Eurographics Workshop on 3D Object Retrieval (2019)
S. Biasotti, G. Lavoué, B. Falcidieno, and I. Pratikakis (Editors)

Sketch-Aided Retrieval of Incomplete 3D Cultural Heritage Objects

S. Lengauer[1], A. Komar[1], A. Labrada[3], S. Karl[2], E. Trinkl[2], R. Preiner[1], B. Bustos[3], T. Schreck[1]

[1]Institute of Computer Graphics and Knowledge Visualisation, Graz University of Technology
[2]Institute of Archaeology, University of Graz
[3]Millenium Institute for Foundational Research on Data, Department of Computer Science, University of Chile

(a) 3D Input (b) Alignment (c) Sketch (d) Inpainting (e) Retrieval Results

Figure 1: *(a) The proposed pipeline takes an incomplete 3D model as input, which is (b) properly aligned, (c) completed by the user with sketching and (d) automatically filled with texture. (e) The rendered 2D image serves as input for a content-based similarity retrieval.*

Abstract

Due to advances in digitization technology, documentation efforts and digital library systems, increasingly large collections of visual Cultural Heritage (CH) object data becomes available, offering rich opportunities for domain analysis, e.g., for comparing, tracing and studying objects created over time. In principle, existing shape- and image-based similarity search methods can aid such domain analysis tasks. However, in practice, visual object data are given in different modalities, including 2D, 3D, sketches or conventional drawings like profile sections or unwrappings. In addition, collections may be distributed across different publications and repositories, posing a challenge for implementing encompassing search and analysis systems. We introduce a methodology and system for cross-modal visual search in CH object data. Specifically, we propose a new query modality based on 3D views enhanced by user sketches (3D+sketch). This allows for adding new context to the search, which is useful e.g., for searching based on incomplete query objects, or for testing hypotheses on existence of certain shapes in a collection. We present an appropriately designed workflow for constructing query views from incomplete 3D objects enhanced by a user sketch based on shape completion and texture inpainting. Visual cues additionally help users compare retrieved objects with the query. We apply our method on a set of relevant 3D and view-based CH object data, demonstrating the feasibility of our approach and its potential to support analysis of domain experts in Archaeology and the field of CH in general.

1. Introduction

With advancing technologies, more and more Cultural Heritage (CH) content becomes available as digital objects. The content is typically given in a number of different modalities, including 2D images, 3D models, sketches, conventional drawings like profile sections and unwrappings, or related texts. At the same time, existing data are in most cases only weakly cross-linked, with data being spread over a vast number of printed publications, web repositories and web pages, virtual museums, etc. Domain expert knowledge is required to find similar objects or detect links, especially if the reference data are only available as illustrations in printed volumes with little or no associated meta data. The comparison of shapes is a fundamental task in archaeological research. As the amount of vases is huge a digitally enhanced system for object retrieval will be welcomed in the archaeological community. A major challenge for a computer-aided search is that only a fraction of the ex-

DOI: 10.2312/3dor.20191057

cavated CH objects are complete, but most of them are present in various degrees of fragmentation or erosion, making it difficult to use them directly as input for shape comparison and search. Our proposed method addresses this issue by defining an appropriate workflow for content-based retrieval of 2D image data from incomplete 3D objects. The workflow is built around a *human-in-the-loop* approach, allowing experts to provide sketch-aids for the query weighting, and visual result comparison.

Missing parts of CH objects can be estimated with high precision by domain experts. The basic idea of sketch-aids is to allow users to create additional object structure, which is filled by a texture inpainting step. Fig. 1 shows an example of the retrieval workflow from an incomplete 3D object to the completed 2D image, which serves as input for content-based retrieval. In the remainder of the paper, we define the approach in detail and show the advantage of sketch-aided queries by several use cases on actual archaeological object data. Compared to purely sketch-based approaches, our concept has the advantage that available texture and shape information is used to enhance the retrieval process. Archaeological research is often interested in finding object analogies, emphasizing or de-emphasizing certain characteristics. Our user interface supports this by detailed shape comparison views and the possibility for interactive weighting of the queries.

2. Related Work

2.1. 3D Object Retrieval and Features

Researchers to date have investigated many approaches to 3D object retrieval, based on various similarity concepts such as shape, structure, appearance, or metadata [TV08]. Similarity can be based on whole objects, or on parts of objects [SPS14]. Retrieval methods can also be distinguished by object modalities. Besides searching with a given 3D object within a set of 3D objects, one can also search with and within views (images) of 3D objects, drawings, sketches, or video. View-based approaches are frequently used, as views can be a common denominator among different modalities. Often, a feature-based approach is implemented, where the objects are described by a feature vector (or descriptor) based on which similarity scores can be computed [Sch17]. Sketching [LLG*14] as a query modality has the advantage, that no query object or view is required, and that it provides a natural user interface. Sketching also is connected to modeling, and previous works have proposed modeling based on example objects found by sketch, or applying generative modeling methods detected from sketches [LF08]. Features can be distinguished as engineered features and learned features. The first are measurements extracted from a 3D object or view, according to expert defined properties. Examples include measures from object surface, volume, or structure (e.g., skeleton) [LLC*15], or view-based features including e.g., depth maps, silhouettes, gradients [DT05], or interest points [Low04].

Recent success of deep learning models for image classification and retrieval has prompted for application of these models also to 3D retrieval. Two main approaches are followed to input 3D object data to deep networks, representing 3D models either as a set of 2D views or as a voxel grid. Generally, models that use 2D views surpass voxel-based models. However, the latter may perform better if more complex neural network models are used, being much more expensive to train and also need a substantial amount of relevant data for training. Ioannidou et al. give a survey of methods using deep learning techniques on 3D models [ICNK17]. In particular, trained deep models have been used for feature extraction, yielding *learned* features, which have been used e.g., for sketch-based 3D Shape Retrieval [WKL15] and to implement a 3D retrieval system [BBZ*17]. Recent results show that if proper training data can be used, learned features can outperform engineered features in terms of retrieval and classification accuracy. However, if a properly trained network is not available, e.g., due to the lack of training data, it can still be more feasible to rely on engineered features that incorporates domain knowledge about the expected target set. The approach presented in this paper is independent from the chosen type of features. In this work, we focus on engineered features (Section 3.4), as preliminary results on learned features from a general-purpose network were not able to produce meaningful results for the kind of domain specific data used in our experiments.

2.2. 3D Objects in Archaeological Research

Although nowadays, 3D data are often incorporated in archaeological documentation and visualization, it is still rarely used for archaeological analysis. One of the key issues in dealing with archaeological objects is a thorough classification of artifacts, which is often affected by the large amount of available objects. Searching for relevant object collections is thus fundamental to archaeological research. They are provided by conventional publications and recently also more often by online repositories. For the research on Greek pottery highlighted in this paper, useful starting points are the series of the Corpus Vasorum Antiquorum (`www.cvaonline.org/cva/`) and the Beazley Archive, Oxford (`www.beazley.ox.ac.uk/index.htm`).

To date, 3D retrieval methods have been used to support applications and research in different projects in the CH field. A recent survey of geometric analysis techniques for CH applications is given in [PPY*16]. In [PSA*17], a workflow for reassembly and completion of incomplete objects is proposed, using similarity in several key stages of the workflow. In [PSM*18], predictive scanning of CH objects is proposed. Based on a partial view acquired from an object, similar objects are retrieved and fitted to complete the partial view, based on local shape descriptors and nonrigid registration. In [BID*17] a pipeline is proposed for reconstructing 3D objects from profile sections as a basis for a reference set for retrieval. In [GMM*17] introduced a preliminary approach for searching views of 3D CH objects in image data, using descriptors based on color and shape. While building on the same problem, our work presents a much more encompassing search approach considering geometry, texture, complementary user sketches, and useful interface extensions for query weighting and result visualization on relevant 3D and image data.

3. Concept for Sketch-Aided 3D Retrieval

Our concept supports a similarity search over a database of 2D images with an incomplete 3D object as query input. Due to the different modalities of query and target (3D+sketch, images), feature extraction is not directly applicable, but several processing

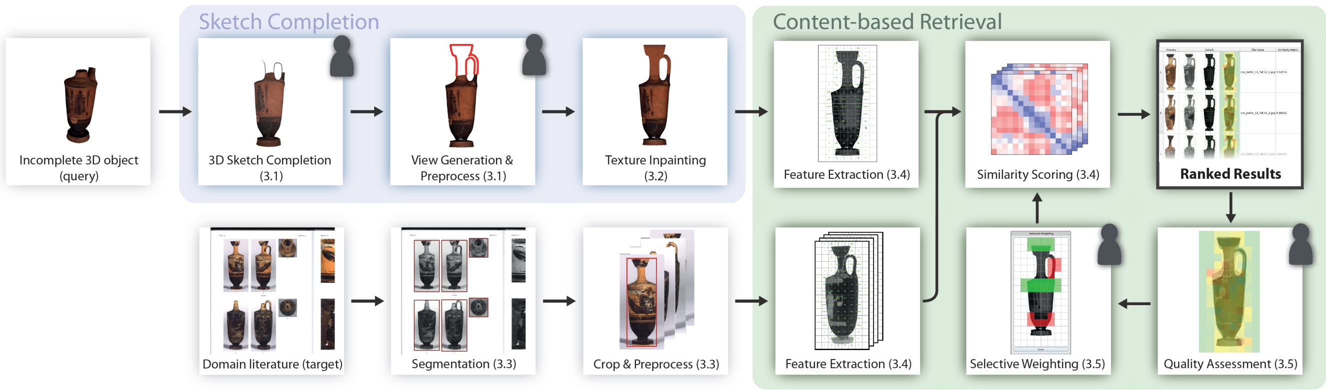


Figure 2: *Concept for implementing shape search using hybrid 3D+sketch queries. The sketch-completed input object is transferred to 2D modality space and compared to the 2D target images resulting in a ranked list of results. Optionally, the results can be iteratively evaluated and refined by region-based selective weighting. Steps requiring user feedback are indicated with a characteristic icon.*

steps are need. Fig. 2 shows our proposed processing pipeline. For the query object, the first step includes determining an appropriate viewpoint and completing the geometry by sketch (Section 3.1). Subsequently, a rendering of the sketch-completed object serves as basis for a texture inpainting step, which is based on previous texture synthesis techniques (Section 3.2). The result is a completed query usable for similarity search (Section 3.4). After the retrieval, the query can be further refined by applying selective weighting to the inpainted query image (Section 3.5). This allows the user to refine the search by shifting the focus to crucial parts of the geometry or disregarding parts that are irrelevant or misleading for the search task at hand. As target objects we consider images in a comprehensive archaeological domain publication series. The generation of the target database involves segmentation of plates from the publications, which typically depict multiple images together with meta data. The extracted object images also require further preprocessing steps, discussed in Section 3.3.

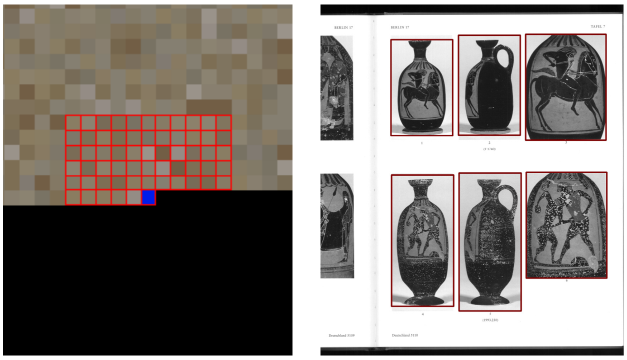
3.1. Contour Completion of Query Object with Sketch

The alignment and sketching is done in a simple 3D editor with orthographic projection. From our collaborating archaeologists, we learned that the depictions in the considered domain publications come more close to an orthographic projection, than to a perspective one. Prior to the sketching process an appropriate viewpoint is determined by rotating the model accordingly. The sketch-plane is defined by the camera's up- and right-vector. The offset in the viewing direction is determined when the first sketch line is drawn (a ray originating from the camera, piercing through the mouse position is intersected with the 3D model). The sketch consists of a cuboid, whose start points are connected to the previous end points if given. An expert user, placing the start and the end of the outline on top of the existing geometry, is assumed. This step is concluded by exporting a non-photorealistic render of the current view.

3.2. Texture Inpainting Step

The input to the texture inpainting step is a 2D image with a sketched contour. The region to be filled, as well as the region containing the necessary contour, are defined by a labelling algorithm together with the user identifying the respective regions. First, the edges of the image are calculated using Sobel edge detection. The edge image is then subjected to Gaussian blur and thresholded to close small gaps and making the edges unambiguous. The resulting image is subjected to connected components labeling [SS01, p.69], consisting of 2 passes. In the first run, the algorithm determines the connectivity by looking at the local 4-connectivity of each pixel, assigning either a new or an existing label to the pixel. In the second pass this data structure is then used to label each pixel.

The labels determine the segment which will be inpainted (target segment) as well as the segment providing texture information (source segment). Occurring gaps between sketch lines and the target segment are counteracted by slightly growing the segment. The inpainting algorithm described by [WL00] was combined with a descriptor loosely based on the RIFT descriptor proposed by [SL05]. For each pixel a color histogram of the neighborhood is computed. The applied descriptor is loosely based on the RIFT descriptor proposed by [SL05] and uses a color histogram over the neighborhood of the center pixel. Different neighborhood sizes were considered (e.g., five-pixel neighborhood in Fig. 3a) before we chose a nine-pixel pixel neighborhood. Black, white and gray pixels in the neighborhood were ignored to discriminate the background and sketch lines. We also found 12 histogram bins to be a good compromise between performance and quality. The source descriptors are organized in a k-d Tree of dimension $k = 12$ bins $\times$ 3 color channels $= 36$. The inpainting is conducted row-wise from the top-most and left-most pixel of the target segment. For each pixel the descriptor of the local neighborhood is calculated and the L2 distance to all source descriptors in the k-d Tree is determined. From those with a distance below an empirically determined threshold, one is determined randomly and the color of its respective source pixel applied to the current target pixel. If no neighbor within this threshold is found, the pixel color of the source descriptor with the lowest distance is taken. Finally, a median filter is applied to remove noise possibly caused by the sampling.



(a) (b)

Figure 3: *(a) Pixel descriptor footprint for a given pixel (blue) (b) Example domain document [ZE18, pl.7], scanned and converted to grayscale for the extraction of target images (framed in red).*

(a) (b) (c)

Figure 4: *(a) Sketch interface. (b) Quality Assessment with query- and target image, superimposed with a heat map indicating the cell-wise difference of the feature vectors (red is large). (c) Selection tool for marking* important *(green) and* unimportant *cells (red).*

3.3. Target Image Extraction and Preprocessing

The target images are extracted from scans of relevant archaeological literature. The goal is to obtain a comprehensive set of target images with each image depicting exactly one object. A first step involves detecting the images present on a given plate, which contains typically between two and eight images, together with text and numbering information. This is achieved with the contour detection algorithm presented in [S*85], after applying a Gaussian Blur and a binary threshold. With further restrictions in terms of width and height of the detected contours, false matches like pagination or text paragraphs have been ruled out. Images depicting close-ups, or paintings on the surface of an object like the two examples on the right side of Fig. 3b, and such depicting objects from a top view have been removed manually from the target set. A typical scanned page with the detected images marked is depicted in Fig. 3b.

To remove the background of the extracted images, a Canny edge detection was used. The resulting edge image gets dilated for 3 iterations, to close the contour of the displayed object. The previously mentioned contour detection is used again, and the contour with the largest area is assumed to be the contour of the object. It is dilated for 10 iterations and then eroded for 10 iteration to smooth out the edges. A Gaussian blur then is applied. This mask is used to separate the foreground and the background. The background subtraction step was taken to detect the outline of the object. Finally, the image then gets cropped so that the depicted object takes up 90% of the width and 90% of the height of the image and is centered.

3.4. Retrieval of Target Images

After the sketch-completion, texture inpainting and preprocessing steps, the query and target collection are available, allowing for a conventional image feature-based similarity search. There exists a wide range of established 2D image features, incorporating global features like Color Histogram, Edge Histogram, Tamura, Color and Edge Directivity Descriptor (CEDD) [LM13, p.30-40], Histogram of Oriented Gradients (HOG) [DT05] and local features like Scale Invariant Feature Transform (SIFT) [Low04]. Global features are computed from the whole image and can be further divided into features based on color, texture and shape. The group of local features relies on "significant" points in the image, which noticeably differ from their neighborhood. We found that color-based features performed poorly for our specific use case of pottery images, as they exhibit mostly undiscriminating color distributions, and many target images are available only in grayscale at the first place.

In preliminary experiments, global shape-based feature descriptors yielded the most promising results, due to both the query and the search space images being depicting with the whole object in view and with a characteristic orientation. Scale- and rotation invariance turned out to be of little concern due to the alignment and position of the reference objects following certain convention (e.g., cylindrical shapes are photographed standing up from one or more sides with a handle, if present, on the left or the right side). We found that the HOG feature descriptor yields appropriate results and fits to our proposed processing pipeline well. HOG describes an image by an array of local histograms of an image, and hence can be locally weighted by a user, if required. We chose an image size of 64 by 128 pixels, resulting in a feature vector of 3780 float values. The feature vector of each image in the search space is used to compare to the feature vector obtained from the query image using the weighted Euclidean L2 single distance metric $d_{L2}(x,y) = \sqrt{\sum_i (x_i - y_i)^2 w_i}$, with globally uniform weights $w_i = 1$. Note that in our approach the used type of features and distance metric are exchangeable. Integration of other features using feature combination or selection methods would be straightforward.

3.5. User Weighting of Target Object Areas

The output of the described retrieval system is a list of images from the search space, ranked by similarity to the query according to above metrics. In its standard configuration, these metrics assign a globally uniform importance to all regions of the query image. This might lead to suboptimal rankings that are due to an improper assessment of importance in different regions of the image. From

Figure 5: *Target search space of more than 300 objects extracted from digitized domain publications [ZE13, ZE18].*

Figure 6: *3D objects of Landesmuseum Kärnten (from left, Inv.-Nr.: 1245, 1248, 1251, 1252, 9049 and 9050) in various stages of fragmentation, sketch-completed by domain experts (top) and the result after texture inpainting (bottom).*

our archaeologist research partners we learned that they would like to emphasize and compare certain parts of their query objects. For example, the head or handle of a vessel may or may not be of importance, depending on the domain comparison task at hand. To allow the user to adjust the result set, the proposed pipeline shown in Fig. 2 includes two further steps to this end:

Quality Assessment. The HOG descriptor provides an array of local image descriptors [DT05]. Hence, we can compute the Euclidean distance between all pairs of cells of two images, and visualize it with a semitransparent heat map, superposed to the input image. The heatmap view allows to effectively grasp local similarities and dissimilarities betweeen a pair of images. An exemplary result is depicted in Fig. 4b.

Selective Weighting. The quality visualization described above supports the user in the next step, which allows for concentrating the focus on specific regions based on his domain knowledge, or disregarding other regions where shape similarity might be less relevant. To this end, our software allows the user to interactively select areas of higher or lower importance based on cells, as shown in Fig. 4c. This input results in a user-defined weight vector w used in the distance metric, where we empirically chose weights $w_i = 5.0$ for "important" cells (green), $w_i = 0.0$ for "unimportant" cells (red) and $w_i = 1.0$ for all other cells (default).

4. Implementation and Application to Archaeological Data

We implemented our proposed sketch-aided approach, and next demonstrate the applicability and effectiveness. We first give details on our implementation (Section 4.1), an overview of the used data experimented on (Section 4.2), and then present an encompassing use-case based evaluation, demonstrating its benefits for supporting archaeological research (Section 4.3). The section closes with a discussion of limitations and extension possibilities (Section 5).

4.1. Implementation

The sketch interface is based on the Ogre3D library, which was extended for our needs. It is capable of loading objects given as PLY-files and supports sketching directly on top of the 3D structure. However, the individual stages of our retrieval pipeline in Fig. 2 are independent from one another. Thus, arbitrary 3D and graphics editors can be used to generate the input for the texture inpainting step. For the content-based retrieval (Section 3.4) from the inpainted 2D image a purposely built separate application was used. It supports a variety of local and global features together with different distance metrics. For the results presented in the following section, our implementation used the OpenCV Open Source Computer Vision library, both for tasks of image preprocessing (Section 3.3) and feature extraction (Section 3.4).

4.2. Data Sets

We worked with research partners from Archaeology (who also co-authored this publication), to obtain 17 3D objects, provided by courtesy of Landesmuseum Kärnten, Austria. They are high-resolution textured 3D scans representing Attic black-figured lekythoi from the first half of the 5th century BC., a very common vessel type in Greece for that period. Seven objects, with Inv.-Nr. 1245, 1248, 1251, 1252, 1253, 9049 and 9050 (respective 3D models added as attachment with file names like C8_450_Klagenfurt_{*Inv.-Nr.*}.ply), were selected and parts of their geometry were removed, to mimic the effects of a poor conservation status, e.g., with broken handle, mouth, etc. Different degrees of fragmentation have been created. The subsequent reconstruction by sketch (see Section 3.1) was done by domain experts in Inkscape according to their knowledge of Greek pottery (in our case Attic black-figured lekythoi [Has36]). Some of the selected objects as well as their sketch-completion and texture inpainting are depicted in Fig. 6. Corresponding 2D images were extracted from the prominent Corpus Vasorum Antiquorum (CVA) domain publication. Our search space consists of 114 images from CVA Berlin 13 [ZE13] and 230 images from CVA Berlin 17 [ZE18], all illustrating lekythoi, in different sub-shapes and styles (black-figure as well as red-figure). Images depicting only parts of an object and images of the object from an uncommon viewpoint were not included in our set of target images. A part of the objects we automatically extracted from CVA is given in Fig. 5. All of them have been subjected to automatic preprocessing steps referred to in Section 3.3.

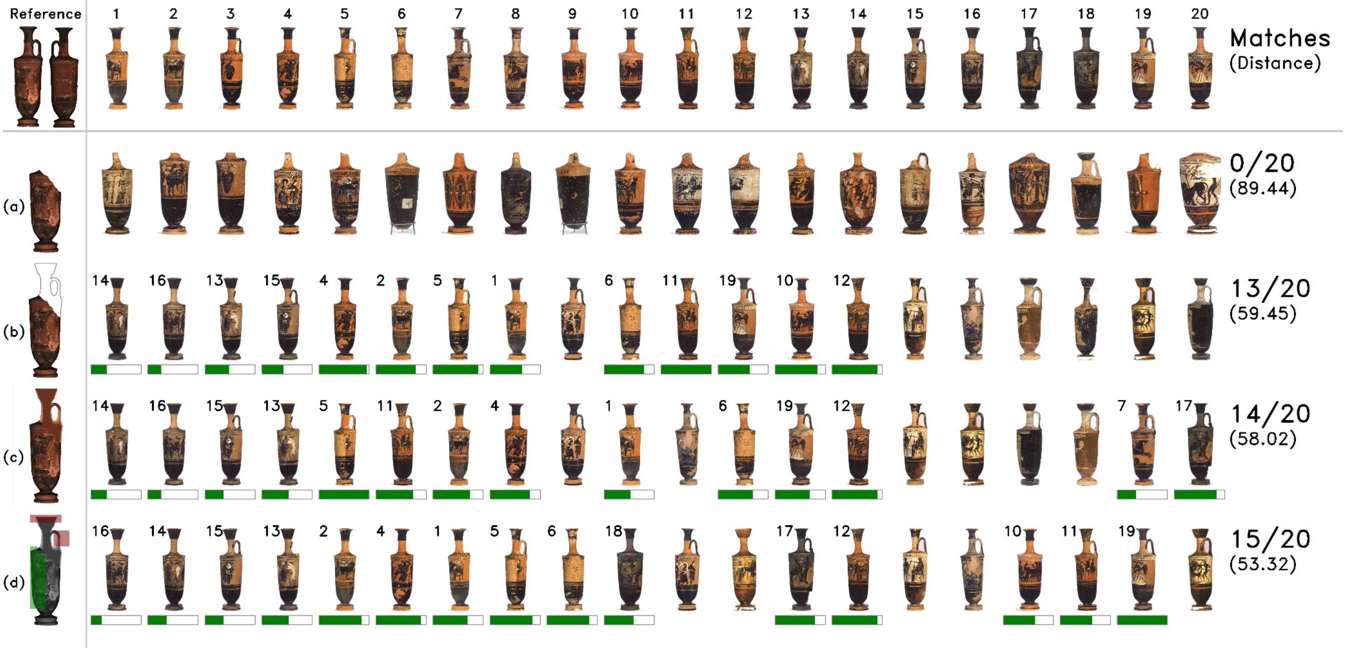


Figure 7: *Ranked result sets from retrievals using an incomplete query object (a) after sketch-completion (b), additional texture inpainting (c) and additional selective weighting (d). The top row shows the reference ranking of object according to archaeological knowledge. If a result object is present in this reference set (*match*), a green bar indicates the accuracy of its ranked position compared to the reference ranking. The right hand side displays the number of such matches with the reference set and the L2 distance between the ranking vectors.*

4.3. Results

We evaluate the benefit of sketch-aiding and feature weighting for the search result in a retrieval system, with a specific focus on domain tasks in archaeological object comparison. To this end, we use one particular vessel (*C8_450_Klagenfurt_9049.ply*) exhibiting both a reasonable amount of missing geometry as well as a realistically placed fracture line, as shown in Fig. 7a. In all our experiments the HOG feature descriptor (Section 3.4) was used.

Reference Ranking. To obtain a reference result set for our evaluation, our Archaeologists determined, from the set of all CVA objects, 20 ranked target objects (based on the incomplete 3D object) most similar by domain consideration. This ranking has been established using a holistic approach by considering the vessel shape and the style of painting in toto. They differentiated the images based on the main Greek painting styles (i.e., black figured, red figured), and then performed an exclusion of specific shapes, i.e., bulgy shaped vessels. Based on a common approach in Archaeology, looking for analogies, the other objects were ranked descending on their similarity to the 3D query object.

Retrieval without Sketch-Aid. Fig. 7b displays the top 20 images of the ranked result set retrieved using the incomplete input object. It can be seen that the applied HOG descriptor is able to retrieve images visually similar to the query image, ranking similarly thin shaped vessels first. However, a quantitative evaluation towards the reference ranking shows that none of these best ranked images is present in the reference set, i.e., they do not reflect the experts' understanding of similarity. This is an expected result, as the HOG shape-based descriptor cannot compensate for missing geometry.

Sketch-Aided Retrieval. Fig. 7c illustrates the ranked results based on sketch-completed query objects. A visual inspection of the top 20 results shows a much stronger resemblance to the original complete object, shown in the upper left part of the image. An improvement of results can be clearly observed with 13 of the top 20 matches corresponding to the reference set. Besides a mere matching count, we also investigated the similarity of the rankings in detail. To this end, we measure for each matching image the deviation of its ranking from its ranking in the reference set, shown as green bar in Fig. 4.3, with a full bar indicating the exact same position. The image also depicts the ranking similarity of the entire set via the Euclidean distance of its ranking vector $r = \{14, 16, 15, \ldots\}$ to reference ranking vector $\{r_i = i\}$, where a maximum distance of 20 is used for images that do not appear in the reference set.

Texture Inpainting. The results from the texture inpainting step is displayed in Fig. 7c. While the majority of the ranking remained unchanged compared to row (b) there is a noticeable improvement with 14 out of 20 matches.

Selective Weighting. By incorporating expert knowledge, the quality of the results can be further refined by adjusting the weighting for different regions, as described in Section 3.5. In the case of our example object, our Archaeologists specified a focus on regions containing original texture and shape, and disregarded the importance of the presence of a handle, knowing from domain experience that similarity relations of lekythoi have to be detected in vessel profiles, whereas handles as attachments are not significant [Has36]. Fig. 7d shows the corresponding weighting (left) and the ranking results after incorporating these weights. We observe a further improvement of the number of matches as well as an even smaller distance to the reference ranking vector. Further results and evaluations are given in supplemental material.

5. Discussion and Future Work

5.1. Applicability

From our use case, we conclude that the incorporation of user sketch-aids and weighting can help to improve and refine domain-specific search tasks on incomplete shapes. With the sketch interface, the domain expert can quickly fill in missing parts, and compare the query visually with target objects. Sketching is an effective tool to test hypotheses, i.e., if certain object shapes or variants are present in a collection. Our experiments showed that sketch-aided retrieval is generally more effective. On the one hand the sketch allows to determine the overall extent of the object more intuitively than just providing a bounding box. This is necessary for proper rescaling, especially so if parts at the top or the bottom of the object are missing. In such cases it is the main driving factor for improvement, as shown in the example in Fig. 7. On the other hand, the sketch is a necessary prerequisite for the texture inpainting where it serves as a boundary. The gradients introduced by this step enable the use of the shape in these previously missing regions, improving the search results even further. Extending the currently used HOG features with texture features could further improve the impact of the sketch-aid on recall. In our tests it was established that the selective weighting provides a reasonable means for domain experts to incorporate their domain knowledge.

A problem may arise if the amount of missing geometry is too large for the expert to recognize a plausible completion variant. The concept of sketch-aided search has shown in our experiments to be of good use in cases where the absence of a part changed the overall shape, e.g., absence of a handle or a spout. In such cases, a distinct improvement of retrieval results could be observed. As expected, sketch-completing only small or irrelevant missing parts showed no obvious advantage over standard retrieval for incomplete objects. We also note that we evaluated our sketch-aided retrieval approach on learned features using a pre-trained general purpose neural network. However, using the query object (a) and (c) shown in Fig. 7, these features produced only between 0 and 2 out of 20 matches. We presume that general purpose pre-trained networks may not produce sufficiently specialized features for the domain specific class of objects used in our evaluations. Future work will thus investigate features from a specifically trained networks.

5.2. Limitations

The presented processing pipeline is tailored based on certain assumptions, which also induce some limitations. While the scale variance of the applied HOG descriptor is in most cases of no concern due to proper cropping and rescaling, it might become an issue for target space images with missing parts, like the lekythos with a missing spout depicted in Fig. 8a. Since the bounding box of the displayed object is the basis for rescaling, the expanse of the complete vessel cannot be estimated in such cases.

(a) (b) (c)

Figure 8: *The concept is limited search space images with parts missing from their geometry (a), rotation of the object itself (b) and bad texture inpainting results under specific conditions (c).*

Query image extraction currently relies on manual view point selection, which can influence the query result, e.g., a vessel having a handle could be displayed with the handle on either side in the search space images (Fig. 8b). Using local image descriptors or multiple views on the query object can improve the matching accuracy of our pipeline.

Texture inpainting faces difficulties if the available regions with original texture are comparatively small or untypical for the region to be filled. Additionally, many of the reference objects have areas with split-offs and restoration work, leading to a distorted texture representation at these regions. Fragments in the inpainted surface may occur if narrowings of the shape are present along the path (top to bottom) like the pedestal of the vessel depicted in Fig. 8c.

5.3. Future Work

Our approach currently supports low-level editing of the query shapes. Improved editing could make use of semantic-sketching techniques [ZLW*14]. However, semantic sketching requires appropriate generative modeling procedures to cover the application domain, which may be expensive to obtain for many different shape types. Another idea is to guide the user while sketching in an online fashion, based on available target data, following a shadow-drawing approach [LZC11]. Also, methods for compensating user-induced inaccuracies in the sketch lines (e.g., gaps between lines) will be investigated. We used incomplete objects that were produced artificially from complete ones, and of a specific domain. We plan to extend the experimental data with more object types and larger numbers of target objects and queries. Considering that more and more repositories of CH objects become available, the need for content-based search tools becomes pressing. We also plan to enhance the result visualization with metadata from the target repositories, like inclusion of domain texts, spatiotemporal information, and other metadata which adds context to the search. In addition, clustering of results for similar object groups is considered useful.

As stated, our approach can accommodate different feature types. It will be interesting to analyze in detail the potential of learned features, and compare their performance with engineered features, given the specificity of the considered search domain (CH objects) and limited available training data.

6. Conclusion

We introduced a new query modality, sketch-aided 3D retrieval. Moreover we presented an appropriate workflow and implementation, which we applied to a representative application in the archaeological research domain, informed by cooperation with domain researchers. Our first results show that the aiding of user sketches to complete and modify the 3D query object, and subsequent texture inpainting, help to improve the search. Specifically, the query modality provides the flexibility to search based on incomplete shapes, and to explore hypotheses of possible shapes in a target repository. While the sketch-completion can be done by anyone, better knowledge will lead to a better sketch and in turn to a better retrieval. Result visualization and weight adaption support the analytical retrieval process. Future work will include an extended experimental evaluation, inclusion of other engineered and/or learned features, and extension of the sketch interface.

Acknowledgments

This work was co-funded by the Austrian Science Fund FWF and the State of Styria, Austria within the project *Crossmodal Search and Visual Exploration of 3D Cultural Heritage Objects* (P31317-NBL) as well as the Millennium Institute for Foundational Research on Data (IMFD), Chile. We thank Pavlos Mavridis, Microsoft Corporation for his valuable suggestions made to this project and the Landesmuseum Kärnten for providing 3D models. We also thank N. Zimmermann-Elseify (Berlin, Antikensammlung) for granting the permit to use her CVA-volumes.

References

[BBZ*17] BAI S., BAI X., ZHOU Z., ZHANG Z., TIAN Q., LATECKI L. J.: GIFT: Towards scalable 3D shape retrieval. *IEEE Trans. Multimedia 19*, 6 (2017), 1257–1271. 2

[BID*17] BANTERLE F., ITKIN B., DELLEPIANE M., WOLF L., CALLIERI M., DERSHOWITZ N., SCOPIGNO R.: Vasesketch: Automatic 3d representation of pottery from paper catalog drawings. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (2017), vol. 1, IEEE, pp. 683–690. 2

[DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *CVPR (1)* (2005), pp. 886–893. 2, 4, 5

[GMM*17] GREGOR R., MAYRBRUGGER C., MAVRIDIS P., BUSTOS B., SCHRECK T.: Cross-modal content-based retrieval for digitized 2d and 3d cultural heritage artifacts. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage* (2017), Eurographics Association, pp. 119–123. 2

[Has36] HASPELS C.: *Attic Black-figured Lekythoi*. Travaux et mémoires / École française d'Athènes 4. "de" Boccard, 1936. 5, 7

[ICNK17] IOANNIDOU A., CHATZILARI E., NIKOLOPOULOS S., KOMPATSIARIS I.: Deep learning advances in computer vision with 3d data: A survey. *ACM Comput. Surv. 50*, 2 (2017), 20:1–20:38. 2

[LF08] LEE J., FUNKHOUSER T. A.: Sketch-based search and composition of 3d models. In *SBM* (2008), pp. 97–104. 2

[LLC*15] LEI H., LI Y., CHEN H., LIN S., ZHENG G., LUO X.: A novel sketch-based 3D model retrieval method by integrating skeleton graph and contour feature. *Journal of Advanced Mechanical Design, Systems, and Manufacturing 9*, 4 (2015). 2

[LLG*14] LI B., LU Y., GODIL A., SCHRECK T., BUSTOS B., FERREIRA A., FURUYA T., FONSECA M., JOHAN H., MATSUDA T., OHBUCHI R., PASCOAL P., SAAVEDRA J.: A comparison of methods for sketch-based 3D shape retrieval. *Computer Vision and Image Understanding 119* (February 2014), 57–80. 2

[LM13] LUX M., MARQUES O.: Visual information retrieval using java and lire. *Synthesis Lectures on Information Concepts, Retrieval, and Services 5*, 1 (2013), 1–112. 4

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 2, 4

[LZC11] LEE Y. J., ZITNICK C. L., COHEN M. F.: ShadowDraw: real-time user guidance for freehand drawing. *ACM Trans. Graph. 30*, 4 (2011), 27:1–27:10. 7

[PPY*16] PINTUS R., PAL K., YANG Y., WEYRICH T., GOBBETTI E., RUSHMEIER H.: A survey of geometric analysis in cultural heritage. *Computer Graphics Forum 35*, 1 (2016), 4–31. 2

[PSA*17] PAPAIOANNOU G., SCHRECK T., ANDREADIS A., MAVRIDIS P., GREGOR R., SIPIRAN I., VARDIS K.: From reassembly to object completion - a complete systems pipeline. *Computing and Cultural Heritage 10*, 2 (February 2017). 2

[PSM*18] PRATIKAKIS I., SAVELONAS M. A., MAVRIDIS P., PAPAIOANNOU G., SFIKAS K., ARNAOUTOGLOU F., RIEKE-ZAPP D.: Predictive digitisation of cultural heritage objects. *Multimedia Tools Appl. 77*, 10 (2018), 12991–13021. 2

[S*85] SUZUKI S., ET AL.: Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing 30*, 1 (1985), 32–46. 4

[Sch17] SCHRECK T.: What features can tell us about shape. *IEEE Computer Graphics and Applications 37*, 3 (2017), 82–87. 2

[SL05] S. LAZEBNIK C. SCHMID J. P.: A sparse texture representation using local affine regions. *IEEE Trans. Pattern Anal. Mach. Intell.* (2005), 1265–1278. 3

[SPS14] SAVELONAS M., PRATIKAKIS I., SFIKAS K.: An overview of partial 3D object retrieval methodologies. *Multimedia Tools and Applications* (2014), 1–26. 2

[SS01] STOCKMAN G., SHAPIRO L. G.: *Computer Vision*, 1st ed. Prentice Hall, 2001. 3

[TV08] TANGELDER J., VELTKAMP R.: A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications 39*, 3 (2008), 441–471. 2

[WKL15] WANG F., KANG L., LI Y.: Sketch-based 3d shape retrieval using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015* (2015), pp. 1875–1883. 2

[WL00] WEI L.-Y., LEVOY M.: Fast texture synthesis using tree-structured vector quantization. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2000), SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., pp. 479–488. 3

[ZE13] ZIMMERMANN-ELSEIFY N.: *Corpus Vasorum Antiquorum Berlin, Antikensammlung 13, Deutschland 93*. München, 2013. 5

[ZE18] ZIMMERMANN-ELSEIFY N.: *Corpus Vasorum Antiquorum Berlin, Antikensammlung 17, Deutschland 102*. München, 2018. 4, 5

[ZLW*14] ZENG L., LIU Y.-J., WANG J., ZHANG D.-L., YUEN M. M.-F.: Sketch2jewelry: Semantic feature modeling for sketch-based jewelry design. *Computers & Graphics 38* (Feb. 2014), 69–77. 7