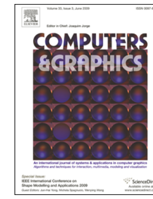
Computers & Graphics ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag



Special Section on SIGRAD 2015

Sifting through visual arts collections Q2

H. Pflüger, T. Ertl Q1

Institute for Visualization and Interactive Systems (VIS), University of Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany

ARTICLE INFO

Article history:
Received 1 October 2015
Received in revised form
21 March 2016
Accepted 24 March 2016

Keywords:
Visual arts collections
Visualization
Projection methods
Clustering
Similarity
Comparison features
Point cloud

ABSTRACT

We introduce a visualization system for large image sets which combines a distance function, a clustering and a projection method. The distance function, the clustering and the projection methods run so fast that they can calculate new results during the interaction with the user and can therefore be adapted dynamically to the context of the investigation and the requests made by the user at any given moment. The system aims to facilitate investigations which take similarity between images in terms of human perception into account. Similarity in terms of human perception is highly context and task dependent and cannot be described by a metric in the mathematical sense. Functions reflecting similarity in terms of Q4 human perception have to be adapted dynamically to the context of the investigation as well as to the tasks assigned at any given time. Our system thus shows that these requirements can be met in principle, Q5 and we propose it as a basis for developing specific applications and suitable surfaces in collaboration with experts for whom such tools are useful, as for instance experts of art theory. Q3

1. Introduction

The number of digital images people can access is enormous and continually increasing. Specialized Web portals such as Google or Flickr are offering billions of digital images. While limiting ourselves to works of visual arts reduces these myriads of images significantly, they still number in the range of several million. The image archive Prometheus (www.prometheus-bildarchiv.de), for instance, has more than one million digital images on display. It is obvious that on this kind of scale, no one human being is able to sift manually through such a number of images. These bodies of images either contain great redundancies, or one has to accept that one can only survey small selections of these bodies. In either case, one is left with the question of how to get an overview of the entire body of images and how to make that body tangible.

The theory of art provides ways of structuring quantities of images. For example, different styles help to classify bodies of images, both in terms of appearance and content. Additionally, image details such as the names of the artists, the images' titles, their years of origin, painting techniques, and sizes, help in the process of categorizing large bodies of images. Such annotations can help making structures and correlations accessible to users and may enable them to find samples suitable for their purposes. Formal metadata of the images (artists' names, titles of images, years of origin, painting techniques, and sizes) usually accompany the works' digital copies, and assignations of the styles and characteristics of the images are available for the most important and best known works of art. However, assigning works of art and choosing representative artworks require expert analysis and are therefore costly; and they fall short when one is dealing with the ever increasing total number of works of art. In addition, such assignations are most often ambiguous and depend on the issue at hand as well as on the context.

A system to analyze and explore large sets of images of visual arts should have the following properties.

- Similarity between images in terms of human perception obviously constitutes an important relationship between images, and we believe that it is essential in the study of works of visual art to consider this property. The computation of visual similarity is usually calculated by means of complex functions (Section 3). Therefore, the image distances for clustering or visualization have to be calculated via a distance function and not via a metric of a feature space (Section 2); in the process, metadata or features retrieved from image annotations or the image itself can be integrated into the distance function.
- There is no way a human being can examine all the individual elements in a large set of images. For this reason, we do not use individual images as objects of an investigation, but rather groups of images. The optimal number of images per group depends on the images which are clustered and on the specific conditions during the investigation. Large clusters may be taken if there are many similar images in the set, or if the task is to get an overview of the total amount. On the other hand, it is necessary to form small groups or even take single images when

E-mail addresses: Hermann.Pflueger@vis.uni-stuttgart.de (H. Pflüger), Thomas.Ertl@vis.uni-stuttgart.de (T. Ertl).

http://dx.doi.org/10.1016/j.cag.2016.03.009
0097-8493/

individual subsets are to be examined in more detail. Therefore, the clustering should be done dynamically.

- Similarity between images depends on the given context (e.g., paintings, drawings, photos, color or black and white images, figurative or abstract images), the task (e.g., getting an overview, exploring structures or relationships, or classifying tasks), and the current state of an investigation. The distance function and the visualization in general should therefore be adaptable during the investigation for any requirements that may arise at any given time of an investigation.
- The system should be a human-in-the-loop system and therefore follow the intuition, tasks, requirements, and precognition of the user, ensuring that that person be given incremental insight into a collection.

The validity of distance functions simulating human perception is weak. Their dependence on the given context and tasks is too high to transfer them into a metric in the mathematical sense. In our opinion, tools for structuring works of art need to be supported by experienced users; the structuring process can only be done interactively with a human being involved in the system, which leads to the above listed conditions. To our knowledge, the usual methods are too sophisticated and costly for such tools. In our work, we present methods that meet the conditions described above: a distance function similar to David Lowe's algorithm SIFT [1], optimized for visual art (Section 3); a simplified k-means++ clustering (Section 4); and a 2D visualization reflecting similarity between images/clusters (Section 5). In Section 6 we show an interface that takes advantage of these methods and enables users to sift through large image sets and visualize structures given by metadata (Fig. 1). We show that the methods are fast enough for the intended applications, that their results are precise enough for the intended applications (Sections 7 and 8), and we show in a user study that useful applications can be derived with the help of these methods (Section 9).

We applied our system to a collection of about 5000 images while staying fully interactive without any significant delay. The collection includes portraits, individuals or groups of people, architecture, landscapes, and abstract paintings, done in different painting and drawing techniques. We also conducted performance tests with up to one million images (frames of cell phone videos). For some of our evaluation test we used a set of 200,000 images (frames of cell phone videos); other evaluation tests run with much lower number of images because of a large computation time (see Sections 7 and 8). The user study was done with 676 images because for this study it was not feasible to conduct the manual clustering of more images without support that had to be done for comparison purposes.

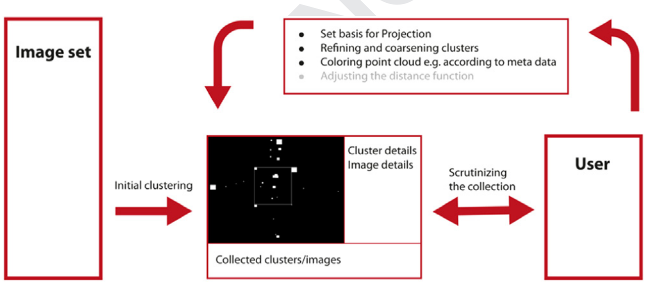


Fig. 1. Schematic workflow of our visualization system.

We thus demonstrate that the proposed system is feasible in principle and that useful applications can be generated with it. Of course, this work represents only a first step and is likely to be regarded as a feasibility study. Specific applications and suitable surface designs must be developed in collaboration with experts from the field of art theory.

2. Related work

The analysis of large quantities of images usually consists of the following steps: defining features that characterize the images; mapping the images in a feature space in which every feature defines a dimension (Fig. 2); and analyzing structures of the data in the feature space.

Images with identical feature values are considered identical, images with similar feature values are considered similar. The distance/similarity between two images can be easily calculated by using a metric of the feature space, as for example the Euclidean distance. This makes it easy to search for identical or similar images, or to find images with specific characteristics. Additionally, there are very efficient methods which use such distance metrics to cluster images, analyze relationships between features, and visualize image sets as point clouds to show structures and distances between the images (see, e.g., [2–5]).

However, we do not consider the procedure we just outlined suitable for our purposes. The properties listed in the last section require high-speed clustering and fast similarity-based visualization on a 2D display. Time complexity has to be at least quasilinear, so that interactivity is also possible for large image sets. There are many methods for clustering (see, e.g., [6]) and for distance-preserving 2D projections (see, e.g., [7–16]). However, all methods which are based on distance functions and have a time complexity faster than $O(n^2)$ work on the basis of a given distance matrix (which requires n^2 image comparisons) or a distance function that obeys the triangle inequality. In [17], Faloutsos shows a method for embedding data in a Euclidean pseudo space, which uses only a distance function and has only a linear time complexity, but the distance function used in this case has to obey the triangle inequality. Unfortunately, the result of distance functions calculating similarity in terms of human perception constitutes no metric in the mathematical sense. In particular, no exact statement can be made about the similarity between image A and image C if only the similarity between A and B and the similarity between B and C are known. We assume that if A is calculated as similar to B, and C as similar to B, then there are aspects that make A similar to C, but even this is not certain. It is possible to construct examples that disprove this. A monochrome red painting, for example, is similar to a portrait painted in red in terms of color. At

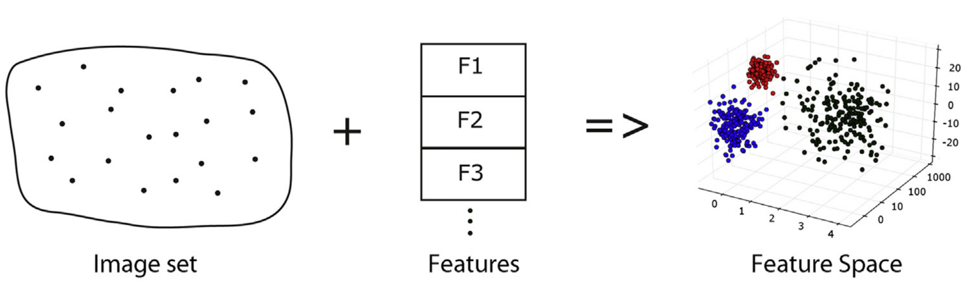


Fig. 2. Images represented by features. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)