RECONSTRUCTING THE DRAWING PROCESS OF REPRODUCTIONS FROM MEDIEVAL IMAGES

Antonio Monroy[1], *Bernd Carqué*[1,2], *Björn Ommer*[1]

[1] Interdisciplinary Center for Scientific Computing, [2] Institute for European Art History
University of Heidelberg, Germany
{amonroy,bommer,bcarque}@iwr.uni-heidelberg.de

ABSTRACT

Based on 14th century paintings and their 18th century manual reproductions we reconstruct the temporal order of the sketching process and analyze how images were altered by artists. Therefore, we present a novel algorithm for decomposing images into groups of parts that were similarly transformed between images. Moreover, a method is proposed that orders all parts based on the relatedness of their transformation. Based on this ordering, a third dimension, which corresponds to the timeline of the drawing process, is added to 2D paintings.

1. INTRODUCTION

Many outstanding medieval images were manually copied hundreds of years later for purposes of reproduction. Copying in some cases worked by placing a thin, semi-opaque sheet of paper on the surface of the original, and sketching the contours. Slight modifications between the images can be observed. From the side of art history several questions arise from this fact: Are these transformations regular by their nature due to technical characteristics of the reproduction procedure or are they focused on specific regions or details due to the interpretation of the original by the artists or their patrons? Are we able to reconstruct the reproduction process based on these modifications? Can we observe some pattern that could lead to conclusions about the cultural context in which these images were reproduced? Based on the comparison of deformations between images from the 14th century and their 18th century manual reproductions we reconstruct and visualize the temporal order of the drawing process. The idea is based on the observation that parts of an image that were drawn at the same time (e.g. a contour) or in close succession (e.g. the eyes and head) should exhibit similar or identical transformations, whereas parts that significantly feature different transformations were either deliberately altered later on or geometrically errors accumulated during the drawing process, and therefore indicate that they were not drawn in close succession. We present an iterative algorithm that automatically decomposes the complex nonlinear transformation between both images into piecewise linear functions. By designing a measure that reflects the temporal order we address the challenge of art history to reconstruct

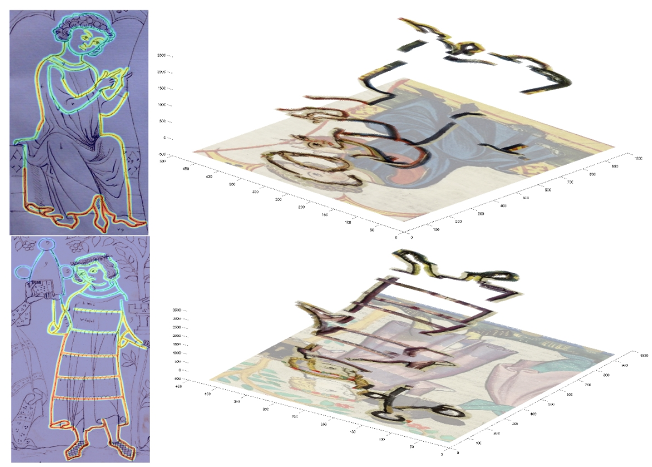

Fig. 1. Temporal ordering of the drawing process obtained by our algorithm. The ordering is encoded in the z-coordinate (right). The left column shows the ordering in color (blue comes first, red last).

the order of the drawing process. This problem is to the best of our knowledge, a novel problem that has not yet been approached. We present conceptual results on images coming from the Codex Manesse ([1]) illustrated between c. 1305 and c. 1340 in Zürich and their reproductions commissioned by Bodmer/Breitinger in 1746/1747 ([2]). The algorithm presented in this paper assists art historians in systematically detecting and quantifying visual transformations. Thus it amends and enhances the partial comparisons provided by traditional methods [3] and provides deeper insight into the reception history of the Codex Manesse, which is an outstanding source for a central topic of Medieval Studies in recent years: the visual interpretation of the Middle Ages in early modern and modern times [4]. Beyond that, the algorithm enhances research on several kinds of transmission processes in the visual arts [5]. As a final result, the approach provides the art historically most likely reconstruction of a drawing process, which was completed centuries ago .

Being a new problem setting in the literature, the work in

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

computer vision that comes closest is that on sparse motion segmentation. [6, 7] present a method for decomposing videos into similarly moving layers. The scene is firstly divided into a regular grid and an affine transformation is calculated for each block. Both methods estimate affine motion models for segments on a regular grid. Due to clutter and missing contours, accurate estimation of small and continuous deviations in transformations be estimated with this approach. Other state-of-the-art approaches [8, 9] need multiple frames to correctly separate the different motions present in the sequence. Finally GPCA [10] is a known general algebraic method that can also be used for segmenting subspace arrangements of motions.

2. APPROACH

Our goal is to localize and quantify the transformations between the medieval image and its modern reproduction and herewith model the temporal order in which images were reproduced. As a first step, we need to find point correspondences between pairs of images. We observe that calculating the optical flow between both images and clustering the resulting vector fields [6, 10] features only insufficient accuracy: localization along the contours is ambiguous (cf. aperture problem) since contours have been distorted (e.g. stretched), junctions are partly missing, and textures alongside the contours have not been reproduced. Moreover, contours in original and reproduction have feature significant variation of thickness which yields additional ambiguity. For this reason, art historians have provided us with landmark correspondences. We then propose an algorithm for estimating how the various image regions where transformed between original and reproduction. Finding the subtle differences in these transformations is a challenging task not only because the differences are comparable small but also because contour points are mostly collinear. In section 2.3 we then analyze the relations between different transformations and find a projection which models the temporal ordering.

2.1. Calculation of affine transformations

Let A be an original image and B be a reproduced version of the original. $\mathcal{X}^A := \{x_i^A\}_{i=1}^n$ are landmark points in image A and $\mathcal{X}^B := \{x_i^B\}_{i=1}^n$ are the corresponding points in B (given in homogeneous coordinates). The transformation

$$T(\vartheta_i^1) := \begin{pmatrix} \sigma_i^x \cos\alpha_i & -\sigma_i^y \sin\alpha_i & t_{x_i} \\ \sigma_i^x \sin\alpha_i & \sigma_i^y \cos\alpha_i & t_{y_i} \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

then describes the deformation each point x_i^A underwent during reproduction onto x_i^B. Obviously this transformation matrix cannot be computed locally (cf. [11]) using Levenberg-Marquard algorithm but it requires an extended neighborhood $\vartheta_i^1 \subset \mathcal{X}^A$ of point x_i^A. Obtaining a neighborhood of contour points that are not collinear is a challenging problem. Art historical analysis of the corpus has indicated that stretching, rotation, and translation are the transformations to be expected and so we can restrict (1) to the corresponding five free parameters. For the iterative algorithm that follows, the neighborhood ϑ_i^1 is initialized to contain x_i^A and its 10 nearest neighbors (10 is twice the number of degrees of freedom). The neighbors are those points $x_j^A \in \mathcal{X}^A$ with minimal distance $d_C(x_i^A, x_j^A)$, where d_C is a contour-based distance measure between landmark points (explained in next section). The choice of the neighborhood is crucial. Given a small neighborhood, the robustness of the transformation $T(\vartheta_i^1)$ suffers since fewer points render the parameter estimation susceptible to noise, and small rotations and scalings cannot robustly be measured locally. Large neighborhoods, however, average over many, potentially unrelated points and thus do not allow to identify local variation in the transformation, i.e., due to local deformations. To overcome this problem we propose an iterative procedure, which updates the neighborhood set and the transformation iteratively. Let $T_i^1 := T(\vartheta_i^1)$ be the initial transformation calculated according to (1). Then we can compute the set of points x_j^A in the image that are consistent with this transformation ($T(\vartheta_i^1)$ maps x_j^A to x_j^B with an error equal or less than ϵ).

$$C_i^k := \left\{ j \mid \frac{1}{|\vartheta_j^k|} \sum_{s: x_s^A \in \vartheta_j^k} \left\| T_i^k x_s^A - x_s^B \right\|_2 \le \epsilon \right\} \quad (2)$$

The iterative procedure starts with $k = 1$. Since errors accumulated continuously during reproduction, we cannot assume perfectly matching groups and thus we allow an error of ϵ. Setting it to be the 5-th percentile over the error $\|T(\vartheta_i^1)x_r^A - x_r^B\|$ of all $x_r^A \in \mathcal{X}_A^1$ yields a good trade-off between exactness of the transformation and the capacity to efficiently refine the transformation during the iteration process. Thereafter we calculate for each $j \in C_i^k$ the transformation $T(\vartheta_i^k \cup \vartheta_j^1)$ with the points $\vartheta_i^k \cup \vartheta_j^1$, i.e., we grow the group ϑ_i^k with the local neighborhood of point j. Let

$$E_{ij}^k := \left| \left\{ s \mid \left\| T(\vartheta_i^k \cup \vartheta_j^1) x_s^A - x_s^B \right\|_2 \le \epsilon,\, x_s^A \in \mathcal{X}_A^k \right\} \right| \quad (3)$$

represent the number of points that are consistent with $T(\vartheta_i^k \cup \vartheta_j^1)$, here we set $\mathcal{X}_A^1 := \mathcal{X}_A$. The refined transformation in the next iteration is updated using

$$\vartheta_i^{k+1} := \vartheta_i^k \cup \vartheta_{\underset{j}{\operatorname{argmax}} E_{ij}^k} \quad (4)$$

$$T_i^{k+1} = T(\vartheta_i^{k+1}) \quad (5)$$

$$\mathcal{X}_A^{k+1} := \mathcal{X}_A^k \setminus \vartheta_i^{k+1}. \quad (6)$$

This iterative update of transformations proceeds until C_i^{k+1} is empty, meaning that no additional points in the image are consistent with the transformation. The result of the algorithm are transformations $T_i := T_i^k$ for each point x_i^A that are globally consistent with other related parts in the image. It is important to realize that in our algorithm each point can be considered in the calculation of more than one transformation.

2.2. Contour distance between Points

In the last section affine transformations were calculated using the contour measure $d_C(x_i^A, x_j^A)$. In this section we de-

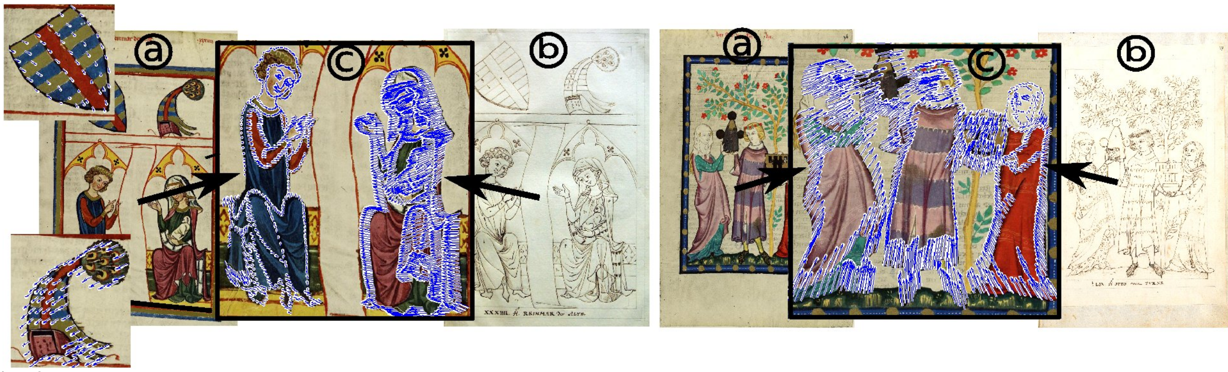

Fig. 2. a) 14th century original image. b) 18th century manual reproduction. c) Transformation using a single affine transformation.

scribe its construction. We recapitulate that the only information available is the landmark positions and their correspondences. No additional neighborhood information among the points is provided. A trivial approach would be to define $d_C(x_i^A, x_j^A)$ as the euclidean distance between points. In doing so, points belonging to different structures (like hair and eyes) can be closer than points on the same contour. This does not reflect the true nature of the continuous drawing process: contours are normally drawn as part of one painting stroke and therefore points on the same contour should be closer to another. To model this concept of neighborhood we construct a graph $G = (\mathcal{V}, \mathcal{E})$ by solving the Traveling Salesman Problem (TSP) using the pairwise distance matrix:

$$M(i,j) := \begin{cases} \|x_i^A - x_j^A\|_2 & : \quad \|x_i^A - x_j^A\|_2 < \delta \\ \infty & : \quad \text{else} \end{cases} , \qquad (7)$$

where δ is a user-given constant. The edges $\mathcal{E}$ indicate how landmark points in image A are connected to each other. Although the TSP is NP-hard, approximate solutions can be found in practice in a feasible time. Here we use the algorithm described in [12].

Given the neighborhood graph G, the distance d_C between points x_i^A and x_j^B, $i < j$ is then defined as:

$$d_C(x_i^A, x_j^A) := f(p), \qquad (8)$$

where $f(p) := \sum_i \|x_{i+1}^A - x_i^A\|_2$ and $p := (x_i^A, \cdots, x_j^A)$ is the shortest path connecting x_i^A, x_j^A obtained by the TSP.

2.3. Temporal Ordering

We aim at reconstructing and visualizing the temporal order in which the reproduction of images were drawn. This problem can equivalently be formulated as the visualization of how different parts in an image are related according to how they were transformed during the drawing process. This equivalence, as stated in the introduction, is based on the observation that parts of an image that were sketched at the same time or in close succession should exhibit similar or identical transformation. One way of measuring the dissimilarity between transformations T_i, T_j corresponding to the points x_i^A, x_j^A is to measure the average distortion that is made if T_j is applied to map point x_i^A instead of using its corresponding transformation T_i:

$$d_T(x_i^A, x_j^A) := \frac{1}{2}\left(\|T_j x_i^A - T_i x_i^A\| + \|T_j x_j^A - T_i x_j^A\|\right) \qquad (9)$$

Since disconnected parts that are far away from another (e.g. eyes and feet) are not likely to have been reproduced consecutively, we additionally need to take into account the contour distance $d_C(x_i^A, x_j^A)$ (sect. 2.1) between points. Thus we obtain the distance measure

$$\Delta_{ij} = \beta_1^{-1} d_T(x_i^A, x_j^A) + \lambda \beta_2^{-1} d_C(x_i^A, x_j^A), \qquad (10)$$

where $\beta_1 = \max_{ij} d_T(x_i^A, x_j^A)$ and $\beta_2 = \max_{ij} d_C(x_i^A, x_j^A)$ ($\lambda = 0.3$ has been determined experimentally and is kept constant in all our experiments). Δ_{ij} measures the dissimilarity between contour points and thus indicates how close the reproduction of point j followed that of i. One natural way to encode the temporal ordering for all points x_i^A is to add a third dimension z_i to the locations x_i^A. The z-coordinate, which represents the temporal ordering, is obtained by projecting Δ_{ij} using a distance preserving embedding in 1D space of the z_i. By applying multidimensional scaling [13] on (10) we obtain a solution for the z_i by solving

$$(z_1, z_2, ...) = \underset{z_1, z_2, \cdots}{\operatorname{argmin}} \sum_{i<j} (\|z_i - z_j\| - \Delta_{ij})^2 .$$

Fig. 3. Inconsistency between parts. A single affine transformation is insufficient to model the distortion of the complete figure, since individual parts have been transformed differently by the artist. Each transformation is calculated using our alg. and ϑ_i^1 being the red segment.

3. EXPERIMENTAL FINDINGS

Due to its importance, we analyze images from the *Codex Manesse* ([1]), illustrated between c. 1305 and c. 1340. Reproductions of these images, commissioned in $1746/47$ (by J.J. Bodmer and J.J. Breitinger), are registered against the originals. Our analysis reveals significant distortions in the

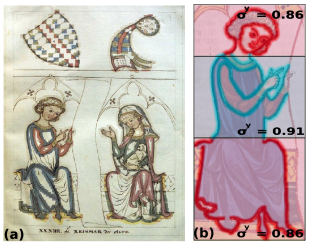

Fig. 4. (a) Transformation of an image using the piecewise affine transformations computed with our algorithm. (b) Average of σ_i^y for different image regions shows that the torso is stretched.

reproductions. Different regions of an image feature different transformations as can be seen from Fig. 2 and 3 where a single transformation does not suffice to bring original and reproduction into correspondence. Only if the shape is decomposed into several affine transformations we achieve a consistent registration (we use ~ 500 points for each shape). An example of this is shown in Fig 4 (a) where both images are brought into alignment by means of piecewise affine transformations. We then average the scaling in y-direction, σ_i^y, from (1) in each of three regions. This analysis shows that the proportions of particular parts of objects were deliberately modified. Fig. 4 (b) visualizes one example, where the torso was stretched in length by 5.5% compared to head and feet. Such a remarkable finding confirms the art historical hypothesis that proportions of human figures were altered to reflect the change of aesthetic preferences.

In Fig. 1 we show the reconstruction of the temporal ordering of the drawing process. It can be observed that areas with the most complex visual structures (e.g. facial features, drapery, hands) were drawn in close succession, whereas the connections between such details exhibit continuous variations over time. A deeper art historical analysis would however exceed the length restriction of this paper.

We note, that a hierarchical clustering of the temporal dimension (z-coordinate) of the landmark points would be equivalent to finding groups of similar affine transformations. Fig 5 (c) - (f) shows some examples of the resulting groupings (using Ward's method). A decomposition into semantically meaningful groups (e.g. head, hands, feet, helmet) is conform to the manner in which the traditional reproduction process worked [5]. This corroborates our method and provide us with deeper insights about the images.

Since motion segmentation algorithms group image regions based on local distortions, we also compare our grouping with the approaches [10, 6] in Fig 5 (a) and (b) (we use the same number of clusters for comparison and due to space restrictions limit ourselves to show one example). The groups from [10, 6] are clearly deficient since the subtle differences in transformations can only be measured based on an accurate neighborhood assignment, as the one proposed in Sect. 2.1.

4. CONCLUSION

Based on the novel comparison of deformations between images from the 14th century and their 18th century manual reproductions we have presented a method that allows us to reconstruct and visualize the temporal order in which reproductions where sketched. For this, we presented an algorithm that discovers local neighborhoods of related points, accurately estimates their transformations, and projects their similarities onto a temporal ordering. We investigated our method on images coming from the Codex Manesse, discovering new insights about the image reproduction process. Our approach clearly outperforms the grouping strategy from [6, 10]. Moreover, our method supports art historians in their analysis of image reproductions. The approach has, for instance, enabled a quantitative investigation of how local object proportions were altered by artists to meet the fashion of their time.

Fig. 5. (a)-(b) Motion segmentation between images using [6] and [10]. (c)-(f) Groupings of timeline using our method. (Best viewed in color)

5. REFERENCES

[1] UB Heidelberg, "Codex Manesse, Cod.Pal.Germ. 848," .

[2] "Ms. G 46. Zentralbibliothek Zürich," .

[3] B. Carqué, "Zur manuellen Reproduktion der Miniaturen," *Der Codex Manesse und die Entdeckung der Liebe*, 2010.

[4] B. Carqué, D. Mondini, and M. Noell, Eds., *Materielle Relikte des Mittelalters in bildlichen Darstellungen der Neuzeit und Moderne*, 2006.

[5] B. Bussmann and A. Hausmann, Eds., *Übertragungen. Formen und Konzepte von Reproduktion in Mittelalter und Früher Neuzeit*, 2005.

[6] J Wang and E. Adelson, "Representing moving images with layers," *IEEE Trans. on IP*, vol. 3, no. 5, 1994.

[7] G. Borshukov, G Bozdagi, Y. Altunbasak, and A. Tekalp, "Motion segmentation by multi-stage affine classification," *IEEE Trans. on IP*, vol. 6, no. 11, 1997.

[8] F. Lauer and C. Schnorr, "Spectral clustering of linear subspaces for motion segmentation," *ICCV*, 2009.

[9] E. Elhamifar and R Vidal, "Sparse subspace clustering," *CVPR*, 2009.

[10] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (gpca)," *CVPR*, 2003.

[11] B. Ommer, T. Mader, and J.M. Buhmann, "Seeing the objects behind the dots: Recognition in videos from a moving camera," *IJCV*, vol. 83, no. 1, 2009.

[12] C. Walshaw, "A multilevel lin-kernighan-helsgaun algorithm for the travelling salesman problem," Tech. Rep., SE10 9LS, 2001.

[13] I. Borg and P. Groenen, *Modern Multidimensional Scaling: theory and applications*, Springer-Verlag New York, 2005.