Top-down Analysis of Low-level Object Relatedness Leading to Semantic Understanding of Medieval Image Collections

Pradeep Yarlagadda, Antonio Monroy, Bernd Carque and Bjorn Ommer

Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany
{pyarlaga,amonroy,bcarque,bommer}@iwr.uni-heidelberg.de

ABSTRACT

The aim of image understanding, which is a long standing goal of computer vision, is to develop algorithms with which computers can advance to the semantic content of images. One ability of such algorithms would be the automatic discovery of relations between different objects in large collections of images. To analyze this relatedness we present an unsupervised and a semi-supervised approach for decomposing the large intra-class variability of object categories. The relations between objects is discovered by mapping all exemplars into a single low-dimensional projection that preserves the structure that is inherent to the category. The analysis reveals subtypes and an automatic classification algorithm is presented that predicts the artistic workshop that has drawn the objects. Finally, an approach for ordering the instances of an object category is proposed that also shows transitions between object instances. Our work is based on late medieval manuscripts from the Codices Palatini germanici.

Keywords: Shape analysis, Top-down analysis of low-level object relatedness, Visual summary of characteristic drawing styles, Computer-based discrimination of artistic workshops, Semantic understanding of medieval image collections

1. INTRODUCTION

Recent digitization projects such as that of late medieval paper manuscripts of the Codices Palatini germanici[1] have produced large collections of images. Obviously no human user can view all of these images at the same time and, thus, relations between different images or the objects within are hard to discover. Revealing the structure that is inherent to a collection of images, i.e., the artistic variations of all instances of an object category such as medieval crowns, is consequently a very difficult task. The mere size of a dataset makes it difficult to see the greater whole. Computers on the other hand can easily handle thousands of images at the same time. Therefore, we present an approach for discovering the structure inherent to a large image collection and we propose an algorithm for effectively summarizing the characteristic object relationships in a single visualization. Object relationships prove to be a useful input to researchers in the humanities for several tasks such as identifying authorship of an art manuscript based on principles of artistic design, understanding the variations of art within a particular school of design and understanding the transition of art from one school of design to another. To help the researchers address the above questions, we present a series of unsupervised and semi-supervised techniques in this paper. We demonstrate the utility of our approach in a case study on the object category of 'crowns' on 27 late medieval paper manuscripts of Upper German origin which are part of the Codices Palatini germanici digitized by Heidelberg University Library.[1]

Starting from the manuscripts, we choose an appropriate feature representation for object instances. Based on the feature representation, we generate a category level object model and utilize the model to extract all the object instances from the database. Next, we automatically extract low-level relationships among object instances based on their feature representation. These low-level relationships are the input to an unsupervised and semi-supervised method that we propose for analyzing the variability of object instances. We generate compact 2-d visual summaries with our methods, which are then utilized by art historians for their analysis.

2. RELATED WORK

Image databases in the field of cultural heritage are normally made accessible via textual annotations referring to the representational content of the images.[2] Therefore, content-based image retrieval depends on either the controlled vocabularies of the used classification systems or the textual content of free descriptions. In both cases only that can be found what has been considered in the process of manual indexing; and it can only be found in the specific form in which it has been verbalized. The inevitability of textual descriptions generates numerous problems, for example concerning the scope and detailedness of the taxonomies, their compatibility beyond linguistic,[3] professional or cultural boundaries, their focus on specific aspects of the content according to specific scientific interests or not least the qualification and training of the cataloguer. One of the most sophisticated classification systems is ICONCLASS.[4] Yet, despite its high level of differentiation it has severe limits in a global perspective because it was developed only to cover Western art and iconography. Therefore its ability to index for instance transcultural image resources such as the database of the Cluster of Excellence Asia and Europe in a Global Context at the University of Heidelberg[5] is limited. Furthermore, object definition schemes are featuring a very limited differentiation. In our showcase 'crown' the hierarchy of objects ends with this general notion and does not offer varying types of crowns. To focus the object retrieval on subtypes is, in contrast, possible in the case of REALonline, the most important image database in the field of medieval and early modern material culture.[6] Here, the controlled vocabulary contains a few compounds like 'Buegelkrone' or 'Kronhut'. But whereas the main division 'Kleidung–Amtstracht' is searchable in German and in English, these subdivisions are available only in German, thus raising difficulties of translation. Problems such as the lack of detail and connectivity are even greater in the case of heterogeneous databases, which are –like HeidICON,[7] Prometheus[8] or ARTstore[9] –generated by the input from different institutional and academic contexts. In such cases, the cataloguing of the image content is almost arbitrary due to the uncontrolled textual descriptions. Finally, a basic problem of all these databases is the fact that –due to the serious efforts of manual indexing in terms of cost and time –the fast-growing number of images that are available in a digital format can hardly be itemized in detail and thus cannot be used efficiently in the long term. To overcome these restrictions, we present a system that directly searches the visual data thereby circumventing the need for detailed textual annotations. In turn, the extracted visual data can be used to discover relationships between different object instances in the images leading to a semantic understanding of image collections.

3. APPROACH

Our work is based on a database consisting of 27 late medieval paper manuscripts from Upper German origin archived by Heidelberg University library. These codices are illustrated with more than 2,000 half or full-page tinted drawings. We chose the object category of 'crowns' because symbols of power are for their identity creating symbolism of high relevance in art history. Moreover, they are represented in significant numbers in the database and the relationships between different crown instances have a high semantic validity. Compared to standard benchmark datasets used in computer vision (e.g.[10,11]), we present a database with a high degree of background clutter, scale variation, and within-category variability. Being close to the needs in the field of cultural heritage, this image collection is highly challenging for categorization algorithms, e.g.,[10] voting methods for detection such as,[11–14] and sliding window based classifiers.[15]

We decompose the images into preparatory drawings and coloration. The preparatory drawings provide a rich source of shape information which we utilize to generate a feature representation for these crowns. These high dimensional feature vectors encode the distribution of the edge orientations of the crowns. We utilize the feature representation to generate an object model and use the model in an object detection algorithm[16] to extract all the crowns in the database.

3.1. Analyzing Intra-category variability

We capture the relationship between various object instances in the database in a single plot by embedding high dimensional Histogram of Oriented Gradient feature vectors into a low dimensional space. Such a plot makes it convenient for researchers from cultural heritage to discover relationships without having to study thousands of images. In a first step, pairwise clustering based on HoG descriptors is employed to discover the hierarchical substructure of crowns. Then we compute the pairwise distances for samples in the vicinity of the

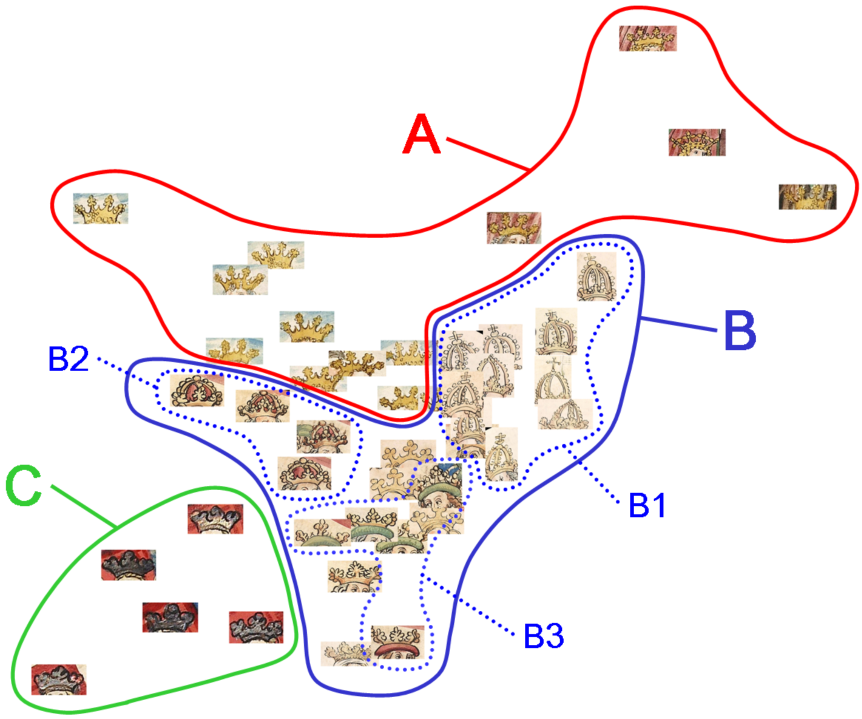


Figure 1: Visualization of Intra-category variability and substructure of crowns. Group A shows the Swabian workshop of Ludwig Hennflin. Group B shows the Hagenau workshop of Diebold Lauber with the subgroups of crowns with arches (B1), crowns with lined arches (B2) and crowns with torus-shaped brims (B3). Group C shows the Alsatian workshop of 1418.

cluster prototypes. Thereafter, a distance preserving low-dimensional embedding is computed to project the 765 dimensional feature vectors onto a 2-d subspace that is visualized in fig. 1. The embedding of the crowns in the two dimensional space is given by locations $\mathbf{x}_i \in \mathbb{R}^2$ which are computed jointly using

$$\mathbf{x}^* = argmin_{\mathbf{x}} \frac{\sum_{i \neq j} (\|\mathbf{x}_i - \mathbf{x}_j\| - d_{ij})^2}{\sum_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|^2}. \tag{1}$$

d_{ij} denote the distances between crowns i and j in the original 765 feature dimensional space.

The plot displays the following three findings: i) the different principles of artistic design ii) the high type-variability within an object category and iii) the accurate separation of different workshops. In particular, our visualization for the category ‘crown’ shows that to the simple crown circlet (A) varied elements like arches (B1), lined arches (B2), torus-shaped brims (B3), hats, or helmets are added. The low dimensional embedding of figure 1 features different principles of artistic design, which are characteristic for different workshops engaged with the illustrations. Group (B) indicates the concise and accurate style of the Hagenau workshop of Diebold Lauber,[17] group (A) the more delicate and sketchy style of the Swabian workshop of Ludwig Henfflin, and group (C) the particular summary style of the so-called ‘Alsatian Workshop of 1418’. This detection of specific drawing styles is a highly relevant starting point to differentiate large-scale datasets by workshops, single teams within a workshop, or even by individual draftsmen.

The unsupervised approach that we have presented above has helped recognizing the category substructure which has then lead to a visualization (fig. 1) of the different artistic workshops that have contributed to the Codex Palatina Germanica. Based on this visualization, art historians have provided us with groundtruth information so that we can conduct a quantitative evaluation: they have labeled the crowns in the dataset with the workshop that they come from based on formal criteria and the semantic context of the text surrounding each illustration in Codex Palatina.[17] There are 137 crowns in our dataset that belong to group A (the workshop of Ludwig Henfflin), 106 crowns belong to group B (the workshop of Diebold Lauber) and 23 crowns belong to group C (the Alsatian workshop). We then incorporate a discriminative approach for predicting the workshop that a crown belongs to. This multi-class classification problem is tackled using the features from before and incorporating a Support Vector Machine[18] in a one-versus-all manner. For evaluation, we apply 10-fold cross-validation: In each round, 50 % of the crowns from each group have been used for training and the remaining 50 % of the crowns are used for testing by holding back their labels. The classification results of the crowns according to the workshops are presented in table 1 in the form of a confusion matrix.

Workshops pred.: correct:	A	B	C
A	0.9836	0.0163	0
B	0.0365	0.9634	0
C	0.0083	0.0083	0.9833

Table 1: Classification results on the crowns from workshops corresponding to groups A, B and C in fig. 1. Columns are the predicted workshop labels and rows are the correct labels. A: Swabian workshop of Ludwig Henfflin, B: Hagenau workshop of Diebold Lauber and C: Alsatian workshop of 1418. The average classification accuracy is **97.67 ± 1.7** %.

3.2. Semisupervised Analysis of Intra-Category Variability

Figure 1 has helped the historians in visualizing the characteristics of different artistic workshops. However, the completely unsupervised mapping, defined by equation 1, from the high dimensional feature space to the 2-d space cannot preserve all the pairwise relationships between the crowns. This is an inherent limitation of any projection from higher dimensional feature space into a lower dimensional space that can be visualized. This limitation is particulary problematic for art historians when trying to infer the object relationship between crowns which belong to the same workshop, since these distances are more affected by the mapping from equation 1.

However,consider the following simple case. An arbitrary crown C has distances d_1, d_2 and d_3 from three crowns C_{R_1}, C_{R_2} and C_{R_3}. Given the distance triplet (d_1,d_2,d_3) we can assign 2-d locations to these four crowns such that the distances between C and C_{R_1}, C_{R_2} and C_{R_3} are preserved. In fact, if we fix the crowns C_{R_1}, C_{R_2} and C_{R_3} as landmark crowns with respect to which we obtain the distance triplets, we can find a 2-d configuration of crowns such that all the distance triplets are preserved. This simple but important insight leads us to a semisupervised approach where the user can choose the landmark crowns, all the other crowns are projected into 2-d space preserving the distance triplets.

We start by obtaining three landmark crowns provided as input by the user. In a first experiment, one crown from each workshop was provided as landmark. Next, we compute the distance triplets for the rest of the crowns in the database. Then we choose the location of the landmark crowns at the three corners of an equilateral triangle in 2-d space (which we refer to as 'probability simplex') such that the side of the triangle is greater than the maximum of the distance triplet values. Next, we find a mapping for each of the crowns into the interior of probability simplex such that the distance of the crown from the three corners of the equilateral triangle is proportional to its pre-computed distance triplet (d_1,d_2,d_3). Figure 2 shows the organization of crowns in a probability simplex.

Figure 2: Configuration of crowns in a probability simplex where the three landmark crowns are chosen one each from swabian, hagenau and alsatian workshops.

To visualize the substructure of crowns from the Hagenau workshop of Diebold Lauber, we choose three landmark crowns from this workshop and compute the distance triplets for all the other crowns from this workshop. The resulting probability simplex is shown in figure 3. Also, the probability simplex for crowns from the swabian workshop is shown in figure 4.

3.3. Ordering Objects

Given any two objects from a dataset the question arises how all the other objects in the database relate to these two exemplars. In particular, (i) can we find instances that help to *interpolate* between the selected reference exemplars, and (ii) can we *order* all those instances? Such an ordering is valuable for art history as it is directly visualizing relationships between the exemplars, it is illustrating smooth transitions in artistic style, and it could even reveal relationships between artists.

Given two crowns from the probability simplex, we compute the geodesic between the crowns (in this case, a straight line joining the two crowns in the probability simplex). Next, we project the rest of the crowns onto this geodesic and measure the distances between the projections onto the geodesics and the instances themselves. We retain the crowns with small distances. Then, we generate a one dimensional ordering of the crowns by showing the user selected crowns at the two ends of the geodesic and the retained crowns at the projected locations onto the geodesics.

Figure 5 shows two examples where two pairs of crowns from Hagenau workshop were provided as user input. Notice that a smooth transition can be observed in the one-dimensional ordering of crowns in both the examples.

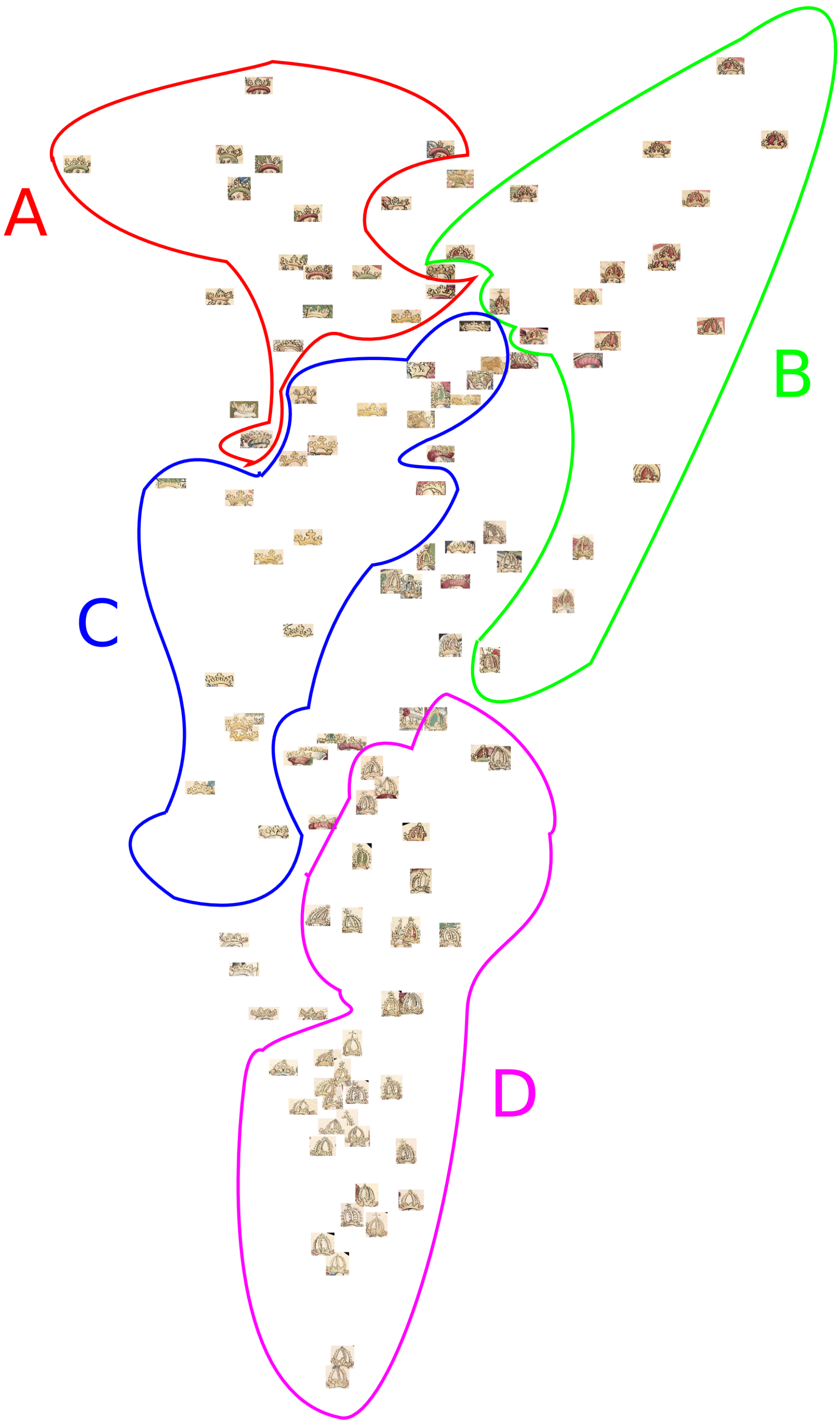


Figure 3: Configuration of crowns belonging to the hagenau workshop in a probability simplex where the probabilities are the normalized distances to three landmark crowns. Four groups consisting of, A : Torus-shaped brims, B : lined arches, C : Simple crowns circlet and D : arches have been identified in the figure.

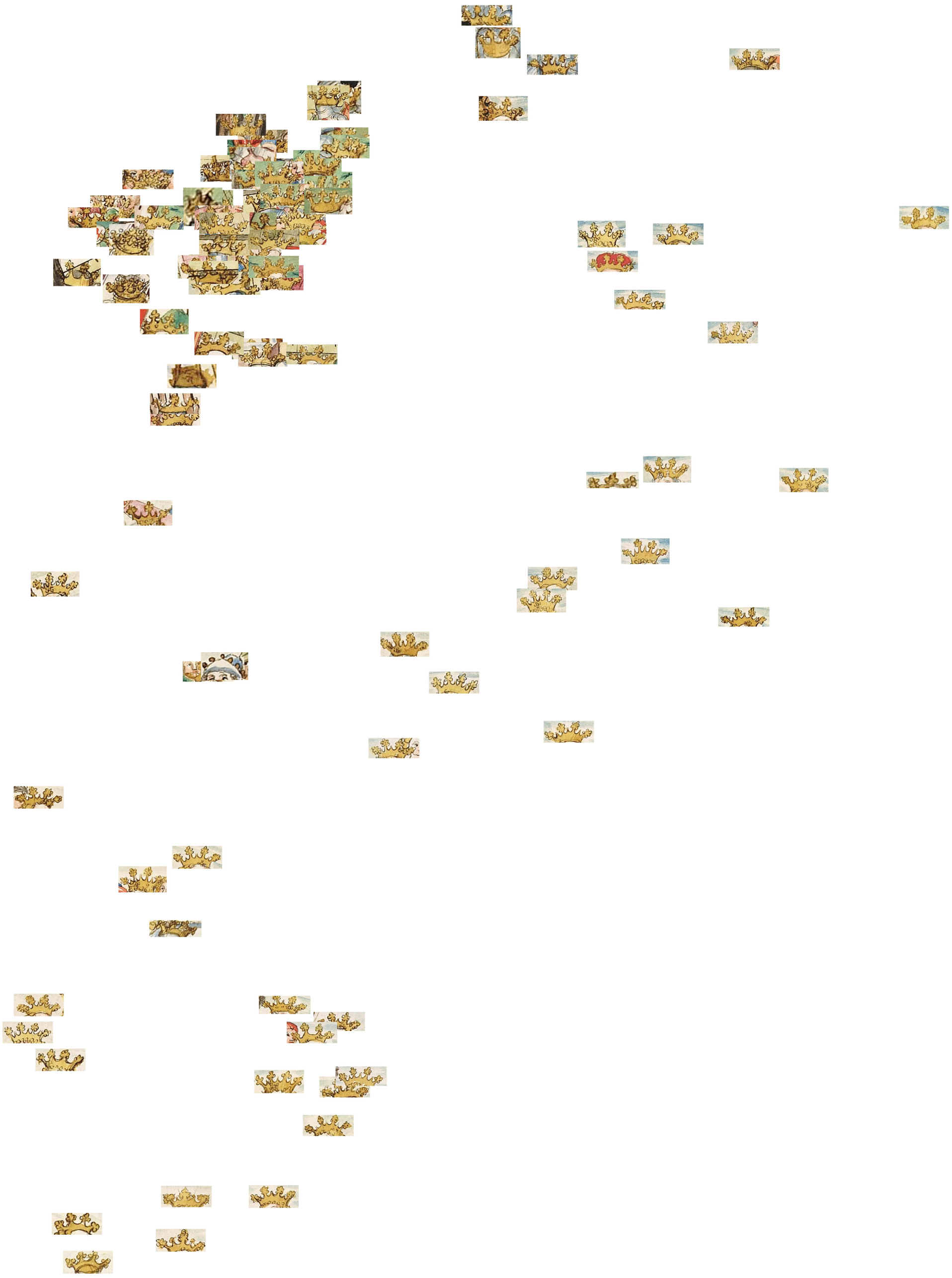
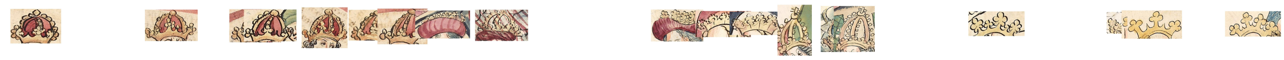
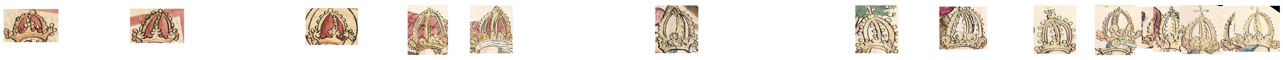


Figure 4: Configuration of crowns belonging to the Swabian workshop in a probability simplex where the three landmark crowns are all chosen from the Swabian workshop.



(a)



(b)

Figure 5: One dimensional ordering of crowns between two pairs of user chosen crowns. All the crowns belong to the Hagenau workshop.

4. DISCUSSION

In a case study on late medieval manuscripts from the Codices Palatinini Germanici we have presented an approach for analyzing the relatedness of object category instances. An unsupervised and a semi-supervised algorithm have been proposed based on a top-down object model. The approach decomposes the large intra-class variability of categories and visualizes the inherent structure of all objects in a dataset within a single 2D projection. The automatic analysis reveals subtypes based on their difference in artistic design and successfully classifies objects by the artistic workshop that has drawn them. Finally, an approach for ordering instances of an object category has been presented which also provides an illustration of the transitions in artistic style that are inherent to the image collection.

Acknowledgements. This work was supported by the Excellence Initiative of the German Federal Government, DFG project number ZUK 49/1.

REFERENCES

1. "http://www.ub.uni-heidelberg.de/englisch/helios/digi/palatina-digital.html."
2. Baca, M., Harpring, P., Lanzi, E., McRae, L., and Whiteside, A., [*Cataloging Cultural Objects. A Guide to Describing Cultural Works and Their Images*] (2006).
3. Kerscher, G., "Thesaurus-Verwendung und internationalisierung in Bilddatenbanken," *Kunstchronik* **57**, 606–608 (2008).
4. van Straten, R., [*Iconography, Indexing, ICONCLASS. A Handbook*] (1994).
5. "http://www.asia-europe.uni-heidelberg.de/research/heidelberg-research-architecture/hra-databases-1/transcultural-image-database/the-image-database."
6. "http://www.imareal.oeaw.ac.at/realonline/."
7. "http://heidicon.ub.uni-heidelberg.de."
8. "http://www.prometheus-bildarchiv.de."
9. "http://www.artstor.org."
10. Fergus, R., Perona, P., and Zisserman, A., "Object class recognition by unsupervised scale-invariant learning," in [*Intl. Conf. on Comp. Vision and Pat. Rec.*], (2003).
11. Ferrari, V., Jurie, F., and Schmid, C., "From images to shape models for object detection," *Intl. Journal of Comp. Vision* , 40–82 (2009).
12. Leibe, B., Leonardis, A., and Schiele, B., "Combined object categorization and segmentation with an implicit shape model," in [*Europ. Conf. on Comp. Vision*], (2004).
13. B.Ommer and Malik, J., "Multi-scale object detection by clustering lines," in [*Intl. Conf. on Comp. Vision and Pat. Rec.*], (2009).
14. P.Yarlagadda, Monroy, A., and Ommer, B., "Voting by grouping dependent parts," in [*Europ. Conf. on Comp. Vision*], (2010).
15. Lampert, C., Blaschko, M., and Hofmann, T., "Beyond sliding windows: Object localization by efficient subwindow search," in [*Intl. Conf. on Comp. Vision and Pat. Rec.*], (2008).

16. Felzenszwalb, P. F., Girshick, R. B., Mcallester, D., and Ramanan, D., "Object detection with discriminatively trained part based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2010).
17. Saurma-Jeltsch, L. E., [*Spätformen mittelalterlicher Buchherstellung. Bilderhandschriften aus der Werkstatt Diebold Laubers in Hagenau*] (2001). 2 vols.
18. Maji, S., Berg, A., and Malik, J., "Classification using intersection kernel support vector machines is efficient.," in [*Intl. Conf. on Comp. Vision and Pat. Rec.*], (2008).