VeCHArt: Visually Enhanced Comparison of Historic Art Using an Automated Line-Based Synchronization Technique

Hermann Pflüger, Dennis Thom, Anna Schütz, Daniela Bohde, and Thomas Ertl, *Member, IEEE*

Abstract—The analysis of subtle deviations between different versions of historical prints has been a long-standing challenge in art history research. So far, this challenge has required extensive domain knowledge, fine-tuned expert perception, and time-consuming manual labor. In this paper we introduce an explorative visual approach to facilitate fast and accurate support for the task of comparing differences between prints such as engravings and woodcuts. To this end, we have developed a customized algorithm that detects similar stroke-patterns in prints and matches them in order to allow visual alignment and automated deviation highlighting. Our visual analytics system enables art history researchers to quickly detect, document, and categorize qualitative and quantitative discrepancies, and to analyze these discrepancies using comprehensive interactions. To evaluate our approach, we conducted a user study involving both experts on historical prints and laypeople. Using our new interactive technique, our subjects found about 20 percent more differences compared to regular image viewing software as well as "paper-based" comparison. Moreover, the laypeople found the same differences as the experts when they used our system, which was not the case for conventional methods. Informal feedback showed that both laypeople and experts strongly preferred employing our system to working with conventional methods.

Index Terms—Visual analytics, user interaction, art history, qualitative evaluation, visual comparison

1 Introduction

Early modern prints open a complex area of study in art history. In the early 15th century, printing techniques such as woodcutting and engraving were used for the production of images. This profoundly changed the way images were produced and perceived. In contrast to paintings, prints are multiples, and they circulated on an anonymous market [1]. Quite a few artists specialized in engraving copperplates. One of them was Martin Schongauer (approx. 1445/50–1491), who published 116 engravings [2], [3], [4]. His compositions were highly appreciated; many were copied by other artists. Today, more than 400 copies based on his prints from the 15th and 16th centuries are preserved [5], [6], [7]. In his lifetimes, copying was a normal artistic practice and not considered illegal plagiarism [8], [9], [10]. These copies are a valuable source for understanding the print culture of the late 15th century. Art historians want to understand how these copies were made, how true to the originals or imprecisely they were executed, and how the compositions were simplified or enhanced.

- *H. Pflüger, D. Thom, and T. Ertl are with the Institute of Visualization and Interactive Systems, University of Stuttgart, Stuttgart 70174, Germany. E-mail: {Pflueger, dennis.thom, Thomas.Ertl}@vis.uni-stuttgart.de.*
- *A. Schütz and D. Bohde are with the Institute of Art History, University of Stuttgart, Stuttgart 70174, Germany. E-mail: {anna-christina.schuetz, daniela.bohde}@ikg.uni-stuttgart.de.*

Manuscript received 3 Aug. 2018; revised 10 Mar. 2019; accepted 22 Mar. 2019. Date of publication 29 Mar. 2019; date of current version 8 Sept. 2020. (Corresponding author: Hermann Pflüger). Recommended for acceptance by K. Mueller. Digital Object Identifier no. 10.1109/TVCG.2019.2908166

This challenge requires a precise comparative examination of the prints, especially in terms of details. Lines are the information medium in historical prints. The craftsmanship and drawing qualities of the artists who made these prints are indicated by their skills in creating images by means of lines. Moreover, the lines of these prints are influenced by features of the printing presses, features and condition of the paper and printing ink, and the state of the printing plates that were used. This is how a formal analysis of the differences in the lines of images that belong together also provides insights into the chronological order in which the prints were made, the restoration of the printing plates, and the entire printing process.

To assess and compare these features, researchers in art history require highly sophisticated methods that help them to detect, document, analyze, and categorize qualitative and quantitative differences among early prints, and to present them via visualizations. Another requirement is the development of techniques that examine and visualize the relationships among the images of a group of images that belong together.

In this paper we present a visual analytics approach that helps users detect deviations between different engravings. To make it easier for users to compare details and to detect differences between two prints, we visualize images in a synchronized fashion and provide meaningful visual cues to indicate possible areas of interest. In order to visually synchronize two images, we have developed a customized algorithm to detect the lines in prints, match the lines of two given images, and align image elements on this basis. Combined with overviews and detail representations as well as direct visual highlights, these algorithms are furthermore used to identify parts of the images where major differences can be expected.

We implemented our approach in the form of a highly interactive visual analysis system that allows users to compare and analyze multiple images in a short period of time. The evaluation of this system, which was conducted as a user study involving both experts and non-experts, shows that we achieve a 20 percent overall performance increase in the number of detected differences compared to traditional software-based and manual paper-based approaches. Furthermore, our results demonstrate that the non-experts can achieve performance levels comparable to those of the experts when using our system. This suggests that our approach could be used to leverage the support of laypeople and volunteers for time-consuming tasks that formerly tied up resources of art history professionals.

In summary, this paper provides three major contributions:

- a novel custom-defined algorithm to detect and align lines in historical prints;
- a visual interface that utilizes this algorithm to enable visual synchronization and automated anomaly highlighting to compare prints;
- results of an evaluation that conclusively demonstrates the applicability and usefulness of the approach.

The remainder of this paper is structured as follows. In Section 2 we first present background information and discuss related research in the areas of visual analytics of art images and the use of computer vision methods in this context. Our interactive visual interface for comparing images as well as the distinct alignment and highlighting techniques will be presented in Section 3. Our own methods for detecting stroke lines and matching similar stroke lines in different images will be described in Section 4. We conclude by reporting on the results of a comprehensive user study that demonstrated the usefulness of the approach in Section 5 as well as by some final remarks and ideas regarding possible future work in Section 6.

2 Related Work

In this section we will discuss existing approaches that are closely related to our work. Our primary contributions cover the areas of the analysis and comparison of visual art as well as methods that integrate computer vision tools with structural image analysis. We will discuss these two areas separately.

2.1 Visual Analysis of Works of Art

Digital archives of artworks do not only provide access to works of art that are difficult or impossible to access in their original versions; they also offer a variety of possibilities of comparing, searching, and communicating art and art contexts. Powerful image retrieval, comparison based on visual similarity, and image search via semantic annotations have therefore been extensively researched in the recent past. For example, Manovich [11] introduces image features that, viewed as semantic features, span a feature space in which images can be meaningfully located, and shows methods to structure and visualize image sets with the help of feature spaces. Manovich and Tifentale [12] introduce a system that allows users to explore a large number of manually annotated selfies. However, semantic annotations are costly and cannot be adapted flexibly to changing questions and image sets. Possible solutions may be found in crowdsourcing (e.g., artigo[1]) or automatic image analysis through object recognition and/or visual similarity [13], [14], [15], [16]. There are also archives that offer a variety of possibilities for communicating art and exploring art contexts; for example akg-images[2] or Google Arts & Culture.[3]

In our case, researchers need tools that do not only provide a broad range of means for image retrieval, but that also enable them to examine and compare digitized prints in great detail. The goals are identifying image differences that are not even apparent from the originals; making the investigation faster or easier; and enabling laypeople to execute certain tasks at expert level. Since these goals cannot be automated easily, visual comparison is used to show properties or relationships of images or image groups. According to Gleicher et al. [17], there are three common approaches used for comparative visualization: side-by-side comparison (juxtaposition); blending (superposition); and explicit difference encoding (aggregation), which means the presentation of the relevant differences. A comprehensive approach to a visual comparison of art image collections, ARIES, was presented by Lhaylla Crissaff et al. [18]. Their system comprises side-by-side comparison and blending in such a way that two images are displayed on top of each other, with the top image being transparent. In both comparative visualizations, zooming is possible, which is essential for examining images in great detail; however, the synchronization has to be done manually. The system of Lhaylla Crissaff et al. also offers explicit difference encoding; however, the differences are calculated only via simple pixel differences shown in a heat map. A range of commercial software offering more advanced pixel-based methods for detecting differences between identical-looking images, such as ImageMagick, PerceptualDiff or Resemble.js, already exists. However, most of these tools usually assume that the two images in question are almost identical photographs. Thus, they apply simple transformations (e.g., rotation, translations, scaling) to find mapping that reduces average pixel differences. They subsequently highlight deviations where corresponding groups of pixels still differ profoundly in terms of color, positional shift, or other low-level aspects. For images made with historical printing plates or copied manually by human artists, though, such pixel-based alignment will naturally fail in most cases. Although the images might look similar to a human observer, they will still be completely different in terms of their digitized representations (see Fig. 1a), and there is no simple function to align them.

Our application is focused on the precise comparative examination of prints, especially in terms of details. In this case it is indispensable that a comparative visualization offer zooming in order to be able to accurately compare details down to single lines. Furthermore, it is necessary that the application supports switching between details and context.

1. https://www.artigo.org/
2. https://www.akg-images.de/
3. https://artsandculture.google.com/

Fig. 1. Comparison of an image section taken from Schongauer's *The Dormition of the Virgin* (right) and its copy made by Monogrammist IE (left): a) side by side; b) one superimposed on the other, colored differently; c) using a lens tool.

There are many interface schemes that allow users to work at, and move between, focused and contextual views of a dataset. Cockburn et al. [19] review and categorize these schemes. They present four approaches: overview+detail, which uses a spatial separation between focused and contextual views; zooming, which uses a temporal separation; focus and context, which minimizes the seam between views by displaying the focus within the context; and cue-based techniques, which selectively highlight or suppress items within the information space.

There is a wide range of applications that deal with the analysis of drawings. They examine individual lines and hatchings in order to determine characteristic properties and to then use these properties for different purposes. Characteristics of handmade hatchings are examined in order to produce artificial illustrations of surfaces in such a way that the illustrations correspond in style and appearance to handmade drawings [20], [21]. The SAR approach [22] converts a drawing into a histogram of stroke attributes that is discriminative of authorship. The approach aims to automatically recognize the authorship of a drawing and to classify image sets accordingly. The approach VAICo [23] supports pixel-based comparison of images. Garces et al. [24] introduce style-based methods for exploring clip art. Lawonn et al. [25] present a method that uses line detection to visualize historical carvings. The method could probably be adapted to visualize historical prints and drawings in order to get impressions of their original shape and appearance. Schmidt et al. [26] introduce YMCA, an application for 3D mesh comparisons. At first glance, this application is closely related to our task. It allows users to compare different polygonal meshes describing a given real object. In our case, however, no real object is given that can be used as ground-truth data. Therefore, the statistical evaluations and visualizations of YMCA cannot be directly adopted for our case. Moreover, such far-reaching analyses, but also the interpretation and classification of historical prints, first require considerable effort to clarify the art-theoretical methodology and questions before suitable methods can be developed. In line with the needs of our target group, our work is therefore focused on examining prints in a purely formal way; questions about authorship, style, and artistic quality are not relevant in our case.

2.2 Computer Vision for Image Analysis

Our application is focused on the precise comparative examination of historical prints. For this purpose, the related elements of the prints to be compared have to be found at high zoom, which is tedious and exhausting if done manually. The strength of our system is that this matching is done automatically. The automatic matching based on the printed lines enables a system to synchronize visualizations of image pairs, and enables the system to do some automatic comparisons, which then provide clues about differences in the prints. As Fan et al. [27] write, line matching is a challenging task for a number of reasons: the inaccuracy of line endpoint locations; the non-availability of strong disambiguating geometric constraint; the lack of rich textures in line local neighborhood; and so on. In accordance with Wang et al. [28], they name four principally different groups of methods:

1) Methods that rely on intensity or color distribution of pixels on both sides of line segments to generate line segment matches.
2) Methods that transfer matching line segments to matching points, because matching points have already been widely investigated. Most of these methods first use some strategies to intersect line segments to form junctions and then utilize features associated with those junctions for line segment matching. Other methods match a large group of points using the existing point matching methods to determine the underlying image transformation, for example, the method of calculating the epipolar geometry of the two camera views of a 3D scene, and using this information for matching the lines.
3) Many line matching approaches that match individual segments based on their position, orientation, and length, and take a nearest line strategy.
4) Some methods that first divide line segments into groups and then generate a descriptor of the configuration of the line segments in each group by calculating the relative positions of these line segments. The descriptor can be used to generate a similarity measure, which in turn can be used for image matching.

In our case, the first two groups of approaches can only be used to a limited extent. We do not have pixel information like color or luminance; instead, we can only use information such as line density, line width, and line intensity. Line density can only be determined for a certain image area, and it accordingly has a low spatial resolution. Line width and line intensity are not very specific for individual lines in engraving, especially in the high line-density areas. Therefore, line density, line width, and line intensity alone do not provide enough information for accurate matching. Furthermore, the approaches in group three offer only limited usefulness in our case, because the line sections, in particular those of the hatching, are too unspecific to be suitable for identification and thus matching. In our application, we use line width, line intensity, and the course of the lines to specify the individual lines, which in turn is the basis for matching. In terms of methodology this is a combination of the methods of groups one and three.

The methods of group four, that is, the grouping of segments, are widely used in object recognition and detection. They are based on perceptual properties such as connectedness, convexity, and parallelism, so that the segments are more likely on the same object. Lines that define the shape of objects in engravings are naturally single lines and can therefore not be treated with such methods. Our application for matching, however, can possibly be combined with the methods given by Wang et al. [28] to improve the matching of hatching. We will review this in our future work.

Another method of comparing images is to calculate the optical flow between two images (see e.g., [29]). This is implicitly a matching of all pixels of two images, and is used to examine consecutive frames in videos. These methods are not designed to handle large image deformation and cannot handle engravings created by different copyists; however, matching prints made with the same or a revised plate can be done effectively with this method. Preliminary tests with this method rendered good results.

Line-matching is based on lines that have to be detected beforehand by a line detector. In our case, the lines that form shadows and three-dimensional effects are pivotal in understanding the techniques, styles, and restoration methods of historical printing plates. We thus need to achieve high precision at detecting lines with the width and distance of only one or two pixels, erratic lines, lines with complex curvatures, as well as broken lines (Fig. 4). Depending on the printing technique, there are many thousands of lines in historical prints. To the best of our knowledge, no specific line-detecting algorithms are currently available to support such an analysis.

Our investigation of line detectors revealed three different principles of detecting lines in pixel images. The first principle is the use of an edge detector to detect potential line edges, and then, in a second step, filtering out the actual lines along with the associated characteristics. The second principle is performing a segmentation to extract pixels that are potentially elements of a line, and then, in a second step, filtering out the actual lines along with the associated characteristics. The third principle is the comparison of the distribution of luminance for each pixel within its surroundings with the typical distribution of luminance around a pixel lying on a line. This directly yields line points, which are connected to line segments in a second step.

In Section 4 we discuss the suitability of the principles for our case in more detail. In the following we give a brief overview of existing methods.

Most computer vision methods for image analysis focus on the characteristics of realistic photographs, satellite imagery, biometric properties, or image sequences. However, a number of scholars have also adapted and applied these methods to the realm of analyzing the features, structure, and composition of historical paintings. David G. Stork [30] presents a comprehensive survey of 125 works in the field. Among the different classes of comparative art analysis identified by Stork, the methods based on brush strokes and marks as elements of analysis and communication are most similar to our work. Here, major challenges are finding brush strokes in paintings automatically and extracting their distinct properties for the analysis and comparison of the paintings [31], [32]. Most of these approaches are based on techniques that detect edges in color, luminance, or structure similar to the popular Canny Edge Detector [33]. Another method of finding and analyzing strokes starts with vectorization to obtain strokes as simple graphical objects so as to derive stroke characteristics from them (e.g., [22]). Vectorization is based on image segmentation. The segmentation of drawings, however, can be directly used for line detection. The study [34] describes a fingerprint recognition approach that accurately detects lines in scanned fingerprints. After the segmentation of the scans, the center lines of the line segments are singled out via skeletization. Gerl et al. [20] and Kalogerakis et al. [21] use the same method to detect lines in man-made sketches used for learning hatching for pen-and-ink illustrations and for example-based hatching. Ziou [35] examined the luminance distribution in the surroundings of centerline pixels and developed a filter for line detection from these findings. Lawonn et al. [25] used a similar approach for detecting lines in carvings, and our method presented here is based on a similar idea.

3 Visualization and User Interface

In order to provide large-scale art-analytical tasks involving many different prints from various artists, the comparison method presented in this work has been integrated with an analytical workbench, VeCHArt, and paired with a sophisticated image database [13]. While this image database is not the subject of our work presented here, we are giving a brief description of the database to provide an overview of the context of VeCHArt:

In our image database, each artwork is considered as a single entity. Each artwork consists of at least one image. In addition, freely definable metadata, detail images, and text can be associated. The additional information can be retrieved at any time. When a new work of art is inserted into the system, the characteristic areas [36] in the image are calculated and stored along with their local image data. These characteristic areas are also used to calculate a visual similarity between the images. An embedded image browser provides the following functions:

- Users can navigate through the entire image dataset or a subset and arrange elements as needed.
- Users can find and order images according to their metadata or via visual similarity search. They can furthermore employ automated clustering to arrange images based on these properties.
- Each stored artwork can be manually annotated with user-defined metadata, detail images, and textual notes.

The user interface is divided into three linked views with manually adjustable sizes (Fig. 2). The topmost view (exploration view) is the main view and provides common exploration methods for image databases. In this view, a user can navigate through the entire image set or a subset, change the sizes of the individual images, and arrange the images as needed. Here, a user can find images according to their metadata. A user can also find and sort images according to their visual appearance, and form groups of images which are visually related. The bottom left view (group view) shows images that have been collected by the system, as they are visually related,

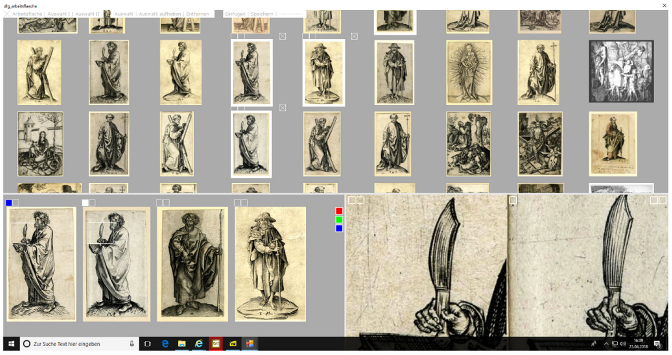

Fig. 2. The three linked views of our image database system. Exploration view (topmost view), group view (lower left view), and detail view (bottom right view).

or have been manually collected by the user. The bottom right view (detail view) shows details of a particular image or specific applications. In our case this view shows the synchronized sections of two images side by side. The group view and the detail view are used for the current application and have been modified accordingly.

3.1 The VeCHArt Visual Analytics System

When we started this project, we carried out extensive interviews with two professional art historians, the co-authors of this paper. The aim was to identify central issues in their scientific work, which is the study of historical printing, and to find out what functionality and properties a system must offer to support this work. We learned that the study of historical printing needs to take into account the historical context of print production. The technology and social significance of prints at that time form the background for the scholarly research of art historians who deal with prints from this period. The main focus of their investigations is on the comparison of the preserved prints. The various copies provide information on the distribution numbers of a given picture motif, thus showing the importance, appreciation, and dissemination of individual images. The changes that were made during the restoration of printing plates and the changes in copies reveal information about changes in the perception and appreciation of individual motifs.

Our interviews with the art historians revealed four requirements for a system that supports the study of historical prints:

- *R1:* Our target group's main task is the comparison of related images, in the course of which these images are not only compared in their entirety, but, with utmost exactness, down to single lines of a width of one or two pixels.
- *R2:* While the smallest details have to be examined, they must also be assessed in the overall context. Therefore, even if details are examined at high zoom, awareness of the overall context should not be lost. That leads to the second requirement, namely, that detail image views must be visually related to the image as a whole.
- *R3:* Historical engravings consist of many thousands of lines. An absolutely accurate comparison down to every single line is time-consuming and tedious. There is therefore a need to receive at least hints at or visual cues alerting users to possibly different areas. If these cues are provided, the differences between two images ideally need to be only examined and confirmed.
- *R4:* Art historians see working with originals as essential. Thus, in order to maintain an unobstructed relation and close resemblance to the originals, the images should not be visually distorted or overlaid with artificial visual elements during handling or synchronizing.

The design decisions for VeCHArt resulted from the requirements listed above. The most important requirement derived from the task of visually comparing picture details (*R1*). In order to accomplish this task, art historians have to examine the corresponding image areas in high magnification. That is why the actual task is to support the comparison of image details by visualization means. According to Gleicher et al. [17], there are three common approaches used for comparative visualization: side by side comparison, superposition, and explicit difference encoding.

When not working with the originals, art historians commonly open the images to be compared in a photo viewer (e.g., Windows Photo Viewer), each side by side. In this way, they can view and compare image details in high magnification. However, if they proceed in this manner, they have to synchronize the two views manually. There are tools (e.g., [18]) that synchronize the two views with manual support. This is a reasonable effort if the two images to be compared can be mapped on each other by an affine transformation, because in this case the images only need to be synchronized once. In our case, however, the single image elements are often somewhat differently placed and differ in shape and size, and the images have to be constantly dubbed. That is not only tedious, but makes it difficult to compare individual elements in terms of size and local shifts. The most important design decision for the side-by-side presentation in VeCHart was synchronizing the zoomed views so that both views always refer to the same image detail.

Superposition is not suitable for engravings. Different drawing techniques, styles, and varied differentiations of details lead to confusing overlays of the lines (see Fig. 1b). Even in the case of prints that were printed with the same printing plate, the lines show deviations because the printing paper deforms differently during drying, which leads to confusing overlays.

In order to be able to synchronize image details in a side-by-side presentation, a system has to be able to recognize the associated image details. The information medium in historical prints are lines. Therefore, the matching of the image details has to work with line matching. We use our line detection/matching method not only to synchronize image details in the side-by-side presentation, but also to automatically detect differences in single lines and line structures, and to provide this information via explicit difference encoding (*R3*). Our approach distinguishes between two types of line differences: one regarding shifts and the other regarding changes in line shape. The image areas with large line differences can be displayed via highlighting in the side-by-side presentation of the details; however, we decided to present them separately in different views. Two reasons were central to this decision. On the one hand, *R4* implies that the target group prefers an unadulterated representation for the visual

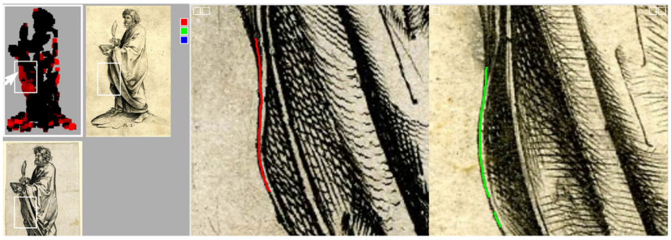

Fig. 3. Local deviations displayed via a heat map in the group view. (*St Bartholomew*; Monogrammist IE and Israhel van Meckenem).

comparison of the image details; on the other hand, a detail view is necessary for the visual comparison, but the representation of areas with strong differences only makes sense in an overall view. Additionally, we calculate how long image areas are being examined, and show this information for the same reasons as before in a separate overall view.

In comparisons with high magnification, there is a risk that users may lose track of the overview, overlook areas of the image, or lose themselves in details. Therefore, it is necessary that our application supports changing back and forth between details and context (*R2*). There are many interface schemes that allow users to work at, and move between, focused and contextual views; however, it actually takes only an overview+detail [19] to get unadulterated views (*R4*). Therefore, in addition to the side-by-side presentation of the image details, we show each image individually as a whole within which the detail area is marked with a frame. The marker can also be used to navigate the detail view (pan and zoom). Navigating by means of the markers has two advantages. On the one hand, it supports a systematic investigation in which all image areas are examined equally thoroughly. On the other hand, it makes it easier to detect local image shifts, since local shifts are not taken into account in the overview display but in the detail view.

In the following we give a short formal description of our user interface. We divided our image comparison user interface into three linked views with manually adjustable sizes (Fig. 2). The topmost view (exploration view) is the main view and provides common exploration methods for image databases. The bottom left view (group view) shows collected images. The bottom right view (detail view) shows details of a particular image pair. Once the user has selected two images out of the group view, these images are displayed side by side in the detail view. In contrast to the images of the group view, where the original scans are shown, the detail view shows the two synchronized images (*R1*). There are two types of synchronization. For one type, if the mouse pointer is not in the detail view, the detail sections are synchronized only based on the basis transformation. This is the affine transformation that is automatically done on image pairs to make the images congruent, but focused on limiting visual distortion within the image content (*R4*). Because of this synchronization, the view displays the same snippet of the two images in the adjusted size, but ignores local shifts or distortions. If there is an area in one of the images that is not given in the other image, for instance because one of the images was cropped, then the area will be marked in color. For the other type of synchronization, where the mouse pointer is within the detail window, the synchronization also takes into account the local shifts and distortions. The snippets are distortion free (*R4*). However, the centers of the two snippets are synchronized, taking into account the local displacements or distortions, that is, the average deviation within the radius of 32 pixels around the center. The synchronization is automatically performed based on the matched lines of the images. In the detail view, a user can also examine single lines. In this case, the line closest to the mouse pointer will be displayed along with the corresponding lines of the other image. At user's request, all detected lines can also be displayed at the same time.

The user can navigate within the detail view in a manner similar to a traditional image viewer (*R1*). Panning can be done by moving the mouse while holding down the left mouse button. Zooming can be done with the mouse wheel. Additionally, it is also possible to navigate the detail view using the minimap representation of the group view (Fig. 3; *R2*). While the mouse pointer is within the marker box, the line of the image nearest the pointer is highlighted in the detail view, and panning and zooming can be done the same way as in the detail view.

At user's request, a visual overview of overall image differences and anomalies is provided (*R3*). This information is displayed separately as a pixel-based heat map [37] in the group view. Direct heat map overlays in the overview representation or in the detail view are not used, since they would lead to unwanted obstruction as described in R4. Just as in the detail and overview representations, the heat map can also be used for navigating the detail view. There are currently three different heat map types available:

- The base transformation aligns the image pair as a whole. The local deviations and distortions can be determined by means of the matched lines. The first available heat map displays these local deviations (Fig. 3). In this case, the average deviation for each point of the image within a radius of 32 pixels is shown. We selected this radius as people see this area with complete acuity while fixating a pixel in an original print at a distance of about 30 cm.
- For the second heat map we calculate a measure of the copy accuracy by means of the percentage of one image's lines that can be assigned to a line in the other image and the quality of the matched line that the matching algorithm determines.
- The third heat map provides a visual log of the ongoing investigation. The durations for which zoomed sections of the image were visible to the user in the detail view are recorded. The total time that each pixel was visible in the detail view is then used as a value shown in the heat map. The third heat map thus illustrates how intensively certain image areas have already been investigated.

3.2 Basic Workflow

The workflow for comparing two prints with our approach is quite simple. If the images to be examined are not yet available in the database, they must first be inserted. In the simplest case, only the scan of the print has to be uploaded. The resolution should be about 300 dpi for adequate performance. In addition, detail images, freely definable metadata/connotations, and textual notes can be entered via the input window. Once all relevant images are in the database, the user

Fig. 4. Detail from Schongauer's *The Virgin of the Annunciation*. Left: the original image; right: the original image with the superimposed automatically detected lines.

can find them according to their metadata and/or visual similarity, and group them in the group view. In this view, two prints can be selected for direct comparison with the comparison view.

The comparison process depends on what the user prefers. The detail view shows a user-defined section of the two prints side by side. As described above, these two sections are synchronized, i.e., they always show the interrelated sections of the prints. The user then navigates the detail views as with a common image viewer. Some researchers only used this possibility to compare two prints, but most researchers also used the heat map analysis data. Those who used the heat maps either used this information to get started with the investigation, or they used it towards the end of the investigation to check whether important differences had been overlooked. Some researchers used the information on how long certain regions in the prints were examined during the study to ensure that all regions were examined with the same effort. Most of them, however, used this information only towards the end of the session to check whether certain areas were not or only insufficiently examined.

4 Line Detection and Synchronization

To address our problem, we initially examined common single line-detection methods (e.g., [38], [39]), but none of these adequately matched the challenge at hand. The existing line-detection methods usually consist of three stages:

1) Detecting pixels that are potentially part of a line.
2) Examining which of the detected pixels actually belong to a currently tracked line.
3) Filtering for the line characteristics relevant for the given application (e.g., course, width, or intensity).

The following difficulties arise if techniques that follow these stages are applied to the specific prints covered in our work.

Stage 1. Most of the approaches are based on techniques that detect edges in color, luminance, or structure [33]; the Stroke Width Transform (SWT) [40] is also based on edge detection. To reduce the influence of noise (e.g., yellowing and stains), one has to apply an intensity threshold to the images before edges are extracted. The lower the threshold, the more lines will be detected, and the result will be increasingly susceptible to noise and to the detection of edges from irrelevant features. Conversely, a high threshold may miss subtle lines, which are relevant in our case. Our prints have both noise and subtle edges (Fig. 4). Our experiments showed that it is difficult, if not impossible, to find automatically suitable thresholds. In particular, the situation is different with each print, so at the very least, one has to find individual thresholds for each print, which makes this approach unsuitable for our purpose.

Another possibility is to segment (vectorize) an image so that the pixels of lines are black and the background pixels are white. With a skeletonization method, the lines can be reduced to the width of one pixel, which greatly simplifies the subsequent line detection (e.g., [20], [21], [34]). In our case, however, the images are usually too noisy for the necessary segmentation. One would have to find at least range-/print-dependent threshold values, which leads to the same problem as before with edge detection.

A third possibility is to use the typical luminance distribution in the surroundings of line pixels to develop special filters for the detection of line points (e.g., [25], [35], [41]). Our approach is close to these methods; it is a simplification, an adaption for cases involving line widths of one to ten pixels and line spaces down to two pixels, where we directly detect the line center pixels and, unlike the methods mentioned above, not line pixels in general.

Stage 2. Lines in images made with historical printing plates do not have smooth shapes. In particular, their edges and courses can be erratic, and the lines are often very close together. The printing technique and the aging process further reinforce this property. Traditional edge linking methods (e.g., [42], [43], [44]) that combine pixels detected by an edge detector (or a detector of line points) thus often stop the tracking too soon or connect the wrong pixels. Furthermore, the hand-drawn lines usually do not resemble parameterizable geometric figures (e.g., straight lines, simple polynomials, or circular arcs), and thus methods that recognize certain line shapes or specific shape templates, such as Hough Transformation, methods that use Gestalt theory [38], or Eigenvalue analysis [43], do not seem suitable to combine detected line fragments into the lines appropriate for our case.

Stage 3. The handmade strokes in our case are irregularly curved lines of varying thickness, partially touching, or crossing each other. It is difficult, if not impossible, to determine the center line and line thickness if only the line edges (line points) are known. An interesting approach has been suggested by C. Li et al. [45]. Using a deep learning approach, it detects the outlines of Manga figures and eliminates lines or points that are only shading. However, in this approach, potential lines must first be detected, and in our case, lines for shading may not be eliminated. In addition, in Manga figures the shape-forming lines are distinctive and contrast strongly with the background, which does not hold true for our shape-forming lines. Therefore, this approach cannot be used in our case.

4.1 Line Detection Algorithm

In order to avoid the aforementioned difficulties, we created a customized line-detection method tailored to the types of images and analytical problems addressed in this work. It is furthermore specifically suitable to be used as a basis for our visual interactive alignment and highlighting.

The requirements for line detection in historical engravings and for analytical purposes are:

- The method should detect most of the hand-drawn lines visible to the expert eye.

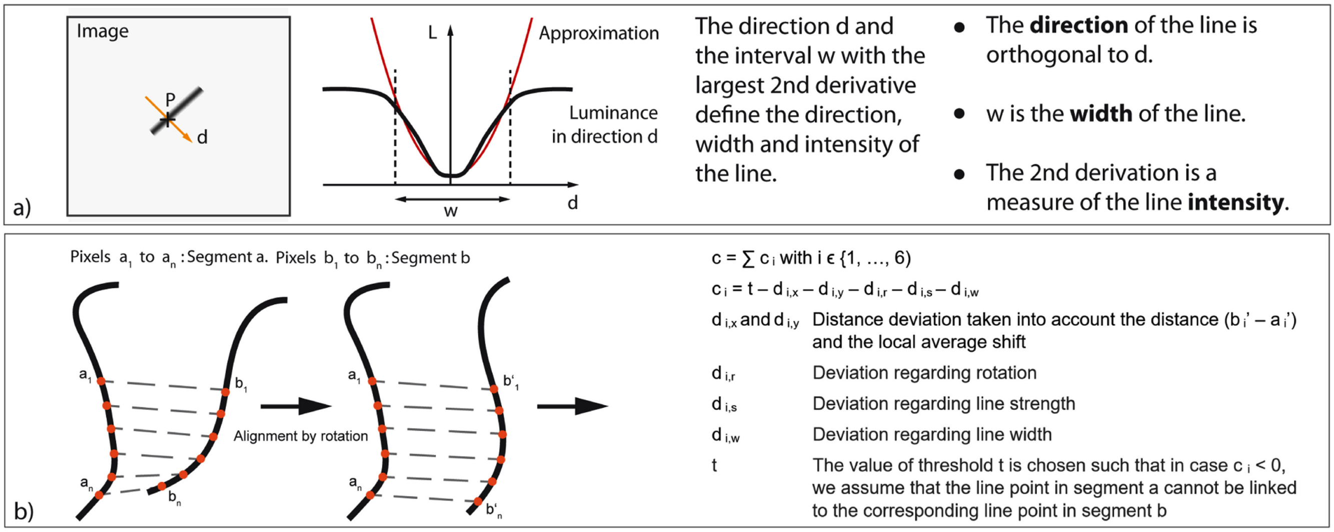

Fig. 5. a) Determining points of center lines using local approximations of the luminance by bivariate quadratic functions. b) Calculating the closeness c of two line segments.

- The algorithm should identify lines of one to ten pixels' thickness and accept line spacing down to two pixels. This requirement is based on a scan quality of 300 dpi, which is typical of scans of historical prints.
- The scanned image material was often yellowed, faded, or soiled. Therefore, the method should ignore anything that is not a hand-drawn line.

To meet these requirements, our method consists of two stages:

1) Finding pixels that are potentially on the centerline of a line, and determining the course, width, and the strength of the associated line.
2) Examining which of the detected pixels actually belong to a line, and defining the resulting lines, with their specific direction, local width, and strength.

Stage 1. In engravings, the variation of the luminance (of the Lab color space) orthogonal to line direction has a typical functional profile [35], with a minimum exactly at the center of the line. We approximate this typical function by a bivariate quadratic function ($f_{(x,y)} = ax^2 + by^2 + cxy + dx + ey + f$). Therefore, the first step of our line detection is to perform five local approximations of the luminance for each pixel in the image: one for the area with a distance of one pixel to the respective pixel; one for the area with a distance of two pixels; and so on. If one of the corresponding functions has a minimum in a spatial direction at the respective pixel (we consider the function with the largest second derivative if there is more than one), one may assume that a line exists orthogonally to this direction. In this case, we build on the following properties, see Fig. 5a, and calculate the direction, width, and strength for the potential line.

- The width of the line corresponds to the size of the area of the respective approximation.
- The strength of the line is directly related to the size of the second derivative of the local polynomial orthogonal to the line.
- At the pixel closest to the center of the line, there is the largest approximated second derivative.

Stage 2. At the second stage of the line-detecting method, individual pixels are merged into lines. Starting with the pixel with the largest line strength, the following steps are performed with each pixel that is potentially located near the center of a line.

If a pixel is already connected to two pixels, then the pixel is already an element within a line and nothing else is to be done with it. If a pixel is connected to only one other pixel or none at all, then the pixel (or the open end of a line) may possibly be extended at one side in local direction. In this case, the immediately neighboring pixels in local line direction are checked as to whether they can be appended. Pixels suitable for attachment are neighboring pixels of similar local direction, line width, and intensity. If one pixel is connected to another, then the line strength of the neighboring pixels orthogonal to the local line direction corresponding to the local line width is set at zero as the latter belong to the current line and are not part of a centerline.

After all pixels that are potentially on the center of a line have been merged as far as possible, the method uses a heuristic to check whether potential small gaps (of a maximum of five pixels) between two lines can be bridged. If two lines intersect, the corresponding point of intersection can be assigned to both lines. Our method detects pixels that are potentially at points where two lines intersect; in these cases there is no unique direction with a minimum of the luminance at the corresponding point, and a pixel can be connected to two different lines.

4.2 Matching Algorithm

Fig. 1a shows a detail of two prints of the same motif, but by two different engravers. A human observer can easily see that most of the lines are directly related to each other as the left print is a copy of the right print. However, Fig. 1b shows that most of the lines are actually different, both in position and in shape. For a human observer, recognizing the connection between the lines requires object recognition and the ability to interpret drawings. For a computer, that would be a difficult and resource-intensive approach. However, it can be observed that characteristic lines of a given engraving are often copied with particular precision and thus these lines are often consistent, at least in terms of course, in different copies. In addition, these characteristic lines are often relatively long. We have thus developed a method of matching the lines of two prints that exploits all these features in order to deliver satisfying results and performance. As our method involves a broad range of 'hard-wired' heuristic optimizations, which are specifically targeted at the task/images at hand, we will only provide a general overview of the basic principles.

Initially, we perform an affine transformation (translation, rotation, or scaling) with one of the images aimed at making the two images congruent. To do this, we divide both images in height and width into three equal parts. This results in nine equal-sized sections in each of the two images. We look at the sections of the second column and the second row. For each of these sections, we calculate the center of all the line points detected there, and calculate an affine transformation that approximately superimposes these line centers. In the following steps, we work exclusively with the transformed image as if it was originally given in this form. Thus, when talking about different positions of image elements or shifts, we mean different positions or shifts after this transformation.

The detected lines are given by successive pixels that define the center of the line. For each pixel of the center line, the line width and the line intensity are recorded. The matching procedure is iterative. At each iteration step, all line segments, starting with the segments of the longest lines, are compared with line segments of the other image. The comparison of two line segments involves the following steps:

1) Closeness of two line segments is calculated so that a high value of this weight indicates a small distance (see Fig. 5b). The closeness weight depends on the spatial distance of the segments; on the local average shift, which is zero at the first iteration step; on the similarity of the segment course; and on the similarity of the width and strength of the segments. In particular, the value grows directly with the length of the parts of the segments that match well. This leads to large closeness weights for long line segments that coincide very well, and to small values for short line sections or line sections that do not coincide well.
2) If a line segment is not already matched, then both segments will be matched. If a line segment is already matched and the new distance is better than the old one, then the old match will be canceled and both segments will be matched.

At the end of each iteration step, the average local shifts for all image pixel are calculated based on the matched line segments within a range of 32 pixels around the respective pixel (human observers usually see this area in full sharpness). Then the next iteration step is carried out.

4.3 Evaluation and Performance

To evaluate both the line-detection and line-matching algorithms, we generated ground-truth data by manually highlighting individual lines of an engraving. We then first identified the lines with our line detection and then assigned them to the manually detected lines using our matching method. We considered detections and matching correct if we found a local deviation of less than or equal to two pixels between automatically detected (center) line points and associated manually determined line points. As a result, we found that 83 percent of the manually detected line pixels had also been recognized and assigned correctly by the automatic detection and matching. This number of line pixels covered 91 percent of the automatically detected line points (approximately 200,000); the remaining 9 percent had a greater deviation than two pixels or could not be assigned.

Evaluating the line matching between prints of different copyists is more challenging because the assignment of lines is more dependent on the viewer's judgment and can hardly be objectively determined. In addition, there are only few possible variations of short lines (that is, lines of about five pixels in length). Therefore, short lines usually have many suitable matching partners, and in print areas with intersecting lines it is often only short line fragments that are detected automatically. When comparing two engravings (considered as belonging together by the experts) with our method, 60 to 90 percent of the detected line points of the engraving with the smaller number of line points could be assigned to line points of the other engraving. When we rated the matching of prints made with different plates, we considered two points as correctly assigned if the deviation between these two points was not greater than five pixels compared to the respective average local shift. The large difference (60 to 90 percent) arises because some of the image pairs were made with the same (or revised) printing plate, while other pairs were made by different copyists, using various ways of copying.

The assignment of long lines was usually plausible, and the assignment of many short lines in a specific area was usually in the corresponding area of the other engraving. In our study, the synchronization of two images by means of the detected and matched lines led to plausible results, and none of the participants questioned the correct synchronization.

When adding a new print to our image database, we automatically perform our line detection and store the result. Thus the line detection has to be done only once. The computing time depends on the number of line pixels and is between 30–120 seconds per image (Intel i7-2600 processor with 3.4 GHz). When a new image is inserted into the image database, similar images are automatically determined, and line matching is performed with the image inserted and the images found. For this purpose we consider two prints to be similar if the number of line points in the image sections of the initial affine transformation (Section 4.2) of the two prints correlates with a factor greater than 0.6. However, at the request of the user, line matching can be performed for any image pair. The results for the corresponding image pairs are stored. The computing time for the matching algorithm depends on the number of line pixels and was between thirty seconds to eight minutes for the images used. If an image is matched with several images, the matching is carried out in parallel.

5 Evaluation

Owing to our co-authors' background in art history, we know that the comparison of historical engravings is preferably carried out on the basis of originals. Working with the originals can be regarded as ideal, because only then is visual access genuinely possible. However, access to originals,

Fig. 6. Details from of nearly exact copies after Schongauer's *The Tribulations of St. Anthony*.

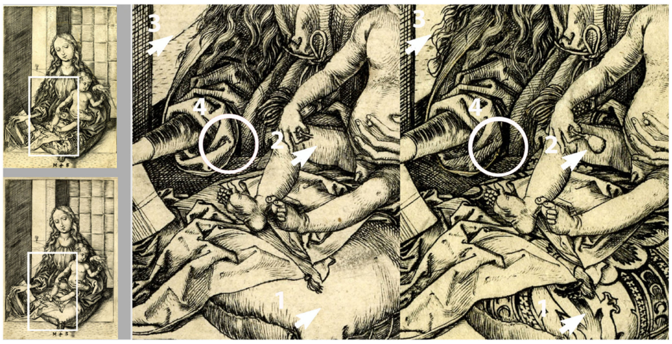

Fig. 7. Two prints of Schongauer's *Virgin and Child with a parrot*.

especially simultaneous access to multiple originals, is only available in a very limited number of cases. For this reason, the direct comparison of images to identify formal image differences is typically done with digitized images. In this case, the usual procedure is to open several images for the comparison in parallel in a photo viewer. Comparing images that way does not seem to be efficient and might benefit from interactive visualization and visual analytical procedures. The approach described in this work is aimed at providing this benefit. In this section we report on a thorough user study that was conducted to evaluate how well this goal was ultimately met.

Before we go on to discuss the results of this user study, we will first illustrate the utility of VeCHart based on two possible findings that can be derived from the images. This will help to illustrate the kinds of analytical challenges on which our approach is focused and make the results of our user study more understandable. In the final part of this section, we will again return to these challenges to discuss them as case studies.

5.1 Example Challenges

Prints made with the same plate are often compared in order to clarify chronology and circulation. Fig. 6 shows two related details of prints made by Israhel van Meckenem and Monogrammist FVB after Martin Schongauer. In this presentation, it can be seen that van Meckenem retouched the plate engraved by the Monogrammist FVB after a composition by Martin Schongauer; it is evident that some areas have been restored in the left-hand print and that the non-revised lines are much fainter than in the picture on the right. This implies that the print on the left was printed in a subsequent run, and that the plate was already in a very worn state when the print was done. This in turn indicates a high circulation and thus high distribution of this print. Additionally, strongly faded areas were indeed restored, but changes of the motif were carefully avoided.

In the second example, both prints were again made with the same plate (Fig. 7). Both plates show two different states of a print. The first state shown on the right is a trial proof. The second state on the left is a revised version. Four significant changes were made with considerable effort. Zooming-in reveals that elements were deliberately removed, rather than having been worn out by use, having disappeared due to damage, or having been added later. The examination of all the differences shows that the plate was not revised to repair fading due to a high number of prints, but that the motif had already changed after a small number of prints.

5.2 User Study

Our study was based on the comparison of ten pairs of engravings from Schongauer and copyists. All images are engravings from the 15th century. The two images of a pair are either copies of the same original, or one image is a copy of the other, or they are from the same plate, which had been restored or edited between the prints. The user task was to find and name the most significant differences in each pair. The named differences were recorded by the experimenter and marked in a printout of the image pair. For the comparison of two images, roughly ten minutes were scheduled, albeit without giving the participants a time limit. The user study involved two groups of participants with ten subjects each. One group (*G1*) consisted only of art historians or students of art history. All participants in this group had considerable experience comparing historical prints with the help of photo viewers. The other group (*G2*) consisted mostly of students recruited from our computer science faculty. These participants usually had some experience in dealing with tools for editing and analyzing images, and all of these participants had at least occasional experience with photo viewers. However, they did not have experience in analyzing historical art in general or historical engravings in particular. They can thus be considered laypeople in terms of historical art research. While solving the given tasks, all participants were asked to provide think-aloud comments, which were recorded for subsequent analysis. In addition, we captured their performances using screen-recording software. Prior to the analytical sessions, all participants had to perform a basic vision test, which all participants mastered without any problems.

In order to evaluate whether our approach leads to actual improvement, the participants performed some comparisons with our method, and others employing the commonly used method: they opened the images to be compared in a photo viewer (in our case, the Windows Photo Viewer), each side by side, while the zooming and panning of the two views have to be synchronized manually.

The participants examined four to six image pairs with our approach and two image pairs using the photo viewer. Finally, two image pairs had to be compared based on printed scans without any digital help in order to obtain an approximate comparison with investigations using originals. The comparison methods alternated between image pairs and participants. In total, each image pair was examined four times using the usual method, four times using the printed scans, and eight to twelve times using our new approach. Furthermore, the participants had to answer a questionnaire for

TABLE 1
Differences per Image Pair

ID of image pair	1	2	3	4	5	6	7	8	9	10
Total number of differences	19	22	30	21	27	26	36	16	26	32
Average number of detected differences	4	12	11	10	13	11	15	5	11	11

additional informal feedback about each of the approaches immediately after the sessions.

5.2.1 Task Performance Results

For the evaluation of the task performances, we collected all the differences that were found per image pair. Since the number of observable differences varied strongly between the pairs, we normalized our results accordingly. For each image pair, we used the average number of differences found by the participants over all three methods as the baseline (Table 1). For each participant and image pair, we recorded the absolute number of detected differences. We then used this number to calculate the (positive or negative) percentage change compared to the baseline. The average percentage changes for all participants as well as for the two individual groups can be seen in Fig. 8.

The chart illustrates that our approach increased the number of detected differences in image comparisons considerably. Looking at all participants in the study, our approach achieved a performance gain of 8 percent over the baseline, whereas the performance of the conventional method with photo viewers was 16 percent below the baseline. If we look at the different groups of participants, we can see that the gain in performance between the new and the conventional methods in the laypeople group (*G2*) was even greater (29 percent). The group of art historians (*G1*), on the other hand, already performed better than the laypeople using the conventional methods with photo viewers or printed scans. However, using the new approach, they were still able to increase their performances by 18 percent compared to the photo viewer and by 5 percent compared to the printed scans. Our number of subjects was of course not sufficient to make accurate quantitative statements. However, even if the confidence intervals are included in the assessment, there is still a clear performance gain with the method presented here compared to the conventional method. The performance gain of the new method over the method using printouts seems a bit

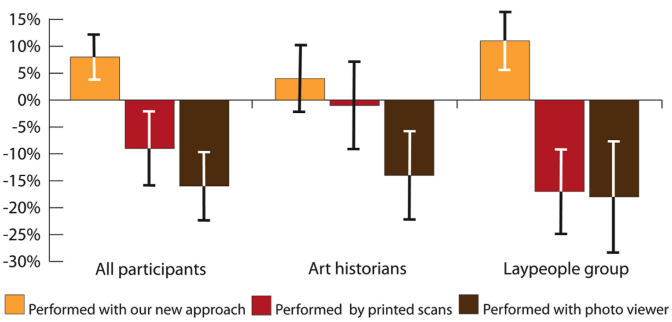

Fig. 8. Detected differences as a percentage change compared to the average number of detected differences (mean values with 95 percent confidence intervals); subdivided according to aids and participant groups.

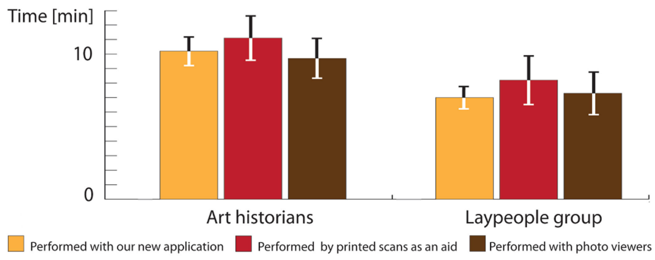

Fig. 9. Time needed to examine a pair of prints (mean values with 95 percent confidence intervals); subdivided according to aids and participant groups.

uncertain, but the fact remains that the use of printouts is not suitable: for one thing, because of the printing costs, and for another, because the examination of very subtle differences would require the use of mechanical magnification aids (see user study 1), which would certainly degrade performance and usability significantly.

The recorded task timings show that our method had no influence on the time the participants took to examine the images (see Fig. 9). The average for both groups was at around 9 minutes per image pair. However, it was slightly different between the group of laypeople (7.5 minutes) and the art historians (10.5 minutes). This might be due to the request to perform think-aloud comments during the tasks. The art historians more frequently discussed nuances of image contents and drawing techniques whereas the laypeople were just focused on detecting the visible differences.

We also examined whether art experts find other differences than the non-experts. To this end, we split the image differences into classes. Class 1 contained all the differences discovered only by one art historian, class 2 contained all differences discovered by two art historians, etc. For all differences in a class, we calculated the average number of laypeople who had also discovered these differences. The result shows that the frequency with which a difference is detected is nearly the same for experts and non-experts. This means that differences that are rarely perceived by experts are also rarely noticed by non-experts, and differences that are perceived by many experts are also reliably perceived by the non-experts. This indicates that non-experts determine formal differences in historical prints with a similar quality as experts.

5.2.2 Think-Aloud Results

Furthermore, the evaluation of think-aloud and screen recordings revealed that the utilization of the analysis tools, the heat maps for local deviations and the display of the regional examination times, varied significantly between the participants. In this respect, though, we did not observe a consistent difference between the two groups. All participants alternately used both ways to navigate: directly in the detail images, or with the marker boxes in the overview images or heat maps. However, the art historians preferred the method of navigating directly in the detail images, presumably because this was the usual way of working with common photo viewers.

The analysis tools can only give hints about possible deviations. Some participants were confused by the fact that some of the highlighted deviations were irrelevant while other, seemingly relevant, deviations were sometimes not

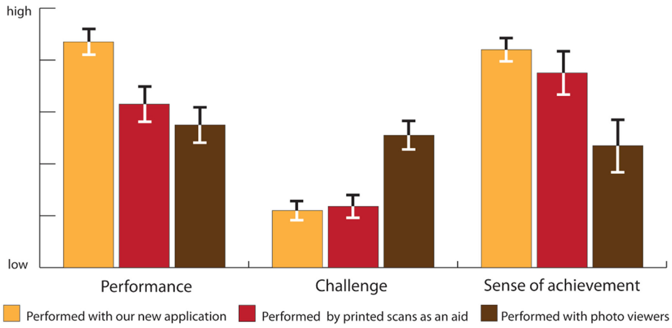

Fig. 10. Subjective assessment of the participants as a result of the questionnaire (mean values with 95% confidence intervals).

highlighted. However, in our estimation, they were quickly able to deal with the situation, which implies that the confusing factors only have an impact at the beginning of the experiment, and then only with regard to the processing time. In particular, none of the participants examined only the marked areas of the image pairs. We also observed that the availability of synchronized zooming sometimes encouraged participants to be highly focused on low-level details while ignoring aspects on the overview level; however, this mainly increased the processing time.

Almost all participants mentioned that the manual synchronization of images with the traditional photo viewer was exhausting and disruptive. At the same time, they did not mention any difficulties with regard to working with our system, and no one questioned the correctness of the synchronization, especially as far as local shifts were concerned.

5.2.3 Questionnaire Results

The evaluation of the questionnaire completed at the end of the study showed that the participants subjectively rated the new approach as more effective and comfortable than working with printed scans and with the conventional method (see Fig. 10). The evaluation of the questionnaire also showed that the participants did not have any major problems in handling the user interface of our system. Usually by around the fourth image pair, they had completely mastered the controls and navigation, and there were no more difficulties concerning usability.

5.3 Case Studies

To illustrate the practical benefits of VeCHArt, we again take a look at the analytical challenges that were introduced in Section 5.1. In the first example (Fig. 6), it was necessary to find restored lines and to estimate the degree of wear of the plate based on a few single lines. The participants of our user study, who examined this image pair by means of the printouts, did not recognize any of the restored regions. These regions were detected only with the help of zoomed-in digital representations. With the automatic analysis of VeCHArt, the revised areas were already recognized and highlighted in the corresponding heat map (top left in Fig. 6). In this case, VeCHArt did not only make it easier to compare the two historical prints of the first example; it also allowed user insights that were not obtained when comparing the prints only with the help of printouts. In addition, the automatic analysis provided clues to important features, and thus made the investigation faster and more reliable.

In the second example, both prints were made with the same plate (Fig. 7). Four significant changes were made with considerable effort. The first two differences are conspicuous and were also recognized by the participants who worked with the help of printouts; but only with the help of zooming does it become apparent that elements were deliberately removed, rather than having been worn out by use, having disappeared due to damage, or having been added later. The third difference, the differences in the hair, was recognized with all three methods of investigation. However, the think-aloud results revealed that with the prints or photo viewer, the difference was often misinterpreted as a difference in the background, which was not the case for VeCHart. The fourth difference was only detected by a small number of participants and only with VeCHart, where the area in question was highlighted in one of the heat maps. In summary, the differences could also be detected with the help of printouts. An accurate interpretation, however, was not possible by these means alone. Moreover, in this example it was only VeCHArt which enabled the user to undertake a conclusive and precise evaluation that led to significant findings about this series of prints.

6 Conclusion and Future Work

Comparing copies or different states of a print is a challenging task for art historians. Only rarely are two versions of a print in the same museum's collection and can be compared by scrutinizing the originals. The usual procedure of opening two images in a photo viewer simultaneously does not seem to be efficient and leaves room for improvement. We have therefore developed an approach in the form of a highly interactive visual analytics system that allows users to compare and analyze multiple engravings in a short period of time. On the one hand, this can enhance the art historian's work of filing the different states of a print with a great degree of precision. On the other hand, understanding how a copyist copied the original enables the art historian to get a deeper understanding of the print culture at hand. Different patterns of deviation will not only provide clues about the techniques of copying (freehand copying, working with a grid or a tracing aid), but also about the way the copyist perceived and interpreted the original. Finally, an art historian can get a fuller understanding of concepts of authorship and the status of prints in the 15th century.

The synchronization method of VeCHArt is new. There is still no method like VeCHArt that performs the side- by-side comparison of engraving details completely automatically. The evaluation in our study showed that the participants rated the new approach as more effective and comfortable than working with printed scans or the conventional method. The evaluation of the questionnaire also showed that the participants did not have any major problems in handling the user interface of our system. The results of the evaluation also demonstrate that non-experts can achieve performances comparable to those of experts when using our system. This suggests that our approach could be used to utilize the support of laypeople and volunteers for time-consuming tasks that formerly tied up resources of art history professionals. Furthermore, the evaluation confirms a significant performance gain over the standard procedure, and the line detection and matching methods make for automatic analyses that provide clues about important features, thus making the investigation faster and more reliable, and, additionally, holding the potential for fully automatic image comparisons; these could be of interest in the consideration of the millions of historical graphics not yet explored. Based on our case studies, we also illustrated the practical relevance of our new approach for art historical research. These case studies demonstrate that a method based exclusively on printouts (or originals) is not enough to clarify all relevant questions, and that, compared to previous methods, our new approach enables a more reliable, efficient, and comfortable evaluation within the workflow of art historians.

However, our evaluation also revealed room for improvement. The easy and comfortable handling of the synchronized zooming sometimes caused some participants to go into too much detail. In order to avoid this drawback, the time a user examines certain image areas should not only be displayed as a heat map, but a warning should additionally be given whenever a user moves too far from the actual task by going too much into detail. The data collected with our user study could be used to determine relationships between relevant image differences and differences in line patterns, which in turn helps improve the automatic analysis of image differences. We are furthermore planning to expand our system so that groups of researchers can work collaboratively on artwork analysis. To this end, corresponding methods have to be developed that enable users to store, classify, edit, and comprehensively visualize image differences that are found both automatically and by teams of users.

7 List of Illustrations

Fig. 1 Martin Schongauer: *The Dormition of the Virgin*, 2nd half of the 15th century, Engraving, 255 x 168 mm, Met Museum, New York, Museum number: 1984.1201.3; Monogrammist IE after Martin Schongauer: *The Dormition of the Virgin*, 1480–1500, Engraving, 249 x 165 mm, British Museum, London, Museum number: 1845,0809.331

Fig. 4 Martin Schongauer: *The Virgin of the Annunciation*, 1470–1480, Engraving, 169 x 117 mm, British Museum, London, Museum number: E,1.77

Fig. 3 Monogrammist IE after Martin Schongauer: St Bartholomew (reverse), 1480–1500, Engraving, 90 x 54 mm, British Museum, London, Museum number: 1926,0713.22.; Israhel van Meckenem after Martin Schongauer: St Bartholomew (reverse), 1470–1480, Engraving, 89 x 60 mm, British Museum, London, Museum number: 1845,0809.339

Fig. 6 Israhel van Meckenem after Monogrammist FVB after Martin Schongauer: *The Tribulations of St Antony*, 1475–1500, Engraving, 284 x 222 mm, British Museum, London, Museum number: 1845,0809.343; Monogrammist FVB after Martin Schongauer: *The Tribulations of St Antony*, 1475–1500, Engraving, 290 x 222 mm, British Museum, London, Museum number: 1854,0614.167

Fig. 7 Martin Schongauer: *Virgin and Child with a parrot*, 1470–1473, Engraving, 159 x 115 mm, British Museum, London, Museum number: E,1.70; Martin Schongauer: *Virgin and Child with a parrot*, 1470–1473, Engraving, 158 x 110 mm, British Museum, London, Museum number: 1895,0915.256

All scanned images are from the image database EasyDB of the University of Stuttgart used with permission.

References

[1] D. Landau and P. W. Parshall, *The Renaissance Print: 1470–1550*. New Haven, CT, USA: Yale Univ. Press, 1994.

[2] L. Schmitt, *Martin Schongauer: Materialien und Anregungen zur Erforschung früher Druckgraphik*. VDG Weimar, 2004.

[3] T. Falk and T. Hirthe, *Martin Schongauer Das Kupferstichwerk. Ausstellung zum 500. Todesjahr*. München, Staatliche Graphische Sammlung, 1991.

[4] H. Krohm and J. Nicolaisen, *Martin Schongauer - Druckgraphik im Berliner Kupferstichkabinett*. Gebruder Mann Verlag, 1997.

[5] L. Schmitt, *Hollsteins German Engravings, Etchings, and Woodcuts 1400–1700; volume XLIX: Ludwig Schongauer to Martin Schongauer*, N. Stogdon, Ed. Harrisonburg, VA, USA: Vision Publisher, 1999.

[6] G. Seelig, *The New Hollstein German Engravings, Etchings, and Woodcuts 1400–1700*, G. Bartrum, Ed. Ouderkerk aan den IJssel, Netherlands: Sound and Vision Publishers, 2001.

[7] J. P. F. Kok and M. Leesberg, *The New Hollstein German Engravings, Etchings, and Woodcuts 1400–1700*, G. Luijten, Ed. Ouderkerk aan den IJssel, Netherlands: Sound and Vision Publisher, 2000.

[8] U. Birgit, A. Tacke, M. Herzog, and S. Heudecker, *Fälschung-Plagiat-Kopie*. Michael Imhof Verlag, 2014.

[9] W. Augustyn and U. Söding, *Original - Kopie - Zitat. Kunstwerke des Mittelalters und der Frühen Neuzeit: Wege der Aneignung - Formen der Überlieferung*. Passau, Germany: Dietmar Klinger Verlag, 2010.

[10] L. Pon, *Raphael, Dürer, and Marcantonio Raimondi: Copying and the Italian Renaissance Print*. New Haven, CT, USA: Yale University Press, 2004.

[11] L. Manovich, "Data science and digital art history," *Int. J. Digit. Art History*, no. 1, 2015.

[12] L. Manovich and A. Tifentale, "Selfiecity: Exploring photography and self-fashioning in social media," *Postdigital Aesthetics: Art, Comput. Des.*, pp. 109–122, 2015.

[13] H. Pflüger and T. Ertl, "Sifting through visual arts collections," *Comput. Graph.*, vol. 57, pp. 127–138, 2016.

[14] Y. Ke, R. Sukthankar, L. Huston, Y. Ke, and R. Sukthankar, "Efficient near-duplicate detection and sub-image retrieval," *ACM Multimedia*, vol. 4, no. 1, p. 5, 2004.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 346–361.

[16] M. Burch, U. Kloos, D. Junger, I. Hagen, R. Horst, L. Brand, A. Müller, and B. Weinert, "artVIS: An interactive visualization for painting collections," in *Proc. 10th Int. Symp. Visual Inf. Commun. Interaction*, 2017, pp. 109–112.

[17] M. Gleicher, "Considerations for visualizing comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 413–423, Jan. 2018.

[18] L. Crissaff, L. W. Ruby, S. Deutch, R. L. DuBois, J. D. Fekete, J. Freire, and C. Silva, "Aries: Enabling visual exploration and organization of art image collections," *IEEE Comput. Graph. Appl.*, vol. 38, no. 1, pp. 91–108, Jan. 2018.

[19] A. Cockburn, A. Karlson, and B. B. Bederson, "A review of overview+detail, zooming, and focus+context interfaces," *ACM Comput. Surveys*, vol. 41, no. 1, pp. 2:1–2:31, Jan. 2009.

[20] M. Gerl and T. Isenberg, "Interactive example-based hatching," *Comput. Graph.*, vol. 37, no. 1, pp. 65–80, 2013.

[21] E. Kalogerakis, D. Nowrouzezahrai, S. Breslav, and A. Hertzmann, "Learning hatching for pen-and-ink illustration of surfaces," *ACM Trans. Graph.*, vol. 31, no. 1, pp. 1:1–1:17, Feb. 2012.

[22] S. Shaheen, A. Rockwood, and B. Ghanem, "SAR: Stroke authorship recognition," *Comput. Graph. Forum*, vol. 35, no. 6, pp. 73–86, 2015.

[23] J. Schmidt, M. E. Groller, and S. Bruckner, "VAICo: Visual analysis for image comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2090–2099, Dec. 2013.

[24] E. Garces, A. Agarwala, A. Hertzmann, and D. Gutierrez, "Style-based exploration of illustration datasets," *Multimedia Tools Appl.*, vol. 76, no. 11, pp. 13 067–13 086, Jun. 2017.

[25] K. Lawonn, E. Trostmann, B. Preim, and K. Hildebrandt, "Visualization and extraction of carvings for heritage conservation," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 801–810, Jan. 2017.

[26] J. Schmidt, R. Preiner, T. Auzinger, M. Wimmer, M. E. Grller, and S. Bruckner, "YMCA your mesh comparison application," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, Oct. 2014, pp. 153–162.

[27] B. Fan, F. Wu, and Z. Hu, “Robust line matching through linepoint invariants,” *Pattern Recognit.*, vol. 45, no. 2, pp. 794–805, 2012.

[28] L. Wang, U. Neumann, and S. You, “Wide-baseline image matching using line signatures,” in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1311–1318.

[29] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in *Proc. Eur. Conf. Comput. Vis.*, May 2004, pp. 25–36.

[30] D. G. Stork, “Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature,” in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2009, pp. 9–24.

[31] R. Sablatnig, P. Kammerer, and E. Zolda, “Structural analysis of paintings based on brush strokes,” in *Proc. Sci. Detection Fakery Art*, 1998, vol. 3315, pp. 87–99.

[32] T. Melzer, P. Kammerer, and E. Zolda, “Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach,” in *Proc. 14th Int. Conf. Pattern Recognit. (Cat. No.98EX170)*, vol. 1, Aug. 1998, pp. 474–476.

[33] J. Canny, “A computational approach to edge detection,” in *Readings in Computer Vision*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 184–203.

[34] BSI, “Fingerabdruckerkennung,” Studie des Bundesamt für Sicherheit in der Informationstechnik, 2004.

[35] D. Ziou, “Line detection using an optimal IIR filter,” *Pattern Recognit.*, vol. 24, no. 6, pp. 465–478, 1991.

[36] H. Pflüger, B. Höferlin, M. Raschke, and T. Ertl, “Simulating fixations when looking at visual arts,” *ACM Trans. Appl. Percept.*, vol. 12, no. 3, pp. 9:1–9:20, Jun. 2015.

[37] M. A. A. Cox and T. F. Cox, *Handbook of Data Visualization*. Berlin, Germany: Springer, 2008.

[38] G. Papari and N. Petkov, “Edge and line oriented contour detection: State of the art,” *Image Vis. Comput.*, vol. 29, no. 2, pp. 79–103, 2011.

[39] L. Quackenbush, “A review of techniques for extracting linear features from imagery,” *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 12, pp. 1383–1392, 2004.

[40] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2010, pp. 2963–2970.

[41] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 1998, pp. 130–137.

[42] C. Akinlar and E. Chome, “PEL: A predictive edge linking algorithm,” *J. Visual Commun. Image Representation*, vol. 36, pp. 159–171, 2016.

[43] D. Guru, B. Shekar, and P. Nagabhushan, “A simple and robust line detection algorithm based on small eigenvalue analysis,” *Pattern Recognit. Lett.*, vol. 25, no. 1, pp. 1–13, 2004.

[44] C. Lopez-Molina, B. D. Baets, H. Bustince, J. Sanz, and E. Barrenechea, “Multiscale edge detection based on Gaussian smoothing and edge tracking,” *Knowl.-Based Syst.*, vol. 44, pp. 101–111, 2013.

[45] C. Li, X. Liu, and T.-T. Wong, “Deep extraction of manga structural lines,” *ACM Trans. Graph.*, vol. 36, no. 4, pp. 117:1–117:12, Jul. 2017.

Hermann Pflüger received the graduate degree in computer science (Diplom-Informatiker) from the University of Stuttgart, Germany, in 2012. He works as a research assistant with the Visualization and Interactive Systems Institute (VIS), University of Stuttgart. His main research interests include digital humanities, computer vision, and visual analytics. A specific focus of his research is on developing new concepts, algorithms, and visualization methods to explore, analyze, and describe visual art collections.

Dennis Thom received the PhD degree in computer science from the University of Stuttgart, in 2015. Since 2011, he has been employed there as an research assistant with the Visualization and Interactive Systems Institute. His research interests comprise the fields of visual analytics, text data mining, and social media.

Anna Christina Schütz received the PhD degree from the Leuphana University in Lüneburg with a dissertation on the engraver Daniel Chodowiecki: Das Scheitern des Charakters am Selbstbewusstsein des Bildes. Chodowiecki zeichnet Goethes Werther. She is now a postdoctoral research fellow with the Institute for Art History, University Stuttgart and works on projects on Early Modern prints and drawings.

Daniela Bohde is full professor of art history with the University of Stuttgart. She is a specialist in early modern art. She received several grants in Germany and the US, among them a Samuel H. Kress Senior Fellowship at the Center for Advanced Study in the Visual Arts (CASVA) at the National Gallery of Art, Washington DC, a fellowship at the Max-Planck-Institute in Florence, and a Heisenberg fellowship. She published widely on Early Modern German drawings and prints.

Thomas Ertl is a full professor of computer science with the Visualization and Interactive Systems Institute (VIS), University of Stuttgart and director of the Visualization Research Center, University of Stuttgart (VISUS). He is coauthor of more than 500 scientific publications. He served as ab editor-in-chief of IEEE TVCG and as chairman of the Eurographics Association. He received the Outstanding Technical Contribution Award of the Eurographics Association and the Technical Achievement Award of the IEEE Visualization and Graphics Technical Committee in 2006.