Which Entry is More Similar? A Non-linear Visualisation of Query Results in Image Retrieval and Image Recognition Problem

Joanna Gancarczyk
University of Bielsko-Biała
BIelsko-Biała, Poland
jgan@ath.bielsko.pl

Anna Olszewska
AGH University of Science and Technology
Kraków, Poland
aolsz@agh.edu.pl

Content based image retrieval (CBIR) has been a subject of exploration in digital humanities since 1990's (Gudivada 1995). Various descriptors were implemented to represent shape, texture and colour content of the image as sequences of numerical values (Zhang & Lu 2004, Veltkamp, Latecki 2006, Zha & Yang 2010). At the same time similarity measures and learning algorithms were designed to enable efficient image classification and retrieval (LeCun 1998). The issue, however, remains in a simple question: which descriptor and which similarity measure best reflects the human perception of similarity of visual objects? And is this the same one, that best responds to the ground truth in a retrieval query?

In this paper, we move for a while away from the very technical issues of shape descriptors definition and verification and we focus on the question how visualisation of the computed data affects the final result of a visual query. We are replacing a traditional, linear presentation of *n* most similar outputs to a set of graph-like and scatterplot based visualisation modes. The research study is performed on a particular example of visual search in large databases of historical watermarks, trademarks and monograms.

We believe, that the approach to search across a digital print room repository involving intuitive user interaction is a step toward fully making use of its potential. We state, that a modern interface, that allows the end user an intuitive navigation through options and partial results is a milestone on the way to fill a technological gap between users familiar with image processing issues and with computer science background and those for whom obtaining an answer for a particular research question is worth more, than understanding how the result was actually computed. Finally, we proof the concept with a proposition of a shape descriptor followed by a set of flexible interfaces designed to display and navigate through the results of a visual query.

CBIR. Data visualisation. Digital print room exploration. Shape descriptors. Pattern similarity measure

1. INTRODUCTION

Historical watermarks, artist's monograms as well as trademarks and collector marks have been a subject of research since the 19th century. Printed catalogues of the collections were published by C. M. Briquet, G. K. Nagler, and G. Piccard. Till now, several attempts have been made to move the content to the digital space, thus making it searchable in more convenient way. However, none of the applications has fully satisfied requirements of robustness, flexibility and intuitive handling. In our paper we propose a set of paths to be followed when handling with a multidimensionality of the feature vector describing the shape of a watermark, displaying results of image retrieval according to a

complex similarity measure and visualising content based connections between images in a given dataset. Our test dataset consists of 250 selected images of historical watermarks, trademarks and collector marks, as described in subsection 1.3.

1.1 Historical Trademarks and Watermarks

Signatures, monograms, trade- and watermarks, used by artists and craftsmen, form a group of "analogue metadata" indispensable for the history, economy and visual studies. Those collections are usually organised according to the traditional areas of art and humanities studies (i.e. art history, paper documents expertise, museology). The curatorship strategies dedicated to the image analysis and

CBIR may enable combining archives and opening new areas of the visual studies. Sets of images accompanied by the high-quality catalogue descriptions are ready to use test groups convenient for the evaluation of visual data curatorship strategies.

Since 19th century numerous documentation had been produced to enable examination of this images (C.M. Briquet, G.K. Nagler, G. Piccard). Until 2016 over 500 000 of items have been translated on-line, forming a cloud of heterogeneous digital repositories. Five volumes printed edition of Georg Kaspar Nagler's 'Die Monogrammisten', brings over 15 thousand of painters' and craftsmen' signatures and monograms, dated up to 1800. On-line extended version of Fritz Lugt's catalogue contains nearly 4700 examples of prints and drawings of the collectors' marks. The largest part of trademark documentation available is dedicated to paper documents (printed images, books and manuscripts), created before 1750. Those repositories report watermarks ('filigranes') and paper structure images published mainly by members of the Bernstein Consortium.
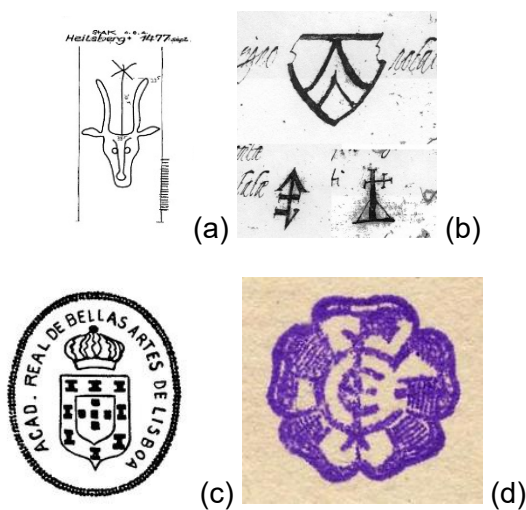

Figure 1: *A watermark displayed in the G.K. Nagler catalogue (a), a trademark displayed in the G. Piccard (b) and collector signs displayed in the F. Lugt repository (c, d).*

Every catalogue brings a set of two-dimensional semi abstract linear shapes, combined with various forms of letters. Most of them are repetitive, mechanically produced: stamped or printed on the surface of the object. In consequence, semantic based search methods of these documents is not necessarily the most effective one. In the former CBIR research projects (Rauber 1997, Eakins 2001) selected groups of this images were tested against various similarity measures and visual descriptors. A ranking of the outcomes was rather simple and unified.

1.2 Human Perception vs. Data Visualisation

There exist significant field of research on the junction of physiology and design studies (Gombrich 1979) or fine arts and mathematics (Bürgisser & Cucker 2013) concluding some fundamental studies and giving hints on visual structures and representations. According to the mentioned works, centrally composed schema follows the basic strategies of pattern filling and makes use of the break-spotting habits observed. We draw on those basic states in our research and propose centralised composition of the display as one of the visualisation tools.

1.3 The Dataset

Our preliminary test set consists of 250 images. In its arrangement, we have tried to resemble variety of the contemporary research repositories containing historical watermarks (150 images) and collectors' marks (70 examples). For this reason, the dataset includes digitalised drawings as well as photos and backlight or x-ray images. The shapes chosen can be divided into 20 unique examples, 10 classes containing 15–30 variants each, 15 smaller groups of 'doubles', acquired with the different imagining techniques i.e. drawing and photography. Division into classes is based on the classification recognised in the Piccard catalogue. Variants of the widely dispersed motives (i.e. watermarks called bull's head, gothic letter "P", lily and the tower) were included. The above-mentioned groups are well known for their variations, that are often hard to be traced.
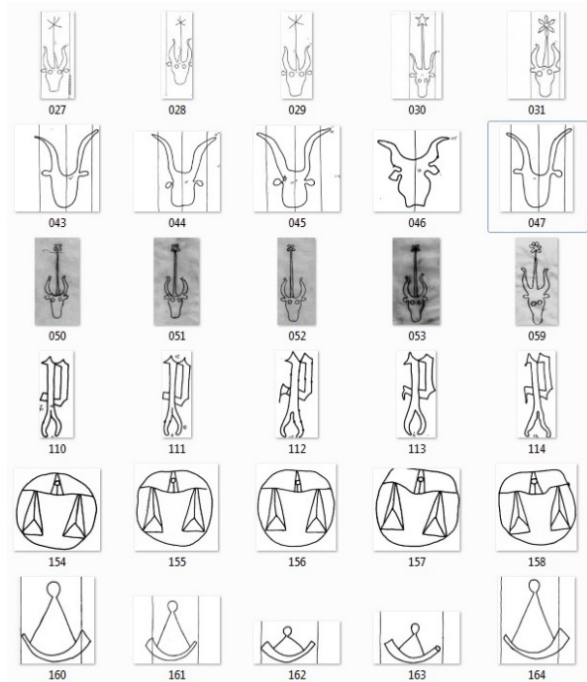

Figure 2: *Examples of test images (primary source in repositories of Frits Lugt Collectors marks, Piccard Online and Bernstein Memory of Paper).*

In Figure 2, some exemplary watermarks from our test database are shown. The original images all derive from open access:

- Frits Lugt *Les Marques de Collections de Dessins & d'Estampes* (http://www.marquesdecollections.fr, Fondation Custodia)
- *Piccard Online* catalogue (https://www.piccard-online.de)
- Bernstein *The Memory of Paper* (http://www.memoryofpaper.eu/BernsteinPortal)

2. THE PROBLEM OF MULTIDIMENSIONALITY

A typical, automatic description of the image content is defined by a *feature vector*. The feature vector contains numerical values of attributes computed on the image by an algorithm on basis of objects size, proportion, feature distribution, colour, brightness – to mention just a few of most frequent descriptors. It is an usual situation, that the feature vector is of one hundred, or often more elements and there is no possibility to plot the values of vector attributes as a scatterplot of n=100 dimensions. Thus, the need of alternative visualisation appears crucial.

In the following chapter, we are going to give a short description of the algorithm used to define the feature vector attributes. Then we discuss the result of a retrieval query and its representation in selected visualisation models.

2.1 Radial Distribution Histogram

Conceptually the proposed shape descriptor is based on the *radial distribution function*, which in statistical mechanics represents the probability of finding an atom in a shell at the distance r of another atom chosen as a reference point.

Let c be the centre of mass of the given figure F and d the most distant point from c, belonging to the figure. Then, we divide the distance r between c and d into n equal parts and construct n circles around c, each of the radius $k*(d /n)$, $k = 1,...,n$. For each circle C_k we compute an intersection with the figure as a sum of all pixels p, such that $p \in F \cap C_k$. A histogram of number of pixels on each circle against the radius r_k, $k = 1,...,n$ defines the shape descriptor discussed in this paper. Figure 3 shows two examples of analysed shapes with superimposed ten concentric circles determined by the centre of mass. The radial distribution is robust to rotation and thus the shape matching is direct and doesn't need additional time consuming operations.
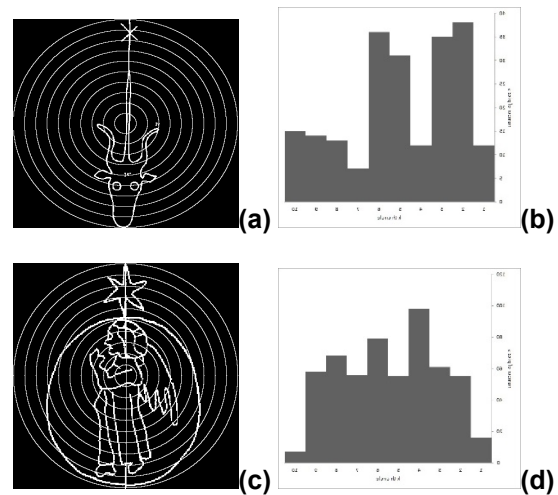

Figure 3: A bull and an angel shaped watermarks analysed by means of the radial distribution algorithm: analysed shapes with superimposed concentric circles (a, c) and histograms achieved by means of the radial distribution function (b, d).

2.2 Dimensionality reduction

Despite the fact that the images forming the test set of our research are described by complete metadata including IPH code (because they derive from curated, well known collections), we do not take the descriptive code into consideration, as the goal was set up as purely visual retrieval and recognition system. But even though all the data concerning image name, provenience, unified category name, dates, etc., was truncated, its dimensionality exceeds possibility to visualise it in a convenient to human perception way. The size of the feature vector, as introduced in subsection 2.1, is 100 attributes and this is not an exception in the task of content based description.

There are several ways to visually represent this kind of complex data by means on human interaction (Ferreira de Oliveira & Levkowitz 2003). On the other hand, algorithms aiming in a reduction of spare attributes let represent the data as a scatterplot in only two or three dimensions.

According to the algorithmic definition, in machine learning and statistics, *dimensionality reduction* (or dimension reduction) methods convert the high-dimensional data set $X = \{x_1, x_2,..., x_n\}$ into two or three-dimensional data $Y = \{y_1, y_2,..., y_m\}$ The aim of dimensionality reduction is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map (Van der Maaten & Hinton 2008).

2.3 t-SNE visualisation

t-Distributed Stochastic Neighbour Embedding (t-SNE) is a technique for dimensionality reduction that is particularly well suited for the visualisation of

high-dimensional datasets. The technique can be implemented via Barnes-Hut approximations, allowing it to be applied on large real-world datasets (Van der Maaten & Hinton 2008).

We perform the visualisation of 250 feature vectors, each containing 100 values referring to the shape representation based on the radial distribution histogram by means of the t-SNE dimensionality reduction algorithm, as implemented in the tSNEJS demonstration by Andrej Karpathy at Stanford (see http://cs.stanford.edu/people/karpathy/tsnejs/).

In Figure 4, each of the images is denoted with a blue dot and the image name. Visually similar images tend to form a "cloud" on the map. The more isolated and compact the cloud is, the higher probability that images belong to the same visual class might be expected. Here we highlight the five-element group of images forming a cloud to the bottom-right from the main concentration area. They all belong to a "scales" class of watermark images. There appears also one "scales" class image away from the centre of its class. In the general case, this might happen in two situations: either the shape descriptor doesn't represent the shape correctly, or, which would be much more interesting, the initial class description of the image (given as a metadata) was 'wrong'. This simple T-SNE's example reveals complexity of the visualisation problems.
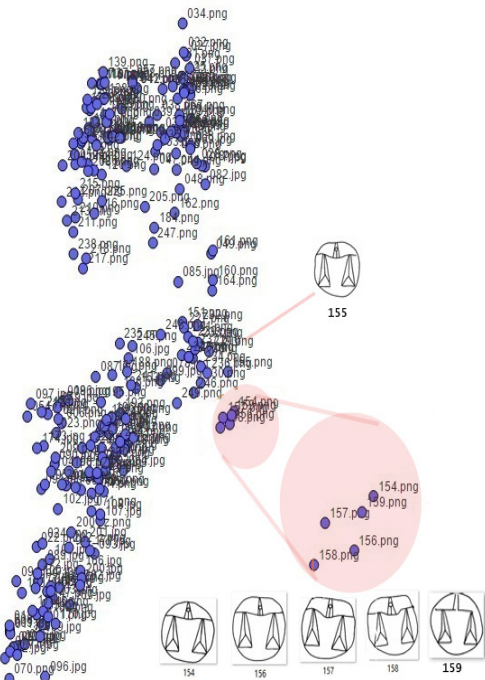

Figure 4: *t-SNE visualisation of the dataset. Each of the images is denoted with a blue dot and the image name. Visually similar images tend to form a "cloud" on the map, and the more isolated and compact the cloud is, the higher probability that images belong to the same visual class might be expected.*

3. VISUAL SEARCH RESULTS REPRESENTATION

The primary goal stated in our project is to define an automated or semi-automated methodology for visual retrieval in large datasets of historical watermarks, trademarks and monograms. Given one query image, we aim to display a collection of most similar images retrieved from the dataset, according to a chosen similarity measure.

3.1 Linear Representation

The linear representation is a "classical" way of representing an ordered set of objects of any kind. If the ordering algorithm is based on one criterion, there is not much to be improved in the scope of its visualisation. Figure 5 shows a visualisation of the result of a retrieval query given by an image of gothic "P". The result is obtained by means of algorithm based on the radial distribution histogram.
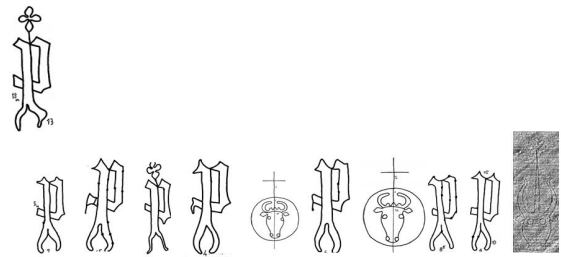

Figure 5*: Linear display of result images in a visual query given by a gothic "P" image.*

The display order is from left to the right side of a screen. As might be observed, most of the retrieved images belong to the expected class, there are however three among ten objects, which were wrongly addressed (the search algorithm tends to return the closest result with high accuracy, the successive results vary).

Misinterpretation of an image in a retrieval process is a complex issue, particularly if the similarity search is based on a weighted feature vector of various image descriptors. The retrieved results may form in that case a sequence of visually inconsistent images, the more diversified the descriptors used to form the feature vector, the more amorphous the resulting set may be.

In the following section, we propose two alternative methods of displaying a result of content based image retrieval. Both are dedicated to (a) involve human interaction in the retrieval process (particularly in the case of diversified descriptors forming the feature vector) and (b) verify the results in a simple, visual way.

3.2 Multidirectional Representation

Non-linear representation of image retrieval outcomes enables parallel presentation of several methods of image comparison, leaving more space for the users to evaluate their accuracy, according to the specific query. The proposed outlay is inspired by Gombrich and Cucker research upon physiology-design and fine arts-mathematics subject, as mentioned in subsection 1.2.

Following observations on pattern-filling strategies of ornamental decoration (Gombrich 1979) we may assume that a legible schema should be a combination of 'central' – larger-scale elements interlaced with more 'scattered' - outlying examples. In this schema, the query image appears in the centre of the page. Each cluster represents separate ranking covering equal space encircling the query example. Images representing outcomes are displayed in various scales, depending on the place they take in the respective ranks. Those that are listed on the top of the lists appear to be larger. They are placed closer to the central image. Those corresponding to a lower position are reduced.

Image size reduction leaves the space for blank background around the images. In consequence,

the outer region of the cluster is interpreted as more loose, scattered structure that tend to be perceived as „secondary" or „distant" region. Change of the scale is being defined as consequent reduction of the image areas. Its value is measured for each cluster, depending on the number of pictures it contains (i.e., 0.2 for the list of 30 images). A "Follow the rank" function enables highlighting the images that correspond to a specific position in a rank. It enables tracking and comparing hierarchy of each cluster.

Figure 6 shows a two-directional distribution of the result of the gothic "P" watermark retrieval. The left and the right side of the diagram corresspond to two different similarity measures applied in the search algorithm.

The goal of this particular representation is to let the user choose an adequate similarity measure by pointing the area, where images that best fit their expectations are distributed. Then the search may be repeated according to the chosen method, intentionally leading to a refinement of the search result.
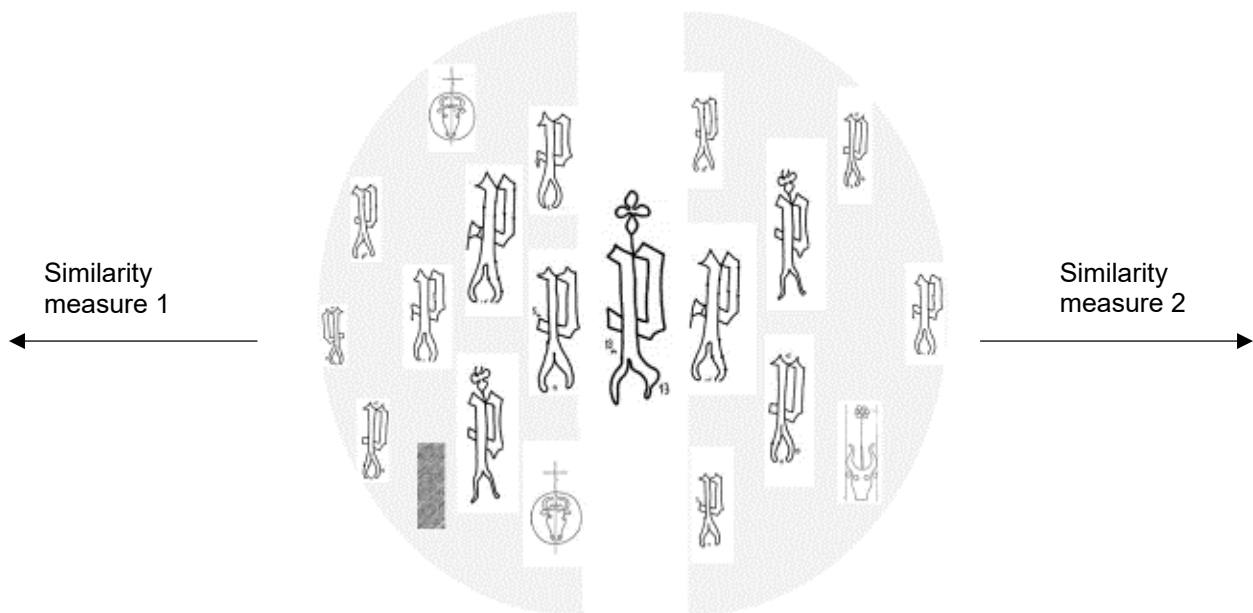

Figure 6: *Non-linear representation of image retrieval outcomes enables parallel representation of several methods of image comparison, leaving more space for the users to evaluate their accuracy. The query image is in the middle.*

3.3 Connectivity Diagram

Another visualisation was designed to perform the evaluation of the visual search results on a labelled training dataset. All images in the dataset are described by their class and name. The image names are distributed evenly around a circle and a

chord is drawn between each image and its five responding images retrieved by a visual query. For clarity, all images and chords in Figure 7 related to the gothic "P" class are indicated by a red colour. As might be observed, most of the connections were retrieved among this one class, there are only few pointing to an image from outside the class. Figure 8 highlights five resulting images

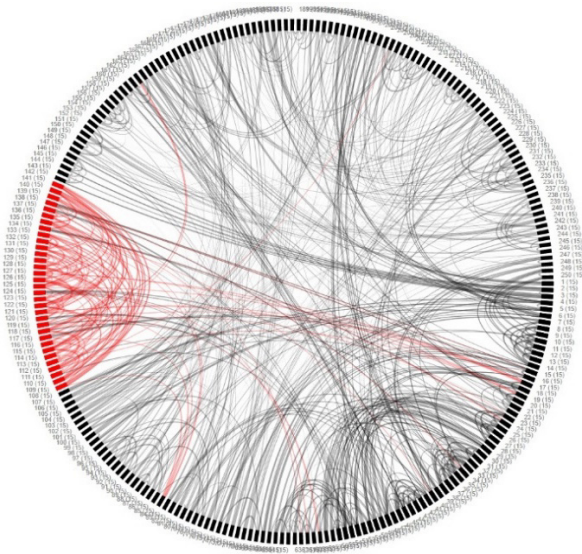corresponding to the given one. All of them belong to the same gothic P class.

Figure 7: *A chord diagram describing relationship among images in a test dataset. Each of the images, names of which are distributed evenly around the circle, is joined with five corresponding images according to a chosen similarity measure.*

The visualisation was performed according to the similarity measure defined by the radial distribution histogram, with use of a java script library (https://d3js.org).

4. CONCLUSIONS

In this paper, we focus on the question: How visualisation in the retrieval process affects the final result of a visual query. The obtained results show, that visualisation mode should not only be considered as a layout for the end user, who's interest is mainly in retrieving a correct answer to their query, but also a verification tool for shape descriptors and distance measures applied in the search algorithm. We propose a series of interfaces based on dimensionality reduction, multi-directional representation and connectivity diagram to enable subdivision of a given dataset, visualisation of the relationship between images and human-system interaction. All the results are obtained on a sample dataset of 250 images of historical watermarks and collector marks, described by a feature vector based on radial distribution histogram and labelled for a verification by a class name.
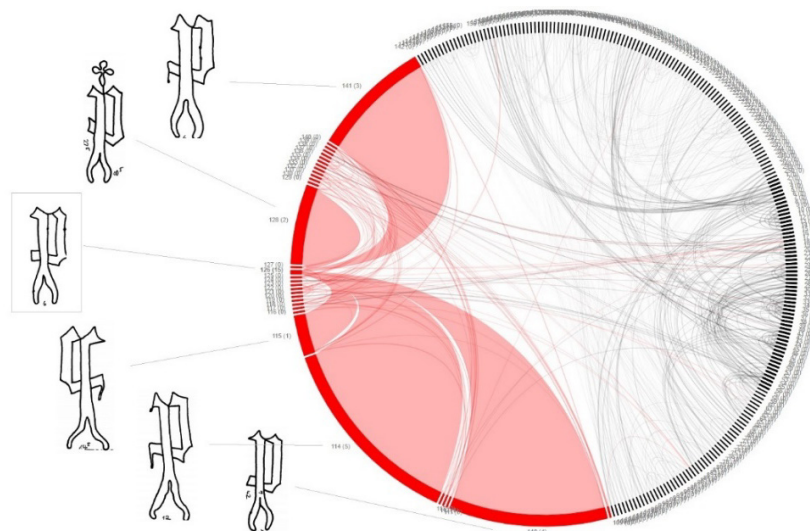

Figure 8: *A chord diagram of dependencies among a given dataset of images. A sample image query (in a frame) and the results of similar shapes retrieval are indicated in the diagram.*

5. REFERENCES

Briquet, C. M. (1923) Les filigranes dictionnaire historique des marques du papier des leurs apparition vers 1282 jusqu'en 1600.

Bürgisser, P. and Cucker, F. (2013) *Condition: The geometry of numerical algorithms*, Vol. 349. Springer Science & Business Media.

Eakins, J. P., Brown, J. E., Riley, J., and Mulholland, R. (2001) Evaluating a shape retrieval system for watermark images. In Bendkowska, A. (ed.), *Digital Art History – A Subject in Transition:*

Opportunities and Problems, CHArt Conference Proceedings.

Ferreira de Oliveira, M. C. and Levkowitz, H. (2003) From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 9(3), pp. 378–394, July.

Gombrich, E. H. (1979) *The Sense of Order: A study in the psychology of decorative art*. Cornell University Press.

Gudivada, V. N. and Raghavan, V. V. (1995) Content based image retrieval systems. *Computer*, 28(9), pp. 18–22.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*.

Lugt, F. (n.d.) Les marques de collections de dessins & d'estampes. The new edition on-line http://www.fondationcustodia.fr/english/merkenboek/default.cfm?gr=menu2 (accessed 24 February 2017).

Nagler, G. K. (1860) *Die Monogrammisten und diejenigen bekannten und und unberkannten Künstler aller Schulen*. G. Hirth´s Verlag.

Rauber, C., Pun, T., and Tschudin, P. F. (1997) Retrieval of images from a library of water-marks for ancient paper identification. *Proc. Second ACM International Conference on Digital Libraries*, ACM, New York, USA, pp.123–130.

Van der Maaten, L. J. P. and Hinton, G. E. (2008) Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9, pp. 2579–2260, November.

Veltkamp, R. C. and Latecki, L. J. (2006) Properties and performance of shape similarity measures. In *Data Science and Classification*, pp. 47–56. Springer Berlin Heidelberg.

Zha, Z. J., Yang, L., Mei, T., Wang, M., Wang, Z., Chua, T. S., and Hua, X. S. (2010) Visual query suggestion: Towards capturing user intent in internet image search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 6(3), p. 13.

Zhang, D. and Lu, G. (2004) Review of shape representation and description techniques. *Pattern Recognition*, 37(1), pp. 1–19.