu, Miguel Atencia, and Gonzalo Joya, On Using Perceptual Loss within the U-Net Architecture for the Semantic Inpainting of Textile Artefacts with Traditional Motifs. In Proceedings of the International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), p. 8 pages. (2022)

[10] Laleh Armi and Shervan Fekri-Ershad, “Texture image analysis and texture classification methods - a review,” arXiv preprint, arXiv:1904.06554 (2019).

[11] Linda G Shapiro and George C Stockman, Computer Vision, vol. 3. (Prentice Hall, Upper Saddle River, NJ, 2001).

[12] Topi Mäenpää, The local binary pattern approach to texture analysis: Extensions and applications (PhD Thesis, University of Oulu, 2004).

[13] Timo Ojala, Matti Pietikäinen, and David Harwood, “A comparative study of texture measures with classification based on featured distributions,” Pattern Recognition, 29(1), pp. 51-59 (1996).

[14] Robert M Haralick, Karthikeyan Shanmugam, and Its’Hak Dinstein, “Textural features for image classification,” IEEE Transactions on Systems, Man, and Cybernetics, 6, pp. 610–621 (1973).

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks,” Communications of the ACM, 60, 6, pp. 84-90 (2017).

[16] David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, and Antonio Torralba, Network dissection: Quantifying interpretability of deep visual representations, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6541–6549. (2017).

[17] Stuart Lloyd, “Least squares quantization in PCM,” IEEE Transactions on Information Theory, 28(2), pp.129-137 (1982).

[18] Dorin Comaniciu and Peter Meer, “Mean shift: A robust approach toward feature space analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp. 603-619 (2002).

[19] Fionn Murtagh and Pedro Contreras, “Algorithms for hierarchical clustering: an overview, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,” 2(1), pp.86-97 (2012).

[20] Casper Fabian Gulbrandsen, Reconstructing the Original: Machine Learning Puzzle Assembly for Matching Archaeological Textile Fragments (Master’s Thesis, Norwegian University of Science and Technology, 2023).

Author Biography

Davit Gigilashvili received his BSc degree in Information Technologies from the University of Georgia (2015), MSc in Color and Applied Computer Science from Jean Monnet University, University of Granada, and NTNU (2017), and PhD in Computer Science also from NTNU (2021). Currently he is a postdoctoral researcher at NTNU Colourlab. His research interests include visual perception of material appearance, image processing, and virtual reconstruction of cultural heritage.

Ha Thu Nguyen is currently a master’s student at the Erasmus+ Joint Master program in Computational Color and Spectral Imaging at Norwegian University of Science and Technology, Norway. She obtained her BSc in Electronics and Telecommunications in 2019 from the Hanoi University of Science and Technology, Vietnam. Her research interests include image processing, image quality assessment, and deep learning.

Casper Fabian Gulbrandsen is a master's degree student in Computer Science at the Department of Computer Science at the Norwegian University of Science and Technology. He received his bachelor's degree in Computer Science in 2021. His research interests include the Internet of Things, machine learning, and web and mobile technologies.

Margrethe Havgar is an archaeologist based at the Museum of Cultural History at the University of Oslo. Her main areas of interest are pre-Christian cosmology, ship iconography and burial customs, human - non-human relations, and archaeological dissemination. She is currently working on her Ph.D. thesis within the multidisciplinary research project “TexRec - Virtual Reconstruction, Interpretation and Preservation of the Textile Artifacts from the Oseberg Find”.

Marianne Vedeler holds a position as professor in archaeology at the Museum of Cultural History, University of Oslo, and is co-collection manager of the museums medieval collection. She has 25 years of experience in the study of textile, costume and identity and has published various books and papers on Viking and medieval Archaeology focusing on textiles, dress and food culture.

Jon Y. Hardeberg is a Professor of Colour Imaging at the Colourlab at NTNU - Norwegian University of Science and Technology, Gjøvik, Norway. His current research interests include spectral imaging, image quality, colour management, material appearance, and cultural heritage imaging; he has co-authored more than 300 publications within the field.