

Special Section on 3DOR 2019

## A sketch-aided retrieval approach for incomplete 3D objects

Stefan Lengauer<sup>a,\*</sup>, Alexander Komar<sup>a</sup>, Arniel Labrada<sup>c</sup>, Stephan Karl<sup>b</sup>, Elisabeth Trinkl<sup>b</sup>, Reinhold Preiner<sup>a</sup>, Benjamin Bustos<sup>c</sup>, Tobias Schreck<sup>a</sup>



<sup>a</sup> Institute of Computer Graphics and Knowledge Visualisation, Graz University of Technology, Inffeldgasse 16c Graz 8010, Austria

<sup>b</sup> Institute for Classics, University of Graz, Austria

<sup>c</sup> Millenium Institute for Foundational Research on Data, Department of Computer Science, University of Chile, Chile

### ARTICLE INFO

#### Article history:

Received 4 September 2019

Revised 29 January 2020

Accepted 5 February 2020

Available online 14 February 2020

#### Keywords:

Image processing

Shapes

Information storage and retrieval

Information systems

### ABSTRACT

With the growing amount of digital collections of visual CH data being available across different repositories, it becomes increasingly important to provide archaeologists with means to find relations and cross-correspondences between different digital records. 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, and often only in incomplete or fragmented state, posing a particular challenge for conventional similarity search approaches. In this paper we introduce a methodology and system for cross-modal visual search in CH object data that addresses these challenges. 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. The proposed approach extends on a previously presented retrieval system by introducing improved retrieval methods, an extended evaluation including retrieval in a larger and richer data collection, and enhanced interactive search weight specification. We demonstrate the feasibility and potential of our approach to support analysis of domain experts in Archaeology and the field of CH in general.

© 2020 Elsevier Ltd. All rights reserved.

### 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 metadata. The comparison of shapes is a fundamental task in archaeological research. An important example is the research area of ancient pottery, where cross-correspondences between different vessels are mostly established via their shape. However, as the amount of digitized vases is huge, the availability of an enhanced system for digital

object retrieval gains increasing importance in the archaeological community. A major challenge for a computer-aided search is that only a fraction of the excavated 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 adding missing shape information, 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, left). A rendering of the reconstructed object serves as input for a content-based retrieval, listing relevant result objects ranked by similarity (Fig. 1, middle). Based on these results, our system supports a detailed similarity assessment based on shape and/or texture by the user, which in turn allows for specializing the query by interactively defining local focus regions (Fig. 1, right).

\* Corresponding author.

E-mail address: [s.lengauer@cgv.tugraz.at](mailto:s.lengauer@cgv.tugraz.at) (S. Lengauer).



**Fig. 1.** Our proposed system. Left: The user specifies missing geometry of an incomplete 3D model by sketch and automatic texture filling. Middle: The retrieval returns similar objects in a ranked list. Right: Using dedicated visualizations, the results can be assessed and iteratively refined by setting foci.

The retrieval system presented in this paper represents an extended and revised approach of the method we previously published in [1]. Specifically, this paper extends on the previous work in the following ways:

1. Improvements in our query completion pipeline, in particular a more faithful texture inpainting method, enhancing the results obtained by our retrieval system (Section 5.2).
2. An inclusion, comparison and evaluation of an alternative feature descriptor for the similarity-based retrieval (Sections 6.2 and 7.3)
3. An updated similarity distance metric, which now allows for reflection invariant retrieval results, e.g. robustly handles cases where the handle of a vessel is depicted on the wrong side. (Sections 6 and 7.3)
4. A more salient and flexible interactive weighting method allowing for pixel-accurate importance foci (Section 6.3)
5. An extended qualitative evaluation with a more diverse search space (over 3300 vessel images as opposed to  $\sim 340$  in [1]) and an analysis of an alternative and more challenging retrieval use-case in the face of only few domain-relevant matches in this search space (Section 7.3).
6. A quantitative evaluation of the robustness of our pipeline in the face of various degrees of fragmentation and various inaccuracies of the sketch completion (Section 7.4).

In the remainder of this 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

Our work builds on methods for 3D object retrieval, as well as methods and applications in Archaeology.

### 2.1. 3D object retrieval and features

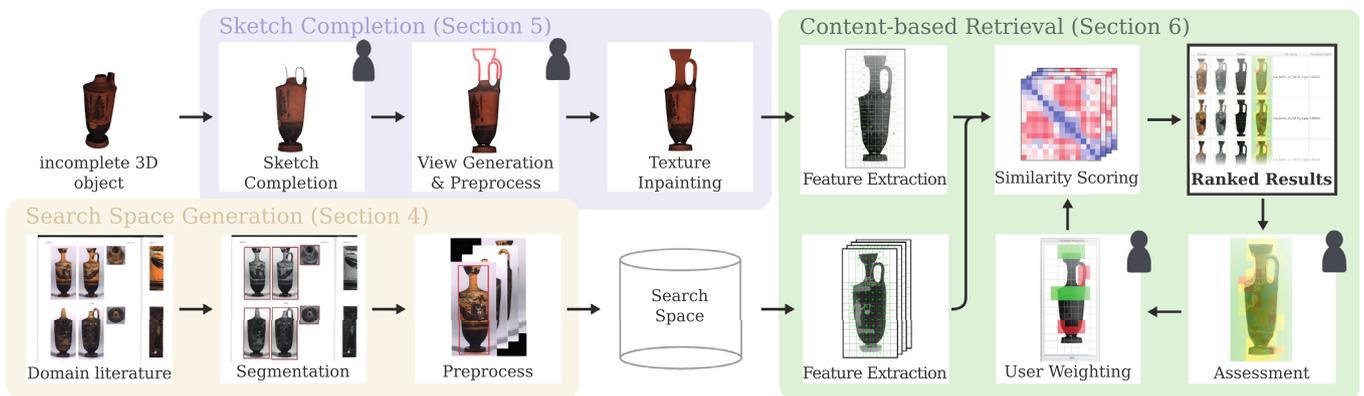
Researchers to date have investigated many approaches to 3D object retrieval, based on various similarity concepts such as shape, structure, appearance, or metadata [2]. Similarity can be based on whole objects, or on parts of objects [3]. Retrieval methods can also be distinguished by object modalities. 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 [4]. Sketching [5] 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 [6]. 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) [7], or view-based features including e.g., depth maps, silhouettes, gradients [8], or interest points [9].

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 needing a substantial amount of relevant data for training. Ioannidou et al. give a survey of methods using deep learning techniques on 3D models [10]. 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 [11] and to implement a 3D retrieval system [12]. 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 6), as preliminary results by learned features from a general-purpose network were not able to produce meaningful results for the 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



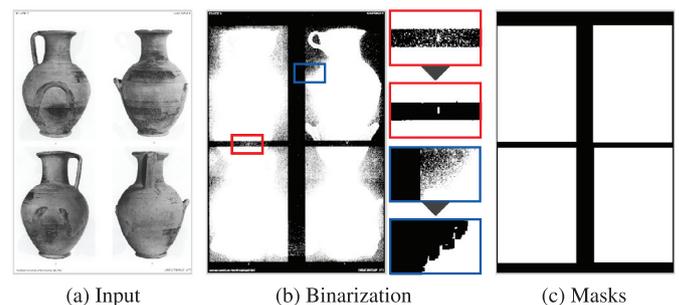
**Fig. 2.** Concept for implementing shape search using hybrid 3D+sketch queries. The sketch-completed input object is transferred to 2D modality space and compared to the 2D target images resulting in a ranked list of results. Optionally, the results can be iteratively evaluated and refined by region-based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

starting points are the series of the Corpus Vasorum Antiquorum<sup>1</sup> (CVA) and the Beazley Archive, Oxford<sup>2</sup>.

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 Pintus et al. [13]. Papaioannou et al. [14] propose a workflow for reassembly and completion of incomplete objects, using similarity in several key stages of the workflow. Pratikakis et al. [15] propose predictive scanning of CH objects. 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. Banterle et al. [16] propose a pipeline for reconstructing 3D objects from profile sections as a basis for a reference set for retrieval. Gregor et al. [17] introduce 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 of 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 steps are needed. 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 5.1). Subsequently, a rendering of the sketch-completed object serves as a basis for a texture inpainting step, which is based on previous texture synthesis techniques (Section 5.2). The result is a completed query usable for similarity search (Section 6). After the retrieval, the query can be further refined by applying selective weighting to the inpainted query image (Section 6.3). 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 given search task. 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 (Section 4.1), which typically depict multiple images together with metadata. The extracted object images also require further preprocessing steps (Section 4.2).



**Fig. 3.** A typical plate [18, pl.2] (a) is binarized by thresholding (b). Noisy artifacts are removed by morphological opening (red section in (b)) and tiny components are merged with morphological closing (blue section in (b)). Bounding rectangles around detected contours (c) are used as extraction masks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## 4. Search space generation

The generation of a comprehensive 2D search space is a major part of our pipeline. In our case the data basis for this purpose is a series of relevant archaeological volumes, present as scans of individual pages with vessel depictions, denoted *plates*. A typical plate exhibits several depictions, possibly multiple of the same object from different views or closeups, together with numberings as given by an example in Fig. 3a. The goal is to get for each plate a set of images corresponding to the depictions where each image displays exactly one object without any additional scene information. This challenge can be broken down to segmentation of a plate into separate images and the subsequent foreground-background segmentation of the object depicted on an individual image.

### 4.1. Plate segmentation

The idea behind our plate segmentation process is to bring the input into a binarized representation where connected components are corresponding to the individual images. Initially, the raw plate image, converted to grayscale, is subjected to a smart histogram adjustment which ignores 4% of the pixel with minimal and maximal values. This ensures that the pixels of the almost white plate background are located on the high end of the color histogram, permitting the application of a tight threshold for generating a binarized image. An exemplaric result is given in Fig. 3b. On the right-hand side it can be seen that in some cases image numberings or noise in the inputs account for sprays of white pixels

<sup>1</sup> [www.cvaonline.org/cva/](http://www.cvaonline.org/cva/).

<sup>2</sup> [www.beazley.ox.ac.uk/index.htm](http://www.beazley.ox.ac.uk/index.htm).



**Fig. 4.** The raw input (a) is subjected to histogram equalization (b) and thresholding (c). With morphological opening relicts of the background scenery are removed (d). The convex hull of all remaining connected components (e) is used for a graph-cut extraction (f).

between images, blending the components together. To counteract this unwanted behavior, an opening transformation with a small rectangular structuring element is applied. A subsequent closing transformation with a slightly larger kernel size reunites components which belong together while still maintaining desired gaps between components belonging to different images. Based on this result the connected components with 8-way connectivity are determined. The bounding box around the component labels are used as extraction mask after ruling out too small components whose bounding boxes take up less 0.5% of the overall plate area.

Additionally to the depictions itself, the associated image number is extracted. Together with the plate number and the book source this uniquely identifies a certain image. In most cases the image numbers are located below and either centered or on the left-hand side of their associated image. A small image segment of the expected location is used as input for the Tesseract Open Source OCR Engine.<sup>3</sup> If no string is found or the string does not fit the expected format of numerical characters, the segment is enlarged until a valid string is found or the segment exceeds a threshold size relative to the plate size.

#### 4.2. Object segmentation

The individual depictions include additional scene information, mostly in form of shadows cast by the displayed object itself, as depicted in Fig. 4a. This is unwanted due to two reasons. First, the scene information introduces additional gradients influencing gradient-based feature descriptors with unwanted information. Second, it prohibits the straightforward detection of the object contour which is the basis for contour-based retrieval approaches.

To overcome this limitation, we apply a foreground-background segmentation, based on a graph-cut extraction algorithm [19].

The idea of this approach is to treat an image as an unconnected graph  $G = (V, E)$  where the vertices  $V$  correspond to pixels and the edges  $E$  are used to model the neighborhood of pixels. The edges carry a weight which is a measure for the dissimilarity between two pixels. The aim of the segmentation is to split the graph in connected components corresponding to the foreground and background pixels while maintaining minimal cutting energy. Graph-cut approaches are typically interactive, as initial seeds are needed to initialize the individual nodes as possible foreground or possible background. Since user interaction is not feasible in our situation (large number of input images) we generate the initial seed automatically from the input.

First of all, the input is subjected to the same histogram adjustment as the plate image to enhance the contrast between object- and background pixels (see Fig. 4b). A thresholding is applied to bring the image into a binarized representation. Inputs are present in grayscale as well as in color. It has been found that for all our colored inputs the blue color channel exhibits the best contrast and has thus been used for thresholding. From the exemplary result in Fig. 4c it can be seen that this is no proper representation of the foreground. On the one hand, there are remnants of shadows in the background that introduce some faulty binarization. On the other hand, some regions on the object surface (especially such displaying motifs) are erroneously marked as background. To overcome the prior issue, a morphological closing operation with a small kernel size is applied as depicted in Fig. 4d. To address the latter, we take the convex hull of all remaining mask pixels as possible foreground (see Fig. 4e). This proved to be a sufficiently good enough representation of object shape for the graph-cut algorithm (see Fig. 4f).

### 5. Sketch completion

This part of our pipeline consists of the completion of the properly aligned 3D input by the user and a subsequent filling of newly generated regions with plausible texture.

#### 5.1. Contour completion of query object with sketch

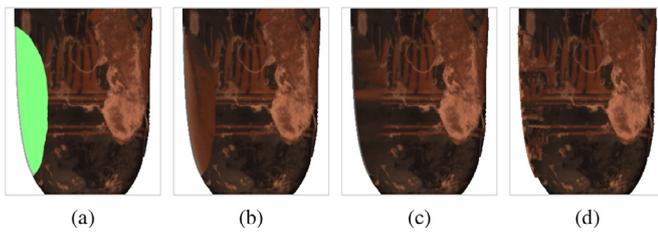
The alignment and sketching is done in a simple 3D editor with orthographic projection. From our collaborating archaeologists, we learned that the depictions in the considered domain publications come closer to an orthographic projection, than to a perspective one. Prior to the sketching process an appropriate viewpoint is determined by rotating the model accordingly. The sketch is drawn on 2-dimensional overlay superimposed to the rendered object.

To add to the usability of the sketching interface, a means to compensate for faulty user input was implemented. This includes 3 main cases: (i) gaps between the sketched lines and the remaining geometry, (ii) gaps between adjacent sketch lines and (iii) overlappings between sketch lines. These cases are handled as follows: In case of gaps between sketched lines and object geometry, the sketched line was connected to the closest point on the geometry. Gaps between sketched lines are closed by connecting the ends of the lines. Overlappings of sketched lines are removed by connecting the lines at the point of intersection and deleting parts not connected to geometry or other lines. When the completion is finished the sketch lines are drawn on a high-resolution rendering of the object. This image is then passed further down the pipeline.

#### 5.2. Texture completion

The input to the texture inpainting step is a 2D image with a sketched contour. For the content-based retrieval, the additionally created object parts need to be filled with texture in a manner that could plausibly resemble the original texture. Using simply a

<sup>3</sup> <https://github.com/tesseract-ocr/tesseract>.



**Fig. 5.** (a): Input with sketch completion defines the area for the inpainting (green). Texture synthesis results with standard texture inpainting (b), texture inpainting with user-defined directivity (c) and patch-filling (d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

single color for the whole region would introduce a gradient at the fracture line and an implausible null gradient in the filled regions. Thus, the aim is to have color and coarseness properties in those regions, similar to the properties of the original object texture.

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. The labels determine the segment which will be inpainted (target segment) as well as the segment providing texture information (source segment).

Previously, in [1], the inpainting algorithm described by Wei and Levoy [20] was combined with a descriptor loosely based on the RIFT descriptor proposed by Lazebnik and Schmid [21]. The basic idea is to compute a pixel descriptor for each pixel of the source label based on the color histogram of a pixels's neighborhood. The target label is then filled linearly from top to bottom and left to right. Same as with the source label, the pixel descriptor of the pixels in the target area is computed as well as the L2 distance to all source descriptors. From those with a distance below an empirically determined threshold, one is selected 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. As it can be seen in Fig. 5b this texture inpainting method introduces vertical directivity, and thereby unwanted gradients, to the filled regions. The desired behavior is that the inpainting originates at the fracture line and propagates towards the sketched outline. To this end we extended our system by a means that allows the user to specify the direction of the filling. The results of this approach can be seen in Fig. 5c.

Additionally, we evaluated an alternative algorithm, based on the concept proposed by Criminisi et al. [22]. The idea is to fill the target region iteratively from the region borders with patches of a fixed size. For each iteration weights are calculated along the filling front. Those are a combination of a certainty term, modelling the uncertainty with growing distance from the region border, and a data term, favoring areas with a strong gradient normal to the region border. The point with the maximal weight is taken as the origin for propagating the filling front. The surrounding of this origin is used for finding a fitting patch from the source segment, which could potentially include the whole image (except the filling region). However, for our application we restricted it to pixels containing original texture information, disregarding any background pixels. The strength of this approach is that gradient information fitting the surrounding image is induced to the target region, leading to a much more realistic texture reconstruction.

Since the patch-filling method [22] proved to produce the qualitatively best inpainting results, we use this method for all subsequent evaluations.

## 6. Content-based retrieval

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) [23, p. 30–40], Histogram of Oriented Gradients (HOG) [8] and local features like Scale Invariant Feature Transform (SIFT) [9]. 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 indiscriminating 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. Two descriptors, yielding appropriate results, are evaluated in more depth: The HOG feature descriptor as well as the shape contour descriptor (SCD) presented by Attalla and Siy [24].

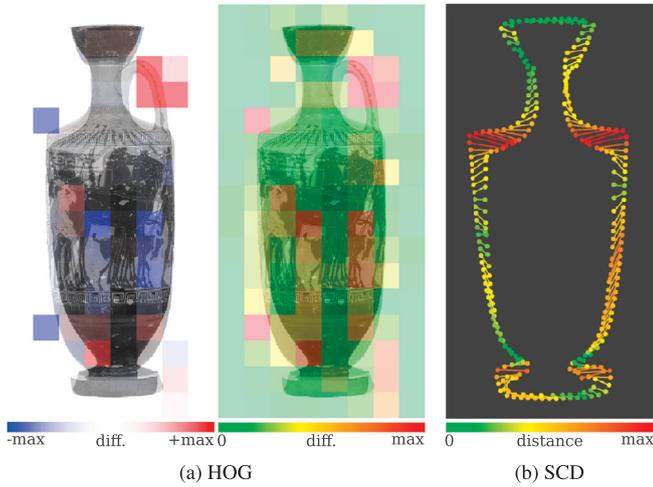
### 6.1. HOG feature descriptor

HOG describes an image by an array of local histograms, and hence can be locally weighted by a user, if required. An input image is divided into a fixed number of equal sized blocks, which are subdivided into cells. For each cell the directivity and magnitude of gradients is calculated.

This requires the inputs to be normalized in a way so that objects take up the same space in the input images. Firstly, we crop the images to the bounding box of the displayed object. Padding is added so that the objects' width and height fills up 90% of width and height of the resulting image in order to create additional context. Finally, the input is scaled to a fixed width and height as required by the HOG descriptor. To this end, additional padding is added, either to the top and bottom or left and right in order to meet the required aspect ratio while maintaining the object's position at the center. We choose an image size of 64 by 128 pixels for slender objects and 128 by 128 for bulgy ones. Moreover, for computing the distance between two feature vectors  $\vec{v}_1$   $\vec{v}_2$ , reflection invariance is obtained by taking the minimum  $d(\vec{v}_1, \vec{v}_2) = \min\{d(\vec{v}_1, \vec{v}_2), d(\vec{v}_{1_{\text{mirror}}}, \vec{v}_2)\}$ ,  $\vec{v}_{1_{\text{mirror}}}$  denoting the feature vector of the horizontally mirrored image.

### 6.2. Shape contour descriptor

The SCD is a descriptor based solely on the contour of a shape, requiring that the boundary of a depicted object is correctly extracted (see Section 4.2). The basic idea is to split the contour into  $n$  equal sized chords. In our experiments a setting of  $n = 100$  showed to be a good choice. For each chord, the descriptor stores three features: (i) the angle between the current chord and the edge from chord start point to the shape centroid, (ii) the distance between the chord starting point and the centroid and (iii) the smoothness, given by the ratio of chord length to arc length of the contour. All three features are normalized with the max values for an individual shape. Due to this normalization, the descriptor is invariant to scale, rotation and translation.



**Fig. 6.** (a): Red-green (left) and seismic (right) color maps indicating the regional differences of query and target. (b): Normalized and aligned query- and target contours with color mapping indicating local differences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 6.3. Distance metric and user weighting of target areas

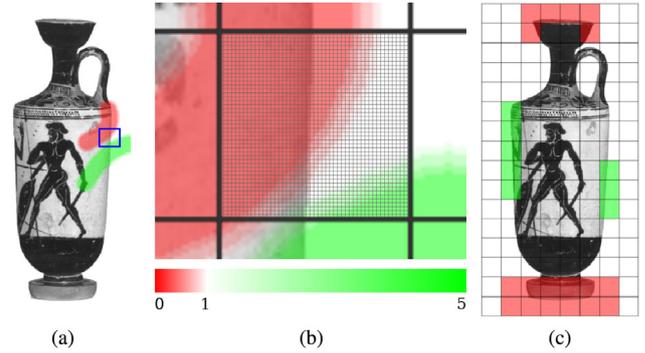
Independent from the applied descriptor, the feature vector of each image in the search space is used to compare to the feature vector obtained from the query image using the weighted Euclidean L2 single distance metric

$$d_{L2}(\vec{v}_1, \vec{v}_2) = \sqrt{\sum_i (v_{1i} - v_{2i})^2 w_i}, \quad (1)$$

with  $\vec{v}_1$  and  $\vec{v}_2$  as the feature vectors of target and query image and 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. 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 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:

**Quality Assessment.** With a dedicated interface we provide a means to the user to establish which regions of an individual target object are similar or dissimilar to the current query, and thus understand why a certain object is at its specific position in the system's ranking. Our visualization is dependent on the applied feature descriptor.

The HOG descriptor provides a vector of local image descriptors of the gridded image [8]. 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. We implemented a red-green color mapping of the absolute differences (Fig. 6a, right) and a seismic color mapping of the signed differences (Fig. 6a, left). The heat map view allows to



**Fig. 7.** (a): Freehand selection of important and negligible object areas by brush. (b) A closeup shows the weights corresponding to a hue and saturation on pixel-basis. (c): The selection can be also done via a grid surface.

effectively grasp local similarities and dissimilarities between a pair of images.

**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 or her domain knowledge, or disregarding other regions where shape similarity might be less relevant. To this end, our software allows the user to interactively specify areas of higher or lower importance with a brush tool, as shown in Fig. 7. In the retrieval this emphasis on certain regions is governed by the weighting factor  $w_i$  of the distance metric in Eq. (1). This requires a mapping of the selection on the two-dimensional image and the feature vector.

The feature vector is the flattened version of the cell-wise descriptor values allowing a direct mapping between position in the feature vector and cell index. Since the selection is freehand with a brush the weights are present not on a cell-, but a pixel-basis. The weight of a cell  $w_{cell}$  is given by the relative fraction of pixel types present in a cell with

$$w_{cell} = \frac{1}{|C|} \sum_{i \in C} w_i, \quad (2)$$

where  $C$  are all the pixel of a cell and  $w_i$  the weight of the pixel at position with an empirically chosen weights  $w = 5.0$  for 'important' pixels (green),  $w = 0.0$  for 'unimportant' pixels (red) and  $w = 1.0$  for all other pixels (default). Values in between are also possible at the edges of a brush stroke due to the linear decline of the brush kernel, visible in Fig. 7b.

With the shapelets feature descriptor the situation is easier. A descriptor value in the feature vector at position  $i$  corresponds to a specific point  $p$  in the input image. This correspondence is used to directly assign the weights to the feature descriptor.

## 7. 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 7.1), an overview of the used data experimented on (Section 7.2), and then present an encompassing use-case based evaluation, demonstrating its benefits for supporting archaeological research (Section 7.3). In Section 7.4 we present quantitative results of an unsupervised experiment with synthetically generated fractures.

### 7.1. Implementation

All steps of our pipeline can be seamlessly conducted with our purposely built software. For the sketch completion a 3D editor is included, which is capable of loading objects given as PLY-files and

supports sketching directly on top of the 3D structure. The rendered result can be exported for the subsequent texture inpainting step. Note however, that 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. Our implementation relies heavily on the OpenCV<sup>4</sup> Open Source Computer Vision library, both for tasks of image preprocessing (Section 4) and feature extraction (Section 6).

## 7.2. Data sets

Within our close cooperation with archaeologists working in the field of ancient pottery and its digitalisation, we have access to several sets of scanned objects from different museums, provided by courtesy e.g. of the Landesmuseum Kärnten (LMK), the Universalmuseum Joanneum (UMJ) or the Kunsthistorisches Museum (KHM). They are high-resolution textured 3D scans representing pottery from the Geometric period (c. 9th - 8th century BC) to the Classical period (c. 5th - 4th century BC). For our qualitative evaluation we focus on a set of Attic black-figured lekythoi from the first half of the 5th century BC., a very common vessel type in Greece for that period. Seven lekythoi from the LMK with Inv.-Nr. 1245, 1248, 1251, 1252, 1253, 9049 and 9050 (respective 3D models 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. In this way different degrees of fragmentation have been created. The subsequent reconstruction by sketch (see Section 5.1) was done by domain experts in Inkscape<sup>5</sup> according to their knowledge of Greek pottery. To expand the search to objects of different vessel shapes and painting styles we have enlarged the query objects (and also the search base, see above) by including the Attic Geometric pottery from the KHM. Corresponding 2D images were extracted from the prominent Corpus Vasorum Antiquorum (CVA) domain publication. Our search space consists in [1] (for the LMK lekythoi) of 114 images from CVA Berlin 13 [25] and 230 images from CVA Berlin 17 [26], all illustrating lekythoi, in different sub-shapes and styles (black-figure as well as red-figure). In the course of the subsequent enlargement of our search base (due to the integration of the Geometric pottery, our search space currently consists of 3340 images extracted from further 12 fascicles (see supplementary material), illustrating a wide range of different vessel types in color as well as grayscale. A small sample of the objects we automatically extracted from CVA is given in Fig. 8. All of them have been subjected to automatic extraction steps referred to in Section 4.

## 7.3. Qualitative evaluation

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), in the following referred to by KF9049, exhibiting both, a reasonable amount of missing geometry as well as a realistically placed fracture line, as shown in Fig. 9a. We compare the performance of the SCD, focusing on an exact match of the object silhouettes, as well as the HOG descriptor, comparing local gradient information of the vessel images.

**Reference Ranking.** To obtain a reference result set for our evaluation, our Archaeologists determined, from the set of all CVA objects, 30 ranked target objects (based on the incomplete 3D object) most similar by domain consideration. This ranking

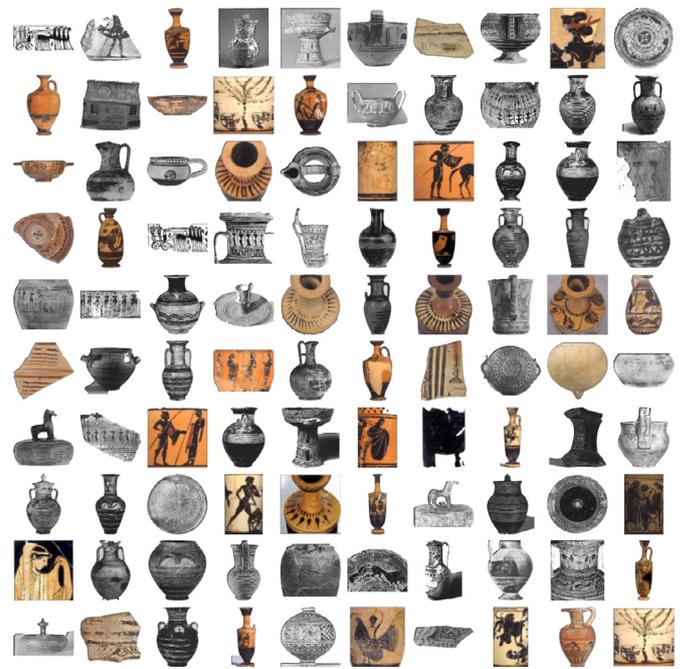


Fig. 8. A small sample of our target search space with more than 3000 images.

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. 9a displays the top 30 images of the ranked result set retrieved using the incomplete input object. The upper row shows the retrieval results using the SCD while the lower row shows the HOG-based results.

It can be seen that both descriptors are able to retrieve images visually similar to the query image, ranking similarly thin shaped vessels first. The rotation-invariance of the SCD also yields some outlier results of flat vessels exhibiting a similar silhouette after an appropriate rotation. 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 descriptors cannot compensate for missing geometry.

**Sketch-Aided Retrieval.** Fig. 9b illustrates the ranked results based on sketch-completed query objects. A visual inspection of the top 30 results shows a much stronger resemblance to the original complete object, shown in the upper left part of the image for both descriptors. An improvement of results can be clearly observed with 15 of the top 30 matches corresponding to the reference set for the HOG descriptor and 11 of 30 for the SCD.

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. The accuracy in this match is indicated by the filling level of the blue bar in the background of the vessel images in Fig. 9, with a completely filled bar indicating the exact same position. If we compare HOG towards SCD results from an archaeological point of view, we observe that HOG performs better, but this is due to the fact that our result reference set is restricted to black figure lekythoi. SCD ranked red-figure lekythoi with a similar proportion of the vessel's shape more than twice as many as HOG.

<sup>4</sup> <https://opencv.org/>.

<sup>5</sup> <https://inkscape.org/>.



**Fig. 9.** 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), the fill level of the blue background indicates the accuracy of its ranked position compared to the reference ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

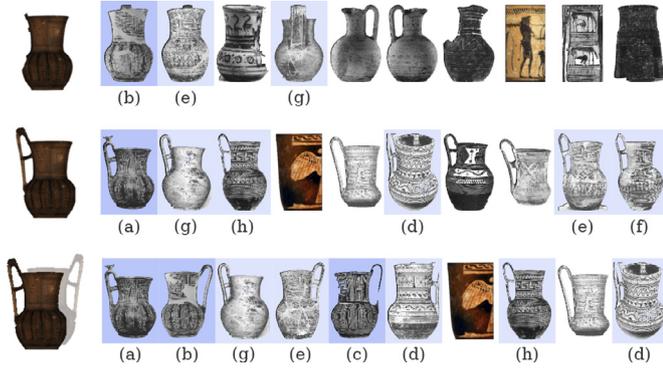
**Texture Inpainting.** After applying an additional texture inpainting step based on patch-filling (see Section 5.2), the HOG-based result can be significantly improved compared to row (b), with 22 out of 30 matches (see Fig. 9c). Note that this inpainting step does not affect the retrieval using the SCD, as this descriptor incorporates solely the silhouette information.

**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 6.3. 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. This is due to their domain experience that similarity relations of lekythoi have to be detected in vessel profiles, whereas attachments such as handles are less significant [27]. Fig. 9d shows the corresponding weighting (left) and the ranking results after incorporating these weights. We observe a further improvement of the number of matches.

**Reflection Invariance.** For scientific archaeological publications like the CVA, pottery objects are normally photographed in a right angle to the middle of the vessel's axis, showing the vessel's shape like a profile drawing. In case of vessels with handles, at least one photo records the object positioned so that the handle is either

on the right or the left side. While in the two CVA Berlin fascicles [25,26] the lekythoi are photographed with the handle on the right side in a standardized way, this is not the rule in the whole CVA and other publications. Therefore our search has to be resistant against these reflections. For evaluating the reflection invariance of our search we choose the Attic Geometric pitcher Vienna KHM IV 1, depicted in Fig. 1, middle. Same as with the lekythoi we introduce a reasonable amount of fracturing by artificially removing the handle (see Fig. 10, top row, left), resembling a realistic state of incompleteness. For the evaluation we take information from published literature in the archaeological domain. The object belongs to a specific group of pitchers with globular body with vertical gadroons and a broad concave neck, all dated to Late Geometric Ib (c. 750–740 BCE). Due to their similar vessel shape and painting decoration three further pitchers are attributed to this workshop, or more specifically to the same painter/potter [28]: Karlsruhe B 2680, Oxford 1894.13 and British Museum GR 1878,0812.8.

Even with the retrieval conducted with the incomplete object our system is able to detect one of the similar pitchers: Oxford 1894.13 (b) highlighted with a dark blue background in the top row of Fig. 10. The matches on rank 2 and 4 (highlighted in light blue) are other pitchers with similar proportion of the vessel's



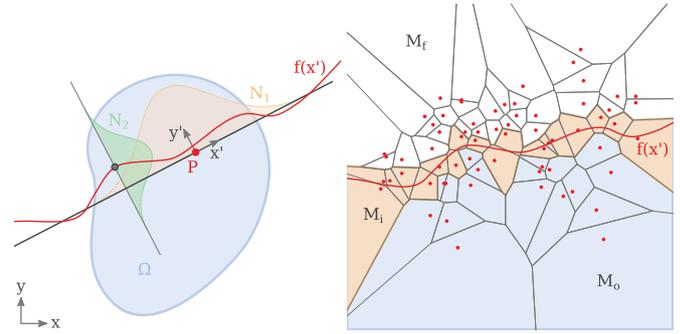
**Fig. 10.** Retrieval results for an incomplete geometric pitcher with the untreated incomplete object (top row), the sketch-completed and inpainted object (middle row) and reflection invariance (bottom row).

shape as Louvre CA 1940 (e) and Tübingen 6214 (g). When applying our sketch-aided retrieval, but without the reflection invariant distance function described in Section 6.1, Karlsruhe B 2680 (a), the best preserved, is ranked on 1 (see Fig. 10, middle row). Additional pitchers are on ranks 2, 3, 6, 9 and 10 with Tübingen 6214 (g), Berlin 31,044 (h), Würzburg H 5171 (d), Louvre CA 1940 (e) and Tübingen S./10 1088 (f). According to the main publication of Vienna KHM IV 1 [28] the Würzburg pitcher is very close to the Vienna pitcher in relation to the vessel's shape and its proportions. Oxford, which was ranked previously on 1 is not within the top 10 any longer since there are only depictions of the front/back side or with the handle on the right side. With a reflection invariant search (bottom row of Fig. 10) all of the three pitchers attributed to the same workshop, Karlsruhe B 2680 (a), Oxford 1894.13 (b) and British Museum GR 1878,0812.8 (c) are ranked within the top 10 together with 4 other pitchers already mentioned above. These are all pitchers from the Late Geometric Ib/Ila period (3rd quarter of the 8th century BC), relevant for the vessel's shape comparison but from different workshops as our Vienna KHM IV 1.

#### 7.4. Synthetic fracture experiment/ quantitative evaluation

Due to several steps of our proposed pipeline requiring user interaction, a manual evaluation with a large number of queries is not feasible. Therefore, we evaluate the sketch completion (Section 5) and the content-based retrieval (Section 6) components of our system by generating numerous synthetic queries exhibiting different degrees of fragmentation and simulating the user input for sketch completion at different levels of sketch accuracy. Two different scenarios are evaluated: the retrieval of KF9049 (see Section 7.3) and the retrieval of arbitrary vessels picked at random from our search space.

**Synthetic Fracturing.** We aim to introduce a level of fragmentation to our depicted objects, resembling the effects of fracture and decay experienced by real world incomplete objects. The basis for our experiments at this point are not 3D scans but 2-dimensional depictions extracted from domain publications (see Section 4). Consequently, physics-driven fracturing methods, which typically require 3D models, are not applicable. A common method to generate plausible fracture patterns for fragile objects is to use a Voronoi tessellation of the object surface [16]. We resort to a similar approach, by cutting through a 2-dimensional Voronoi tessellation superimposed on the image. The Voronoi seed points are placed normally distributed around a cutting curve function  $f$ , splitting the tessellation into two disjoint subsets. To generate a cut exhibiting a high amount of randomness and jitter,  $f$  is defined



**Fig. 11.** Left: The fracturing is based on a cutting function  $f(x')$ , originating at a random point  $P$  within the object pixels  $\Omega$ . Points placed with distribution  $\mathcal{N}_1$  and  $\mathcal{N}_2$  serve as seed for a Voronoi tessellation (right).

as a series of 8 superimposed sinusoidal functions

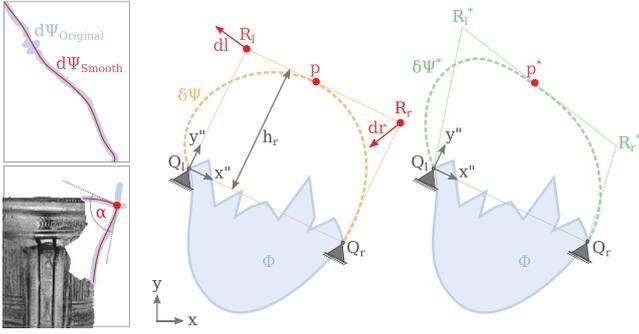
$$f(x') = \sum_{i=1}^8 m_i \sin(a_i x'), \quad (3)$$

with different random amplitudes  $m_i$  and frequencies  $a_i$ .  $x'$  and  $y'$  are the canonical basis of a coordinate system rotated by an arbitrary angle  $\gamma \in [-\pi, \pi]$  with respect to the images coordinate system. The system's origin  $P$  is placed at random within the subset  $\Omega$  of image pixels belonging to the displayed object (see Fig. 11).

A total of 100 Voronoi points  $p(x'_s, y'_s)$  are seeded around  $f(x')$  with coordinates  $x'_s \sim \mathcal{N}_x(0, 0.15h)$ ,  $y'_s \sim \mathcal{N}_y(f(x'), 0.1h)$ , ( $h$  denoting the image height), resulting in a high cell density in proximity of the cut. An example of an arbitrary fracture is displayed in Fig. 11, right. Within the local coordinate system,  $f(x')$  divides the resulting Voronoi cells into three sets: fracture cells  $M_f$  with all pixels above the curve (white), object cells  $M_o$  with all pixels below (blue), and cells intersecting the curve ( $M_i$ , orange). The final image fracturing is performed by setting all fractured object pixels  $\Psi = \Omega \cap M_f$  in the image to white color. Note that the amount of removed vessel surface (fracture rate) is limited to an interval of 5% to 95% by discarding cuts, resulting in less or more fracturing.

**Sketch-line Distortion.** In the interactive retrieval workflow, sketch lines indicating the outline of the missing parts of the fragmented objects are provided by the user. For our unsupervised evaluation system, we aim at generating queries with sketch-lines resembling user sketches with different levels of accuracy and distortion. To this end, we generate the synthetic sketch-lines based on the object part  $\Psi$  removed in the previous fracturing step. As its boundary  $\delta\Psi$  represents a 'perfect' sketch of the missing silhouette, a series of transformations is applied which aim to mimic the traits of a human-created sketch. We identified three characteristics of a user created sketch which we approximate artificially:

1. Sketch-lines are generally smooth continuous strokes without much scribble. In contrast to that, the contours of our extracted object depictions exhibit high-frequency components, either due to artifacts of our segmentation process or due to the abrasive surfaces of the artifact itself. As a countermeasure a curve smoothing based on the concept of Mokhtarian and Abbassi [29] is applied, as indicated on the left side top of Fig. 12.
2. Strong curvatures are oftentimes drawn by two lines meeting at the turning point (see Fig. 12, left bottom). We approximate said behavior by first detecting high curvature points along the contour using the algorithm by Rosenfeld and Johnston [30]. After discarding points with curvature above a threshold angle of  $\pi/3$  and a subsequent non-maxima suppression, the remaining points are used to split  $\delta\Psi$  into segments which are then smoothed independently. As a result we get a con-



**Fig. 12.** Left: Sketch lines are subjected to a smoothing (top) while strong curvatures are preserved (bottom). Middle/Right: Linear distortion of sketch-lines is governed by a control rectangle  $Q_l, Q_r, R_l$  and  $R_r$ .

tour exhibiting both, smooth curvatures but also sharp bends, resembling points where two sketch lines meet.

3. We experienced that even for experts it can be hard to accurately sketch the exact proportions of the once complete object. Let  $\Phi = \Omega \setminus \Psi$  be the set of pixels containing the remainder of the fractured object, and  $Q_l, Q_r$  its intersection points with the missing silhouette  $\delta\Psi$  (see Fig. 12). We observed that sketch distortions aggravate with increasing distance of sketch contour to  $\Phi$ , and are almost zero near the bases  $Q_l, Q_r$ . We simulate this issue by applying a random linear distortion that similarly increases with the distance to the original object part. To this end, we define a control rectangle given by the base points  $Q_l, Q_r$ , and two reference points  $R_l, R_r$  at distance  $h_r$  (Fig. 12, middle). Applying random displacement vectors  $\vec{d}_l, \vec{d}_r$  to these reference points, allows us to map the linear distortion of this control rectangle to each point  $p \in \delta\Psi$  via bilinear interpolation of the displacement vectors (see Fig. 12, right). Note that the displacement vectors at  $Q_l, Q_r$  are  $\vec{0}$ .

As a measure of the overall distortion caused by these transformations the mean Euclidean distance  $d_{dist}$  between the original contour  $\delta\Psi$  and the distorted contour  $\delta\tilde{\Psi}$  is calculated.

*Experiment Setup.* We evaluate the performance of our retrieval based on (1) the fractured image, (2) the fractured image with distorted sketch lines  $\delta\tilde{\Psi}$  and (3) the thereof reconstructed image with texture inpainting.

To determine the quality of a specific retrieval with result vector  $\vec{v}_{ret}$  with respect to a reference result vector  $\vec{v}_{ref}$  we use the normalized mean Euclidean distance of the first  $n$  elements given by

$$\bar{D}(\vec{v}_{ret}, \vec{v}_{ref}) = \frac{1}{n^2} \sum_{i=0}^n D(v_{ret_i}, \vec{v}_{ref_j}), \quad (4)$$

with

$$D(v_{ret_i}, \vec{v}_{ref_j}) = \begin{cases} |i - j| & , \text{ if } \exists j \in [0, n] : v_{ret_i} = \vec{v}_{ref_j} \\ n & , \text{ else.} \end{cases} \quad (5)$$

*KF9049 with different fractures.* In the case of KF9049 we can rely on the reference ranking introduced in the previous section with a length  $n = 30$ . 200 different fractures and sketch-line distortions have been generated. Their normalized mean distance according to Eq. (4) for the fractured, sketch-completed and inpainted queries are given in Fig. 13. Each marker encodes the orientation of the applied cut function  $f$  and the degree of fragmentation. The queries with the incomplete object perform clearly worst with the majority of the data points being close to maximal distance. Those exhibiting a lower distance also have a low fracture rate, indicating that the retrieval with the incomplete object worsens with increasing fracture rate. The figure clearly

shows that adding the sketch outlines improves the results drastically while the texture inpainting leads to an additional minor improvement. As expected, the results generally degenerate with increasing sketch-line distortion as well as with increasing fracture rate. Moreover, there seems to be no correlation noticeable between cut plane orientations and retrieval accuracies.

*Arbitrary vessels.* For the evaluation with arbitrary queries we manually selected 55 images from our search space depicting a whole vessel. For each of them about four fractures have been automatically generated on average, resulting in a total of 200 sample queries. For these cases there is no reference ranking by experts available. Instead the ranking that would result from the query with the non-fractured object is taken as reference. The mean distance of the  $n = 100$  first elements over the fracture rate is given in Fig. 14.

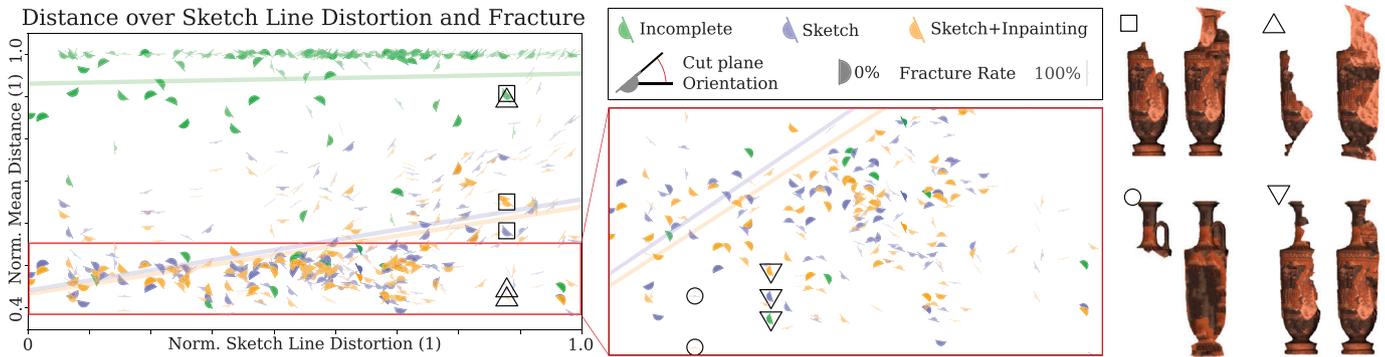
For the incomplete queries the same behavior as with KF9049 can be observed. The sketch-completed and inpainted versions perform clearly superior but in contrast to KF9049 the distribution seems to be wider spread. It is interesting to observe that when considering a diverse set of objects from the search space, both sketch-completed and inpainting-completed queries exhibit roughly similar trends as shown for KF9049 in Fig. 13. We take this as an indication of the robustness of our method.

## 8. Discussion and future work

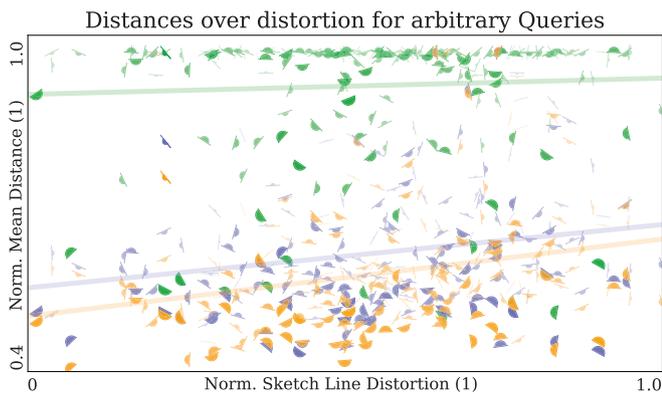
### 8.1. Applicability

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 us 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. 9. 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. Our tests also demonstrate that SCD can be quite useful in case we want to show similar vessel's shape independently of the painting style, e.g. for a diachronic analysis of a vessel's shape development in different regions of Greece. More concrete applications of SCD on archaeological pottery are the subject of future work.

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, the reconstruction of 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. 9, these features produced only between



**Fig. 13.** Left: Normalized mean distance over normalized sketch-line distortion for 200 fractures plus distortions of KF9049 with incomplete, sketch-completed and inpainted queries. The markers indicate the orientation of the cutting function  $f$  and the degree of fragmentation. Middle: Close-up on the trend of sketch-completed and inpainted queries. Right: Fragmented and inpainted version of specific queries. The results show that the inpainting-based search (orange marks and regression) performs best, producing rankings most similar to the expert-generated reference ranking. The sketch-only queries (blue marks and regression) perform slightly worse. Using only the incomplete queries (green) gives highest distances as expected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 14.** Normalized mean distance over normalized sketch-line distortion for 55 arbitrary vessels and different fractures with incomplete, sketch-completed and inpainted queries (Legend as defined in Fig. 13.)

0 and 2 out of 30 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 specifically trained networks.

### 8.2. Limitations

The presented processing pipeline is tailored based on certain assumptions, which also induce some limitations. In literature objects are usually depicted in front of a bright background. If those assumptions are not met (see Fig. 15, left) the automatic object segmentation (Section 4.2) can fail. While our pipeline is able to cope with an incomplete query object, the same does not hold for objects in our search space with missing parts like the lekythos with a missing sprout depicted in Fig. 15, right. Such an object will not be found when querying with a complete input since the overall shape differs significantly.

### 8.3. Future work

Our approach currently supports low-level editing of the query shapes. Improved editing could make use of semantic sketching techniques [31]. 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



**Fig. 15.** Left: the automatic search space generation is limited by publications not following our expected conventions. Right: for retrieval depictions of objects with missing geometry pose a problem.

in an online fashion, based on available target data, following a shadow-drawing approach [32].

We also plan to enhance the result visualization with metadata from the target repositories, like inclusion of domain texts, spatio-temporal information, and other metadata which adds context to the search. With this in mind we want to upgrade the data model of our search space from image basis to an object basis, which would, on the one hand, prevent the occurrence of the same object multiple times in a result set (see Fig. 10), and on the other hand, allow for combined queries with multiple depictions from different views. 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.

## 9. Conclusion

We have 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. The paper extends on a previously presented retrieval system, by introducing new technical findings as well as a revised and extended evaluation. Our results show that the aiding of user sketches to complete and modify the 3D query object and subsequent texture inpainting can significantly 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 itself is intuitive to be applied by any user without special skills, domain knowledge of expected object shapes will lead to a better sketch and in turn to a better retrieval result. Finally, result visualization and weight adaption support the analytical retrieval process. Future work will include other engineered and/or learned features, and extension of the sketch interface.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### 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 as well as the Kunsthistorisches Museum for providing 3D models.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cag.2020.02.001](https://doi.org/10.1016/j.cag.2020.02.001).

### References

- [1] Lengauer S, Komar A, Labrada A, Karl S, Trinkl E, Preiner R, et al. Sketch-aided retrieval of incomplete 3d cultural heritage objects. In: Biasotti S, Lavou G, Veltkamp R, editors. Eurographics Workshop on 3D Object Retrieval. The Eurographics Association; 2019. p. 17–24. doi:[10.2312/3dor.20191057](https://doi.org/10.2312/3dor.20191057). ISBN 978-3-03868-077-2
- [2] Tangelder J, Veltkamp R. A survey of content based 3D shape retrieval methods. *Multimed Tools Appl* 2008;39(3):441–71. doi:[10.1007/s11042-007-0181-0](https://doi.org/10.1007/s11042-007-0181-0).
- [3] Savelonas M, Pratikakis I, Sfikas K. An overview of partial 3D object retrieval methodologies. *Multimed Tools Appl* 2014:1–26.
- [4] Schreck T. What features can tell us about shape. *IEEE Comput Graph Appl* 2017;37(3):82–7. doi:[10.1109/MCG.2017.41](https://doi.org/10.1109/MCG.2017.41).
- [5] Li B, Lu Y, Godil A, Schreck T, Bustos B, Ferreira A, et al. A comparison of methods for sketch-based 3D shape retrieval. *Comput Vision Image Underst* 2014;119:57–80.
- [6] Lee J, Funkhouser TA. Sketch-based search and composition of 3d models. In: *Proceedings of the SBM*; 2008. p. 97–104.
- [7] 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. *J Adv Mech Des Syst Manuf* 2015;9(4). doi:[10.1299/jamdsm.2015jamdsm0049](https://doi.org/10.1299/jamdsm.2015jamdsm0049).
- [8] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proceedings of the CVPR*; 2005. p. 886–93.
- [9] Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 2004;60(2):91–110.
- [10] Ioannidou A, Chatzilari E, Nikolopoulos S, Kompatsiaris I. Deep learning advances in computer vision with 3d data: a survey. *ACM Comput Surv* 2017;50(2):20:1–20:38. doi:[10.1145/3042064](https://doi.org/10.1145/3042064).
- [11] Wang F, Kang L, Li Y. Sketch-based 3d shape retrieval using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Boston, MA, USA, June 7–12, 2015*; 2015. p. 1875–83. doi:[10.1109/CVPR.2015.7298797](https://doi.org/10.1109/CVPR.2015.7298797).
- [12] Bai S, Bai X, Zhou Z, Zhang Z, Tian Q, Latecki LJ. GIFT: towards scalable 3D shape retrieval. *IEEE Trans Multimed* 2017;19(6):1257–71. doi:[10.1109/TMM.2017.2652071](https://doi.org/10.1109/TMM.2017.2652071).
- [13] Pintus R, Pal K, Yang Y, Weyrich T, Gobbetti E, Rushmeier H. A survey of geometric analysis in cultural heritage. *Comput Graph Forum* 2016;35(1):4–31. doi:[10.1111/cgf.12668](https://doi.org/10.1111/cgf.12668).
- [14] Papaioannou G, Schreck T, Andreadis A, Mavridis P, Gregor R, Sipiran I, et al. From reassembly to object completion - a complete systems pipeline. *Comput Cult Herit* 2017;10(2):1–22.
- [15] Pratikakis I, Savelonas MA, Mavridis P, Papaioannou G, Sfikas K, Arnaoutoglou F, et al. Predictive digitisation of cultural heritage objects. *Multimed Tools Appl* 2018;77(10):12991–3021. doi:[10.1007/s11042-017-4928-y](https://doi.org/10.1007/s11042-017-4928-y).
- [16] Banterle F, Itkin B, Dellepiane M, Wolf L, Callieri M, Dershowitz N, et al. Vas-esketch: automatic 3d representation of pottery from paper catalog drawings. In: *Proceedings of the 14th IAPR international conference on document analysis and recognition (ICDAR)*, 1. IEEE; 2017. p. 683–90.
- [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*. Eurographics Association; 2017. p. 119–23.
- [18] Catling H, Mannack T. *Corpus vasorum antiquorum oxford, ashmolean museum 4, great britain 24*. London: Oxford University Press; 2010.
- [19] Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts. In: *ACM transactions on graphics (TOG)*, 23. ACM; 2004. p. 309–14.
- [20] 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: ACM Press/Addison-Wesley Publishing Co.; 2000. p. 479–88. doi:[10.1145/344779.345009](https://doi.org/10.1145/344779.345009). ISBN 1-58113-208-5
- [21] Lazebnik S, Schmid C, J P. A sparse texture representation using local affine regions. *IEEE Trans Pattern Anal Mach Intell* 2005;1265–78. doi:[10.1109/TPAMI.2005.151](https://doi.org/10.1109/TPAMI.2005.151).
- [22] Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans Image Process* 2004;13(9):1200–12.
- [23] Lux M, Marques O. Visual information retrieval using java and lire. *Synth Lect Inf Concepts Retriev Serv* 2013;5(1):1–112.
- [24] Attalla E, Siy P. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recognit* 2005;38(12):2229–41.
- [25] Zimmermann-Elseify N. *Corpus vasorum antiquorum berlin, antikensammlung 13, Deutschland 93*. München: C.H. Beck; 2013.
- [26] Zimmermann-Elseify N. *Corpus vasorum antiquorum berlin, antikensammlung 17, Deutschland 102*. München: C.H. Beck; 2018.
- [27] Haspels C. *Attic black-figured lekythoi Travaux et mémoires / École française d'Athènes 4*. "de" Boccard; 1936.
- [28] Lang-Auinger C, Karl S, Kratzmüller B. *Corpus vasorum antiquorum Wien, kunsthistorisches museum 6, Österreich 7*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften; 2019.
- [29] Mokhtarian F, Abbasi S. Shape similarity retrieval under affine transforms. *Pattern Recognit* 2002;35(1):31–41.
- [30] Rosenfeld A, Johnston E. Angle detection on digital curves. *IEEE Trans Comput* 1973;100(9):875–8.
- [31] Zeng L, Liu Y-J, Wang J, Zhang D-L, Yuen MM-F. Sketch2jewelry: semantic feature modeling for sketch-based jewelry design. *Comput Gr* 2014;38:69–77. doi:[10.1016/j.cag.2013.10.017](https://doi.org/10.1016/j.cag.2013.10.017).
- [32] Lee YJ, Zitnick CL, Cohen MF. ShadowDraw: real-time user guidance for free-hand drawing. *ACM Trans Graph* 2011;30(4):27:1–27:10. doi:[10.1145/1964921.1964922](https://doi.org/10.1145/1964921.1964922).