Image and Vision Computing 32 (2014) 414–423

Contents lists available at ScienceDirect

Image and Vision Computing

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

Morphological analysis for investigating artistic images[☆]

Antonio Monroy [a], Peter Bell [a,b], Björn Ommer [a,*]

[a] HCI & Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany
[b] Institute of European Art History, University of Heidelberg, Germany

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form 5 February 2014
Accepted 2 April 2014
Available online 13 April 2014

Keywords:
Shape analysis
Registration
Grouping
Model selection
Stability analysis
Cultural heritage

ABSTRACT

This paper describes an approach for automatically analyzing the alterations of an original artwork during its reproduction. The overall deformation of the artwork is modelled by a piecewise linear model, where regions of the artwork that feature similar alterations are automatically inferred and assigned to the different model components. Model complexity, that is, the required number of affine components required for registration, is automatically estimated using a statistical stability analysis. The main challenge is to simultaneously solve three tasks: (i) inferring the correspondences between both shapes, (ii) identifying the groups in the image that share the same transformation, and (iii) estimating the transformation of these groups. Our approach is tested on controlled scenarios as well as on real historical images.

1. Introduction

Although some stylistic movements in art like impressionism or pointillism define themselves by color, shape has been the predominant way to perceive an artwork. The theory about the primacy of shape can be traced back to Giorgio Vasari (1511–1574) who propagated the line drawing as the predominant technique of all visual arts. His use of the term *disegno* (conceptual design) can be read as the assignment of ideas to shapes. This “shaped idea” is represented through shapes in preparatory drawings, in the artwork itself as well as in drawn reproductions. Based on this observation, changes in shape between artworks and their reproductions or preparatory drawings can be associated with changes in ideas and concepts that reveal artistic choices and stylistic variations. Thus, the analysis of these changes helps art historians to understand the impact of historical influences in the creation and reproduction of art. However, in many cases, these alterations between shapes are very subtle and thus, it becomes extremely difficult, even for experts, to determine the nature and extent of the deformations suffered by different parts within an artwork. The automatic solution of such a shape analysis poses an ambitious computer vision task, and its solution is the focus of the present paper. The nature of the artwork deformations analyzed in this work arises either due to deliberate alterations or due to geometric errors accumulated during the drawing process. For instance, a typical example for a deliberate alteration between a preparatory drawing and the finished work is a subtle conceptual change that induces small alterations in the relative position of extremities in a human pose. These conceptual changes may have personal, cultural, or historical reasons, and thus, it is of interest for art historians to recognize the parts that feature similar transformations and to determine to which extent these parts differ from other regions in the image.

1.1. Piecewise transformation model

The system presented in this contribution addresses the description of an overall non-linear deformation as featured between an original artwork and its reproduction, and at the same time, it gives insights about the local structure of the shape deformation. Shape transformation models within computer vision can be classified into linear and non-linear models. Since *global* linear models cannot be used for describing complex shape changes due to their limited description power, a common choice for describing non-linear changes has been the usage of splines like the TPS [5]. However, besides requiring the estimation of a high number of parameters (proportional to the number of points in the shape) to determine the model, its complexity is regularized by a single manually set parameter for the entire shape. The global nature of this parameter makes it impossible for the model to locally adapt its complexity according to the shape deformation. Therefore, the present paper presents a piecewise linear registration model that adapts the complexity of each component according to the

[☆] This paper has been recommended for acceptance by Thomas Pock.
[*] Corresponding author at: Speyerer Straße 6, 69115 Heidelberg, Germany. Tel.: +49 6221 547853.
E-mail address: ommer@uni-heidelberg.de (B. Ommer).

http://dx.doi.org/10.1016/j.imavis.2014.04.004
0262-8856/© 2014 Elsevier B.V. All rights reserved.

shape deformation in the underlying region. Moreover, the assignment of regions in the shape to different model components induces a clustering which is used in turn to visualize the structure and geometry of the deformation introduced by the artist during the reproduction procedure.

1.2. Automatic complexity estimation

However, a challenge of using piecewise linear models is to automatically determine the number of components required for registration. In the absence of prior knowledge about the shape deformation, the answer to this question represents an important part of the analysis. Nonetheless, an indispensable requirement for selecting the number of components is the robustness of the registration solution. The present paper considers this robustness or stability from a statistical point of view. A stable registration solution for a given number of components is understood as a solution that is reproducible on different subsampled versions of the shape and does not too sensitively dependent on the sample set at hand. Thus, the "correct" number of transformations is defined as the number that yields the most stable solution capable of handling the trade-off between a too rigid transformation and an overparameterization of the transformation model.

1.3. Historical analysis of image reproductions

Finally, we utilize the proposed approach to analyze prominent reproductions from different periods of art history. At first, images coming from the Codex Manesse illustrated between ca. 1305 and ca. 1340 in Zürich and their reproductions commissioned by Bodmer/Breitinger in 1746/1747 are considered. This image collection is important for art history since the Codex Manesse is the single most comprehensive source of Middle High German Minnesang poetry [3] and represents an outstanding source for understanding the visual interpretation of the Middle Ages in early modern and modern times. Whereas the tracings from book illustrations like the reproductions of the Codex Manesse exhibit only slight changes, the differences between a drawing and a mural painting are obviously greater. Therefore, we also analyze parts of Michelangelo's ceiling fresco in the Sistine Chapel (1508–1512) with sketches, which were made in the artists surroundings, probably after Michelangelo's own preparatory drawings or by Dutch artists after the original artwork had been completed.

2. Related work

In the study of Monroy et al. [18], the temporal drawing process of how an image is reproduced was analyzed. It was assumed that parts drawn in closed succession in the reproduction exhibit similar deformations between the images. A limitation is the manual location and matching of landmark points. Furthermore, the approach lacks a unified model since two different clustering algorithms were applied for estimating the parameters of local affine transformations assuming perfect point correspondences, thus making this procedure very susceptible to noise. The present paper formulates a single optimization problem where affine transformations are estimated and points are grouped within the same procedure.

In the work of Monroy et al. [17], we proposed to solve for the groups and affine transformations by formulating a single optimization problem that was solved using deterministic annealing (DA). However, at the beginning of the optimization procedure, shape points were assigned with almost the same probability to the initial affine transformations. Thus, after updating the transformations, all affine parameters became equal and the algorithm got trapped in a local minimum. A further limitation, which is also shared by Monroy et al. [18], was the inclusion of a Euclidean distance term in the energy function to force the compactness of the groups. Thus, a bias in the solution was introduced since groups were clustered due to proximity and not depending on the registration quality. In the study of Monroy et al. [17], we also assumed for simplicity to have fix point correspondences between shapes, and their calculation was not related to the main optimization procedure. The current approach substitutes the DA technique by a linear program (LP) formulation for assigning points to groups. Moreover, we eliminate the Euclidean distance term in the energy function, and groups are found only by the accuracy of registration. In addition, our method also optimizes point correspondences between shapes along with the groups and the transformation within the same procedure.

In the field of sparse motion segmentation for instance, Wang and Adelson [25] presented a method for decomposing videos into similarly moving layers. This method estimates affine motion models for segments on a regular grid. Due to clutter and missing contours, the accurate estimation of small and continuous deviations in transformations cannot be estimated with this approach. In the study of Delong et al. [8], a regularized energy function was minimized with Graph-Cuts ([2]), which also included a pairwise regularization and thus a bias in the result. This regularization led in practice to poorer registration quality since parts in the shape belonging to different model components were mixed. Furthermore, Komodakis et al. [12] presented an LP formulation of a central clustering in which the number of clusters is determined indirectly by a hard to determine penalty term for each data point. Lazic et al. [14] also indirectly determined the number of clusters through the weighting of the different randomly subsampled linear subspaces. Normally, (rigid) motion segmentation can be seen as an application of the more general task of subspace segmentation [14,26]. This latter task commonly assumes that the data points lie on several distinct *linear subspaces* [9,26,7,24,11]. However, the linearity assumption does not hold in our setting: Whereas shape points lie in a 2D vector space, each of the shape parts that were similarly altered by the artist are represented through elements of the affine group. Therefore, the task consists not only of clustering points that define a linear subspace, but three tasks need to be solved jointly: the correspondence between both shapes, the groups in the image that share the same transformation, and the estimation of the transformations of those groups.

In the field of computer graphics, Sýkora et al. [21] embedded each shape in a lattice consisting of several connected squares and registered them by estimating a rigid transformation for every square. Since the registration is only on the level of rigid squares, a grouping into flexibly shaped regions with related modifications is not part of this contribution. Furthermore, Sýkora et al. [21] are not able to handle deformations that do not preserve local rigidity (e.g., scaling or shear), and it requires a significant overlap between shapes for registration. Additionally, in our setting, background clutter needs to be handled, whereas the method of Sýkora et al. [21] is only applied to cartoons without any clutter. Another interesting related work is by Commowick et al. [6], which presented a piecewise affine regularization method for medical image registration. The drawback of this method is that the affine-registered areas need to be estimated manually by the user. Related to piecewise affine registration, Hongsheng et al. [10] recently introduced a matching algorithm based on affine transformations calculated on a triangulation of the shape. In this case, to match articulated objects, it is required to manually select the groups and their articulation in order to match the scene images. Two different works that are related to estimating transformations between artworks are by Chang and Stork [4] and Usami et al. [22]. While Chang and Stork [4] tried to ensure consistent perspective in art images, Usami et al. [22] aimed to dewarp image reflections shown in convex mirrors within very specific paintings. Common non-linear registration algorithms like Chui and Rangarajan [5] or Myronenko and Song [19] are also not suited to the purpose of the present task. Whereas Chui and Rangarajan [5] used a thin plate spline (TPS) to model the transformation, Myronenko and Song [19] estimated a displacement vector for each point in the shape. In both cases, these models introduce artifacts in the registration as observed by Monroy et al. [17], which is undesirable for art comparison.

3. Approach

In this paper, shapes are represented through landmark points (given in homogeneous coordinates), which are regularly sampled along extracted contours of the corresponding image in an automatic manner (see Section 4.4 for more details). The shape of the original artwork is referred to with the matrix $X \in \mathbb{R}^{m \times 3}$ and the shape of the reproduced artwork with the matrix $Y \in \mathbb{R}^{n \times 3}$.

3.1. Problem statement

The main challenge consists of simultaneously solving three tasks. First, the correspondences between both shapes have to be inferred. Second, the groups in the image that share the same transformation need to be found. Finally, the transformations of those groups and thus the overall deformation model need to be estimated. The missing groups correspond to image regions that are reproduced similarly by the artist. Therefore, each of these groups is modeled through an affine transformation capable of transforming the group from the reproduction into the original painting. The advantage of using a piecewise-affine transformation model is that it allows to describe a non-linear transformation in a more parsimonious manner; that is, less parameters are required for describing the overall transformation. At the same time, the components in the model associated with different regions in the shape give insights about the structure and geometry of the artistic deformation.

Formally, the problem consists of estimating a binary data assignment matrix $C \in \mathbb{B}^{n \times m}$ of n points belonging to the first shape to m points in the second shape. At the same time, a binary matrix $M \in \mathbb{B}^{n \times k}$ of n points to k groups needs to be calculated together with different affine transformations $T^{\nu} \in \mathbb{R}^{3 \times 3} (\nu = 1,\ldots,k)$ for each group. Thus, the overall registration error made by a solution $(M,C,T^1,\cdots,T^k)$ can be written as

$$E_{\text{reg}} := \sum_{i,\nu=1}^{n,k} M_{\nu i} \left(\underbrace{\sum_{j=1}^{m} C_{ij} \|x_j - T^{\nu} y_i\|^2}_{=: r_{\nu i}} \right). \tag{1}$$

An important observation is that although the global deformation between both artworks is expected to be non-linear, regions between both images that were copied without any or little alteration by the artist are transformed homogeneously, and therefore, these parts can be described using a single affine transformation. Thus, for any two points y_i and y_j within such an affine-transformed shape part together with their respective correspondent points x_a and x_b, the distortion between the vector from y_i to y_j and the vector from x_a to x_b is expected to be small (and minimal in the presence of a rigid transformation). Similar to Berg et al. [1], this distortion can be measured by

$$d\left(y_i, y_j; x_a, x_b\right) := \gamma d_a\left(y_i, y_j; x_a, x_b\right) + (1-\gamma) d_l\left(y_i, y_j; x_a, x_b\right), \tag{2}$$

$$d_a\left(y_i, y_j; x_a, x_b\right) := \left(\frac{\alpha_d}{\left\| s_{ij} \right\|} + \beta_d \left| arcsin\left(\frac{\hat{s}_{ab} \times s_{ij}}{\|\hat{s}_{ab}\| \left\| s_{ij} \right\|} \right) \right| \right), \tag{3}$$

$$d_l\left(y_i, y_j; x_a, x_b\right) := \frac{\left\| s_{ij} \right\| - \|\hat{s}_{ab}\|}{\left(\left\| s_{ij} \right\| + \sigma_d \right)}; \tag{4}$$

$$s_{ij} := y_i - y_j, \quad \hat{s}_{ab} := x_a - x_b. \tag{5}$$

While the first term $d_a(y_i, y_j; x_a, x_b)$ penalizes the change in direction, the second term $d_l(y_i, y_j; x_a, x_b)$ penalizes the change of length between two pairs of points in both shapes. The constants $\alpha_d = \beta_d = \sigma_d = 0.5$ allow more flexibility for nearby points, and the constant $\gamma = 0.3$ weighs the angle distortion term against the length distortion term. All these parameters have been kept fixed in all experiments, thus showing the robustness of the solution with this set of parameters, despite large variations in the input. We use this measure to further enforce the matching consistency between both shapes, and thus the energy term Eq. (1) to be minimized is extended to

$$\begin{aligned} \underset{M,T^{\nu},C}{min}\, E_{\text{tot}} := & \sum_{i,\nu=1}^{n,k} M_{\nu i} \left(\sum_{j=1}^{m} C_{ij} \left\| x_j - T^{\nu} y_i \right\|^2 \right) \\ & + \underbrace{\sum_{\nu=1}^{k} \sum_{i,j=1}^{n} \sum_{a,b=1}^{m} M_{\nu i} M_{\nu j} C_{ia} C_{jb} d\left(T^{\nu} y_i, T^{\nu} y_j; x_a, x_b\right)}_{=: E_{\text{quad}}} \end{aligned} \tag{6}$$

$$s.t. \sum_{\nu=1}^{k} M_{\nu i} = 1 \quad (\forall i = 1, \cdots, n) \tag{7}$$

$$\sum_{i=1}^{n} C_{ij} = 1 \quad (\forall j = 1, \cdots, n), \tag{8}$$

$$C_{ij} \in \{0, 1\}, \quad M_{\nu i} \in \{0, 1\} \tag{9}$$

where k is the complexity of the piecewise model (i.e., the number of affine transformations desired for registration). This parameter will be set automatically based on the stability analysis described in Section 3.3. While the constraint Eq. (7) forces each point to be assigned to a single group, the constraint Eq. (8) ensures a many-to-one matching between both point sets yielding robustness in cases of missing points. Important to remark is that whereas Berg et al. [1] minimized the pairwise distortions for all points in the shape together, our model minimizes the pairwise distortions within each of the groups defined through the matrix M.

3.2. Optimization strategy

The general setting of jointly solving for M, C, and T^{ν} is hard. This is reflected in the above problem formulation (Eq. (6)), where solving for the matrix C exactly is already NP-hard [1]. A practical solution to minimize the above energy is to assume an alternating procedure. Departing from an initial solution, the above energy function is reduced by first calculating the matrix M and the transformations T^{ν} (assuming the matrix C is given) and solving for the matrix C (assuming M, T^{ν} are given) thereafter. This procedure is iterated until the matrix C and M do not change.

3.2.1. Problem formulation using a superset of affine transformations

Estimating the matrix M and the different affine transformations T^{ν} (given the matrix C) are closely interrelated problems, and their solution poses a challenging issue. Because T^{ν} $(\nu = 1,\ldots,k)$ can only be estimated when the assignment of points to k groups (given by the matrix M) is known, each of the groups ν is defined by the fact that all points within it can be registered using a single affine transformation T^{ν}. Thus, the rationale of our previous work [17] was to approach this problem by first proposing a *single* initial clustering (i.e., a matrix M) based on the Euclidean proximity of the shape points. Thereafter, based on this matrix, the estimation of the affine transformations

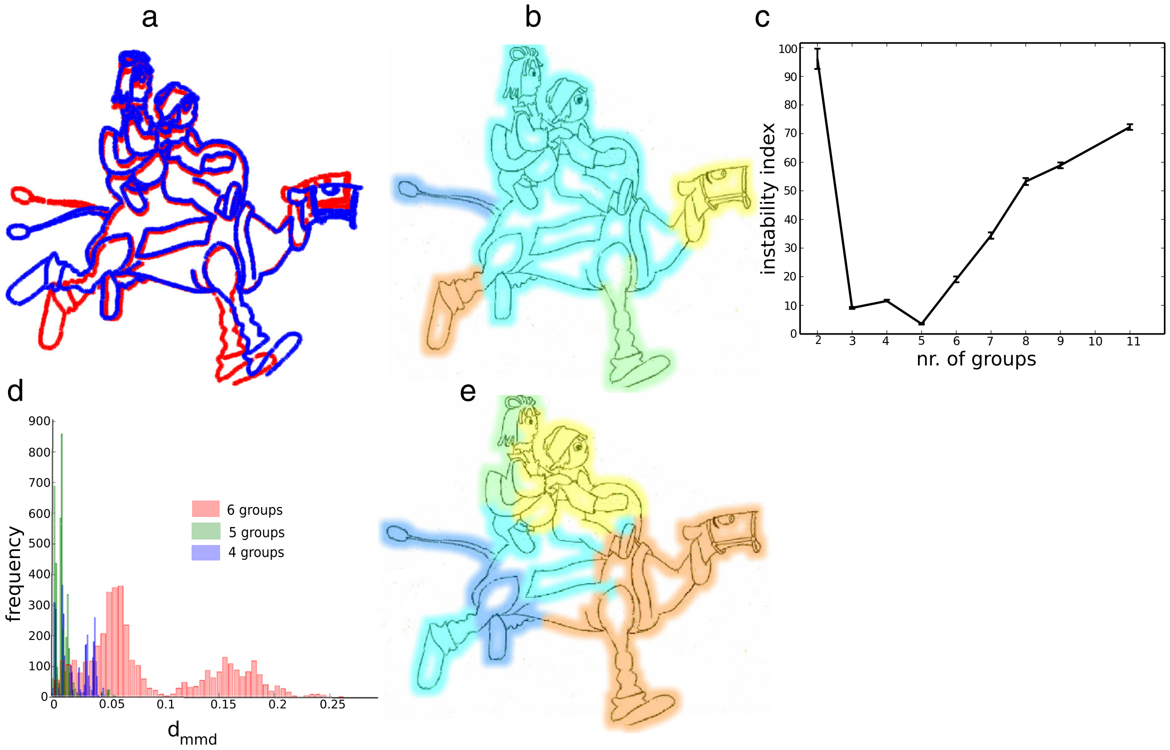

Fig. 1. Results on synthetic data. (a) Original image in blue and distorted image in red. (b) Groups found by our algorithm of Section 3.1 (each color corresponds to a different group). (c) Instability analysis for different numbers of groups. (d) Distribution of pairwise distances 18 for the most stable solutions. Since the distribution is not normal, taking the (normalized) sample mean as in Luxburg [23] is not appropriate.(e) Resulting clustering if the algorithm of Monroy et al. [17], which utilizes a *single* initial *k*-tuple of affine transformations is used instead of our LP-based method described in Section 3.2.1 and shown in panel b.

T^ν was alternated with the actualization of matrix M until local convergence was reached. However, this procedure turned out to be very susceptible to the initialization of the matrix M. We show this fact in Fig. 1(e) where a textitsingle initial k-tuple of affine transformations led to a wrong clustering, where parts in the shape corresponding to different affine deformations were mixed into the same group. This paper studies an orthogonal approach for solving the aforementioned problem leading to better results as shown in Fig. 1(b) (see experimental section for more details). Instead of proposing an initialization for the matrix M or a *single* k-tuple of affine transformations and thus risking a wrong initialization, we construct a large superset of affine transformations

$$T_{\text{pool}} := \left\{T^\nu \middle| T^\nu \in \mathbb{R}^{3\times 3},\ \ \nu = 1, \cdots, l\right\}, \tag{10}$$

where $l > > k$. For this purpose, the shape Y is subdivided into non-overlapping small segments, each of them containing at least 6 non-collinear points. For each segment an affine transformation is estimated and added to the superset T_{mboxpool} (we assume to have an estimate of matrix C). At this point, if there are not enough samples to estimate an affine transformation due to an occlusion, the remaining points in that region will be considered as outliers, and the region will not be matched. Thereafter, each segment is merged with its nearest neighbor, and an affine transformation is calculated for the merged segment, which in turn is added to T_{pool}. For the nearest neighbor estimation, the distance between two segments is defined as the Euclidean distance between their centers of mass (i.e., the average of the segment points). This merging is repeated until the whole shape is merged into a single segment. Thereafter, using this superset T_{pool}, our algorithm optimally selects a subset of k transformations that best register the shape and use these active transformations to estimate the matrix M. Based on this matrix, the active transformations are then updated in turn. Thus, the original problem (Eq. (6)) is transformed into its final form:

$$\underset{M,W,C,T^\nu}{min} \underbrace{\sum_{\nu=1}^{l} w_\nu \left(\sum_{i=1}^{n} M_{\nu i} r_{\nu i}\right)}_{=:E_{lin}(W,M,C,T^\nu)} + E_{\text{quad}} \tag{11}$$

$$s.t.\ \sum_{\nu=1}^{l} w_\nu = k, \tag{12}$$

$$n * w_\nu - \sum_{i=1}^{n} M_{\nu i} \geq 0\ \ (\forall \nu = 1, \cdots, l) \tag{13}$$

$$w_\nu \in \{0, 1\} \tag{14}$$

plus Eqs. (7)–(9). Here the binary vector $w_\nu = 1$ indicates that the νth element of the set T_{pool} is being used and otherwise $w_\nu = 0$. While Eq. (12) guarantees to obtain the desired number of transformations k, Eq. (13) avoids the assignment of points to inactive transformations $w_\nu = 0$. This becomes clearer by remarking that Eq. (13) is fulfilled whenever the logical constraint $w_\nu = 0 \Rightarrow \sum_{i}^{n}{=}_{1} M_{\nu i} = 0$ is met.

3.2.2. Finding correspondences

We first describe how to estimate the correspondence matrix C between shapes Y and X assuming the knowledge of the groups M and the transformations T^{ν} (i.e., transformations T^{ν} where $w_{\nu} = 1$) Thus, Eq. (11) can be alternatively formulated as

$$\min_{z} \quad \sum_{\nu=1}^{k} z^{T} D^{\nu} z; \;\; s.t. \;\; Az = 1, \;\; z \in \{0, 1\}. \tag{15}$$

In this case, z is an indicator vector such that $z_{ia} = 1$ if point y_i is matched to point x_a and otherwise zero. In this formulation, the original matrix C is implicitly included in the vector z. Furthermore, each matrix D^{ν} contains the values $d(T^{\nu}y_i, T^{\nu}y_j; x_a, x_b)$ corresponding to the group ν and otherwise zero. Whereas the diagonal of D^{ν} consists of the linear terms of Eq. (11), the many-to-one constraints of matrix C are expressed through the matrix A. In order to solve for each group independently, vectors u^{ν} are defined which contain all entries of the form $z_{i\bullet}$ for which $M_{\nu i} = 1, w_{\nu} = 1$. Using this vector, we obtain the following local problems:

$$\min_{u_{\nu}} \quad u^{\nu T} D_{|u^{\nu} u^{\nu}} \tag{16}$$

$$s.t. \quad A_{|u^{\nu}} u^{\nu} = 1, \;\; u^{\nu} \in \{0, 1\}, \;\; (\forall \nu : w^{\nu} = 1) \tag{17}$$

where $D_{|u^{\nu}}$ is the submatrix of D^{ν} containing only pairwise distortions related to points belonging to group ν (the non-zero submatrix of D^{ν} in Eq. (15)) and $A_{|u^{\nu}}$ is the many-to-one constraint submatrix of A for the corresponding points. Each of the subproblems in Eq. (16) is then approximated using the integer projected fixed point (IPFP) algorithm for graph matching ([15]). To estimate the initial matrix C required by the IPFP algorithm, both shapes X and Y are registered using a single global transformation (e.g., using a global affine transformation [19]), and for each point in Y, its correspondent point is given as the nearest neighbor point in X. Although we cannot guarantee finding a global minimum for problem (15), we are able to reduce the energy (Eq. (11)) at each iteration (given the matrices M, W) since the solution of each subproblem (Eq. (16)) reduces the total energy of the joint problem (see, e.g., [15]). In practice, this is confirmed through the improvement of the matching accuracy (see Fig. 2(a) in the Experimental section).

3.2.3. LP-based solution for transformations and assignment of points to groups

In this section, we describe how to estimate the active transformations (i.e., the vector W), assign points to the corresponding transformations (through the matrix M), and update them afterwards (we assume to have the matrix C). Jointly optimizing W and M in $E_{\text{lin}} + E_{\text{quad}}$ (Eq. (11)) is very hard due to the non-linearity of both terms. Thus, to render optimization feasible, we focus on the minimization of the term E_{lin}. Disregarding the term E_{quad} at this moment of the optimization is justified by the fact that the term E_{lin} controls the overall registration error since the matrix M defines the support of the different transformations, which are indirectly given by W. This error is what we intend to minimize, while E_{quad} is mainly required (cf. [1]) to obtain better landmark correspondences C, which are now given at this point. A further difficulty is given by the binary constraints on M and W. For instance, if the elements w_{ν} are relaxed to $w_{\nu} \in [0,1]$, the constraint $\sum_{\nu=1}^{n} w_{\nu} = k$ becomes a soft-constraint. Therefore, despite fulfilling this constraint more than k elements, w_{ν} can become greater than zero due to the relaxation. Thus, Eq. (13) will assign points to more than k transformations yielding a wrong solution to the joint problem. However, this last problem is alleviated if we adopt an alternate procedure to minimize E_{lin}:

- Solve $\min_{W} E_{\text{lin}}$ subject to Eq. (12). During the first iteration, all elements of matrix M are set to one and the transformations to build r are taken from T_{pool}.
- Assign points to active transformations solving the linear program (LP) $min_{M} \sum_{i,\nu=1}^{n,k} M_{\nu i} r_{\nu i}$ subject to the constraints $\sum_{\nu=1}^{k} M_{\nu i} = 1$ (for all $i = 1,\ldots,n$) and $M_{\nu i} \in [0,1]$. Here the matrix $M \in \mathbb{R}^{k \times n}$ only indicates the assignment of points to the k active transformations (and not to the l elements in T_{pool}).
- Update the active transformations T^{ν} using M and W. This is done in an exact manner using weighted least squares ([20]). The exact solution for the transformations is an improvement over [18,17], where the transformations were only approximated using the Levenberg Marquardt algorithm.

Our complete method is summarized in Algorithm 1.

3.3. Choosing the right number of clusters

In this section, we describe how to automatically determine the complexity of the model, that is, the number of affine transformations required for registration. The underlying idea is to measure the fluctuations in the registration results when random subsamples of the shapes are considered. For a given number of clusters k, our algorithm is run on $b_{\max}$ subsampled versions of the original shape Y (specifically, 60% of the points in the shape are randomly subsampled each time). Thus, we obtain the clustering results $\hat{M}_b \in \mathbb{R}^{n_s \times 1}$ ($b = 1, \cdots, b_{max}$, $n_s = \lfloor 0.6 * n \rfloor$), where $\hat{M}_b$ indicates the cluster number for each point in shape Y. Since the $b_{\max}$ clustering solutions are calculated on a subset of the points, they are extended to the whole shape using nearest neighbors

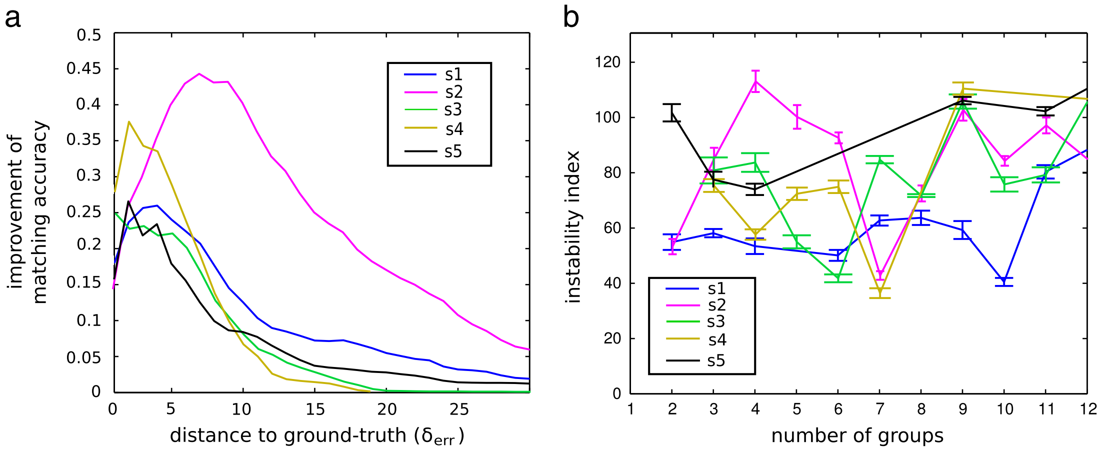

Fig. 2. Results for the Codex Manesse corpus. (a) Improvement in the matching accuracy between the last and first iteration of our algorithm: we plot the difference between the matching accuracy curves of the last and the first iteration. Matching accuracy is the percentage of correspondences where the ground-truth correspondent point lies at most δ_{err} from the predicted correspondent point (see Section 4.2). (b) Instability analysis for all shapes in the corpus and showing the standard deviation for each k.

for the missing points. The extended clustering solutions are referred to by $M_b \in \mathbb{R}^{n \times 1}$. Thereafter, pairwise distances between the different cluster solutions are calculated in order to evaluate the fluctuations in the results induced by the random subsampling. This is done using the minimal matching distance

$$\hat{d}_{\text{mmd}}\left(M_i, M_j\right) = \underset{\pi}{min} \frac{1}{n} \sum_{i=1}^{n} 1_{\left[M_i(i) \neq \pi\left(M_j(i)\right)\right]}, \tag{18}$$

where the minimum is taken over all permutations π of the k labels. In other words, $\hat{d}_{\text{mmd}}\left(M_i, M_j\right)$ measures the percentage of points that changed the assignment (up to a permutation). However, in order to avoid bias when the number of clusters k is increased, $\hat{d}_{\text{mmd}}$ is normalized similar to the study of Lange et al. [13] with the median $r(n)$ of pairwise distances between random labelings. Thus, the fluctuations in the clustering results can be measured by

$$d_{\text{mmd}}\left(M_i, M_j\right) := \frac{\hat{d}_{\text{mmd}}\left(M_i, M_j\right)}{r(n)} \tag{19}$$

In the case of stable clustering solutions, the pairwise distances $d_{\text{mmd}}(M_i, M_j)$ are expected to be near zero. In contrast, unstable solutions yield variations in the clusterings and large distances (see Fig. 1(d)). Therefore, we measure the instability of a solution by approximating the empirical distribution of pairwise distances $d_{\text{mmd}}(M_i, M_j)$ through a histogram $h \in \mathbb{R}^{\text{nbins} \times 1}$ over the distances and define as a measure for the instability the sum of weighted counts:

$$\text{instab}(k) := \sum_{i}^{\text{nbins}} h(i) * c_h(i), \tag{20}$$

where $h(i)$ is the absolute count and $c_h(i)$ is the value of the histogram bin i. Since the number of runs $b_{\max}$ is the same for every value of k, the absolute counts of the histogram can be used without introducing any bias. This measure penalizes distances which are far from zero and, thus, corresponds to unstable clustering solutions for a certain value k. Therefore, the ideal most stable number of affine transformations required for registration is defined as

$$k_{\text{opt}} := \underset{k}{min}\, \text{instab}(k). \tag{21}$$

Consequently, using such an instability measure also helps to approximate the different local deformations between the images. While a region with affine deformation can be represented using a single affine transformation, the stability analysis selects a larger number of affine transformations to approximate non-affine deformations.

Algorithm 1. Summary of the algorithm presented in this paper.

```
    Input: original image $I_X$, reproduction $I_Y$
    Output: $k_{\text{opt}}$, $T^{\nu}_{\text{end}}$, $M_{\text{end}}$, $C_{\text{end}}$, $(\nu = 1, \cdots, k_{\text{opt}})$
1   $\hat{X} \in \mathbb{R}^{n \times 3}$, $\hat{Y} \in \mathbb{R}^{m \times 3}$ ← Landmark points sampling
2   $C_{\text{init}}$ ← Initial global affine registration
3   for $k = 1, \cdots, k_{max}$    ▷ Number of groups
4     for $b = 1, \cdots, b_{max}$ ▷ Iteration for subsamplings
5       $X \in \mathbb{R}^{n_s \times 3}$, $Y \in \mathbb{R}^{m_s \times 3}$ ← Subsampling of $\hat{X}$, $\hat{Y}$
6       $T_{\text{pool}}$ ← Initial pool of transformations
7       do
8         $M_{old} \leftarrow M^b,\ C_{old} \leftarrow C^b$
9         $W^b \leftarrow \min_W E_{lin}(M, W, T^{\nu})$ ▷ Section 3.2.3
10        $M^b \leftarrow \min_M E_{lin}(M, W^b, T^{\nu})$
11        $T^{\nu}_b$ ← Weighted least squares given $M^b$, $W^b$
12        $C^b \leftarrow \min_C E_{lin} + E_{quad}$ ▷ Problem (16)
13      while$(C^b \neq C_{old} \wedge M^b \neq M_{old})$
14    end for
15    $\text{instab}(k) = \sum_i^{\text{nbins}} h^b(i) * c_h^b(i)$, $(b = 1, \cdots, b_{max})$▷ Eq. (20)
16    $k \leftarrow k + 1$
17  end for
18  $k_{\text{opt}} \leftarrow \min_k \text{instab}(k)$ ▷ Eq. (21)
19  $M_{\text{end}}$, $C_{\text{end}}$, $T^{\nu}_{\text{end}}$ ← Repeat steps (6-13) once using $k \leftarrow k_{\text{opt}}$, $X \leftarrow \hat{X}$, $Y \leftarrow \hat{Y}$
```

4. Experiments

4.1. Synthetic data

We first evaluate our algorithm on two frames of a synthetic image sequence. Fig. 1(a) shows both frames in red and blue, respectively. The head, both legs, and tail were modified through affine transformations, and thus, the global non-linear deformation between frames is known. In this case, around 4000 points were used to describe the shape and were uniformly sampled along the contours of the image. In order to carry out the stability analysis (see Section 3.3), 60% of the points were uniformly subsampled and the algorithm was run $b_{\max} = 60$ times for each given number of clusters k (on average the algorithm converged within 5 iterations each run). This resulted in 3600 pairwise distances for each k. This experiment was repeated 20 times, thus yielding the instability plot of Fig. 1(c). The algorithm determined $k = 5$ to be the most stable number of groups. As Fig. 1(b) shows (each color represents a single group), the corresponding groups are consistent with the manually introduced deformations. This experiment shows not only how our algorithm registers both shapes but also how the inferred groups describe and visualize how the different local parts in the shape were truly deformed. Regarding this synthetic experiment, the distribution of the pairwise distances $d_{\text{mmd}}(M_i, M_j)$ (Eq. (19)) for the most stable number of groups ($k = 4,5,6$) is also shown in Fig. 1(d). Luxburg [23] (p. 5) mentions that a simple (normalized) mean over pairwise clustering distances $d_{\text{mmd}}(M_i, M_j)$ is commonly used as instability measure. This methodology presupposes that the distribution of the pairwise distances is normal, and thus, the instability measure weights every pairwise distance equally. However, in Fig. 1(d), we show that the distribution of pairwise distances is in general not normal. Therefore, our measure in Eq. (20) is more appropriate to describe the shape of the distribution since it weights the pairwise distances proportional to their occurrence. Finally, the benefit of our LP-based method (Section 3.2.3) for calculating the affine transformations and the assignment of points to them are evaluated by comparing our method with an alternative procedure based on the algorithm previously explored by Monroy et al. [17]. Instead of using a pool of affine transformations T_{pool} and the LP-based method described in Section 3.2.1, we provided a single initial k-tuple of transformations by locally grouping points in a greedy manner based on their proximity and registration quality. This resulted in a deficient initialization which the successive updates of groups, transformations, and correspondences could not correct. While in Fig. 1(e) we can observe how parts in the shape corresponding to different affine components were mixed into the same group resulting in a clustering which is not consistent with the ground truth, our current method (Fig. 1(b)) groups the different shape parts correctly.

4.2. Reproductions of the Codex Manesse

In the study of Monroy et al. [18], we collected a corpus of 5 shapes coming from the Codex Manesse (reproduced between ca. 1305 and ca. 1340) and their corresponding reproductions commissioned in 1746/47 by J. J. Bodmer and J. J. Breitinger. Since ground truth for the correspondences between the shapes is known, it is possible to measure the registration quality of our method and compare it with other state-of-the-art algorithms. In Table 1, we show the mean squared error (MSE) of the registration for all shapes. The number of affine transformations for all models was automatically determined by our algorithm. Furthermore, we were also interested in measuring the ability of humans to perceive deformations in different parts of a shape. Therefore, we developed an interactive registration tool that was used by 5 experts to manually select the regions in the shape that according to their perception

Table 1
Reproductions of the Codex Manesse. Mean squared error (MSE) of the registration using ground-truth correspondences provided by Monroy et al. [18]. The complexity of the piecewise affine transformation is automatically provided by our method.

	Registration quality (MSE)						
Shape ID (no. groups)	[25]	K_{means}	Ward	[18]	Human	[8]	Our method
Shape 1 (10)	49.36	37.20 ± 2.25	35.46	34.71	57.91 ± 9.93	25.04	24.89
Shape 2 (7)	109.26	80.98 ± 5.30	84.19	131.07	194.33 ± 6.34	260.60	78.55
Shape 3 (6)	24.11	35.77 ± 1.27	36.15	45.62	37.06 ± 5.61	24.68	21.41
Shape 4 (7)	28.57	37.37 ± 0.99	39.26	37.68	44.21 ± 7.97	35.77	28.37
Shape 5 (4)	52.12	57.52 ± 4.83	52.66	66.89	60.23 ± 1.03	67.84	45.86
Average	52.68	49.76 ± 2.92	49.54	63.19	78.74 ± 6.17	82.78	39.81

shared the same transformation. At the beginning of the experiment, both shapes were registered using a single affine transformation. Thereafter, each time a new group of points was selected, the overall shape registration was updated, enabling each user to see the result of his selection. Moreover, it was always possible to correct a group selected before. The average of the MSE over the 5 experts in the experiment is shown in Table 1 under the row *human*. From the large MSE, it becomes clear that the task of an art historian to manually analyze a shape to understand the drawing process is extremely difficult. Thus, a computer-based procedure is essential. The entries K_{means} and Ward in Table 1 correspond to a piecewise affine registration based on the clustering of the displacement vectors between both shapes using K_{means} and Ward's method, respectively. We have observed that clustering the error vectors featured only insufficient accuracy: contours have been distorted (e.g., stretched), and junctions are partly missing and thus affine deformations cannot be described by clustering the displacement term of the deformation. A similar method to Wang and Adelson [25] was also reimplemented (second row of Table 1). For this method, not the displacement vectors but the parameters of the affine transformations contained in T_{pool} were directly clustered instead. Thereafter, we greedily iterated between the assignment of *each point* to the centroids (i.e., the affine transformation representing a group) based on its registration error and the refinement of the centroids themselves. The clustering of transformation parameters resulted in being unstable since they strongly varied depending on the locality of their support. Furthermore, the greedy assignment of points to transformations was also not optimal.

In Table 1, we also added the output of the algorithm from Monroy et al. [18] for comparison. In this case, we observed that areas in a shape were grouped based on their proximity due to the pairwise Euclidean distance term used in their objective function. This bias was also observed in the results of [8], where the assignments to transformations were also regularized by a Euclidean distance-based term in their energy function. This fact had an important impact on the registration since parts of the shape featuring different transformations were forced to be registered together and a bigger MSE was produced. Finally, it is important to remark that all of these methods with exception of the presented one only partially solved the full task since the correspondence between shapes is not calculated. Furthermore, it is not possible to automatically determine the model's complexity as we do in our method.

Since we have ground truth for the correspondences, we also measured the improvement of the matching quality between shapes induced by our algorithm. For this, we measured for each point y_i in shape Y the error produced between its estimated corresponding point $\sum_{j=1}^{m} C_{ij} x_j$ (as induced by the binary matrix C in our algorithm) and its true correspondent point x_i^{gt} (as provided in the ground-truth) in shape X. We then measured for a threshold δ_{err} (which we then vary in turn) the percentage of points, where the estimated correspondences lie at most δ_{err} from the ground truth. This yields a matching-accuracy curve depending on the parameter δ_{err}. In Fig. 2(a), we show the relative improvement in the matching accuracy between the last and the first iteration (initialization) of our algorithm. Thus, optimizing the correspondence matrix C together with the groups M is beneficial for the registration process. In Fig. 2(b), we show the stability of the solutions for the whole corpus of shapes and observe that a local minimum value, indicating a stable solution for all shapes in the corpus is always obtained.

Finally, in order to compare the registration quality of our method with the CPD algorithm of Myronenko and Song [19], we focused on complex medieval scenes. For this, we used reproductions of the codex of Eike von Repgow's Sachsenspiegel ('Mirror of the Saxons') composed ca. 1220–1235 in eastern Saxony, see Fig. 3 for one example. The average computation time of our algorithm for such complex scenes

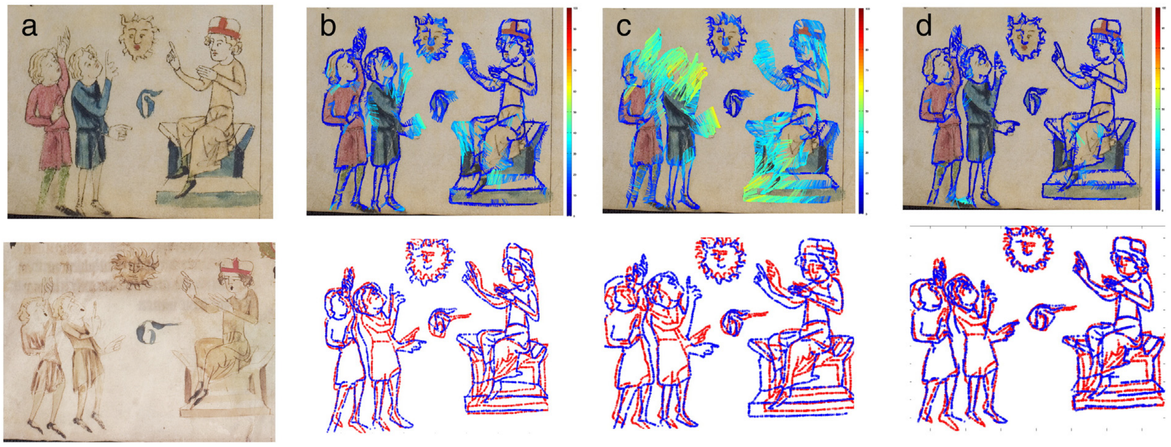

Fig. 3. Registration quality for complex scenes. (a) Reproductions of the codex of Eike von Repgow's Sachsenspiegel ('Mirror of the Saxons') composed ca. 1220–1235 in eastern Saxony. (b) Registration using the CPD algorithm of Myronenko and Song [19] (with an RMSE of 11.45 for this image). (c) Rigid registration (RMSE 17.48) (d) Registration results using our method (RMSE 7.78).

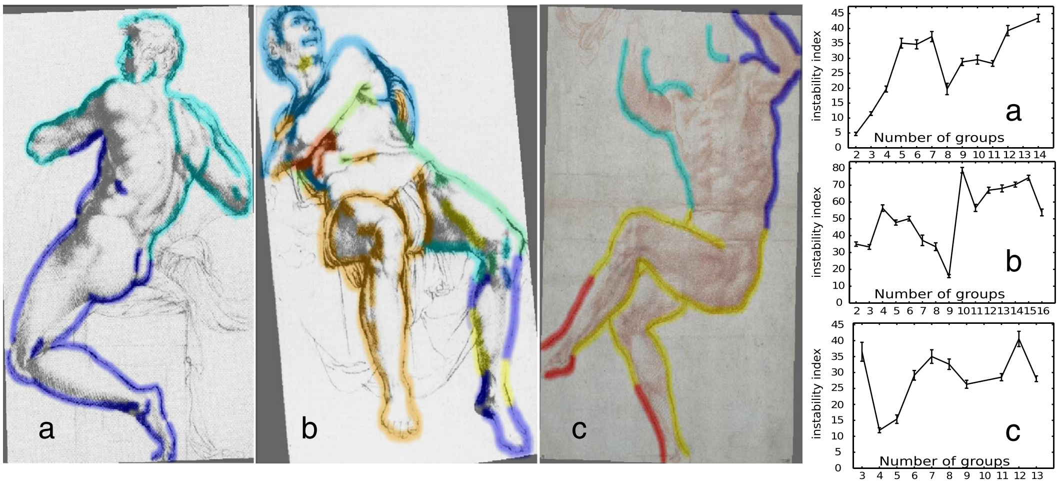

Fig. 4. Analysis of the drawing process. First to third column: Different shapes were reproduced based in semantic entities (e.g., legs, arms, etc.). The grouped parts are mostly anatomically or technically sensible, whereas the parts that are split in different groups show a complex deformation for that area. Last column: corresponding instability analysis together with standard deviation.

is on the order of a minute. While the CPD algorithm of Myronenko and Song [19] (using the default parameters) obtained a root MSE (RMSE) of 12.64 ± 11.96 for the registration error over 3 scenes, our method improved the registration with an RMSE of 8.79 ± 5.8. In contrast to this, registering the scene with a rigid transformation resulted in a poor RMSE of 20.65 ± 15.57. We observed that the improvement of our method over the CPD algorithm was mainly due to CPD regulating its complexity through a global parameter for the whole image, whereas our method has a greater flexibility since it adapts its complexity based on its piecewise nature according to the underlying deformation.

4.3. Michelangelo reproductions

We also focused on the analysis of Michelangelo's ceiling fresco in the Sistine Chapel (1508–1512) and compare distinctive shapes with sketches, which were made by artists surrounding Michelangelo, probably after preparatory drawings or by Dutch artists after the original. The reason is that the differences between a drawing and a mural painting are greater than the tracings from book illustrations like the reproductions in the last section. Our aim here is not to reconsider the connoisseurs' controversy about the attribution of these drawings but to show how our automatic approach is used to analyze the reproduction process of an artwork, which in turn is noteworthy for an art-historical analysis.

The first column in Fig. 5 shows the original fresco images. The second column shows two reproductions and a preparatory drawing. All three images in the second column seem to be reproduced exactly from the first column images. However, after applying an overall rigid transformation, we see that the drawings feature important differences and show non-linear deviations from the fresco. This can be seen in the third row of Fig. 5, where the color of the arrows indicates the magnitude of the induced rigid registration error. Using our method, it is possible to discover a structure in the overall deformation by observing the resulting groups obtained by our algorithm (see Fig. 4). For instance, the Ignudo (i.e., the male nude flanking the Creation of Eve) in Fig. 5(a) features only two relevant deformations: while the upper and lower part of the body can be exactly registered to the other image, both parts together yield a non-linear deformation. From an artistic point of view, this inconsistency can be explained by noting the difficulty of bringing both body parts into an appropriate distance and angle to each other by the artist during the reproduction of the fresco and alterations between these parts can easily be introduced in this procedure. Furthermore, the Prophet Jonah in Fig. 5(b) features very interesting groups: while the left leg fits using a single transformation, the right leg decomposes mainly into three groups which correspond to the observation that this body part substantially differs from the leg in the fresco. in Fig. 5(c) we can observe how the torso decomposes into the right and left arm indicating a deliberate amplification of the articulation in the sketching. Since our energy cost does not introduce any proximity term that could bias the result, it can be concluded that the artist approached the reproduction by independently reproducing smaller parts corresponding to semantical entities. From an art historical point of view, whereas these parts can be considered as technically sensible, regions in the shape that were split in different groups indicate a possible difficulty of reproducing that area for the artist.

4.4. Implementation details

For shape drawings, we have to extract and deal with different contour thickness and texture. Hence, contours are extracted by convolving the image with Laplace of Gaussian (LoG) Filters of varying sigma ($\sigma = 0.8 + j * 0.4, j = 1,\ldots,9$) and then take the maximal response over all sigmas for every pixel. This kind of filter is suitable since it allows obtaining a single response for lines of varying thickness and ensures in praxis a good contrast between ridge response and background. Finally, non-maximum suppression followed by hysteresis thresholding is applied to obtain a single binary response. For images where shape is encoded through texture and color boundaries, we use the Pb code ([16]) for edge extraction, which weights edge signals proportionally to their strength.

In Section 3.2.2, we have estimated correspondences for each group independently. When the group is too large (e.g., more than 1/5 of all points in the shape), each group is subdivided into smaller pieces based on a bottom-up contour grouping (using the Euclidean distance), and then the point correspondence for each subgroup is independently estimated. However, we force the groups to reach a minimum size to guarantee a robust matching.

5. Conclusion

This paper has presented a novel approach for the analysis of alterations between artworks and their reproductions. Therefore, the overall shape deformation is represented by decomposition into a piecewise affine model. Model complexity was automatically estimated using a statistical stability analysis. The present contribution jointly estimated the correspondences between shapes, the affine structures in the shape,

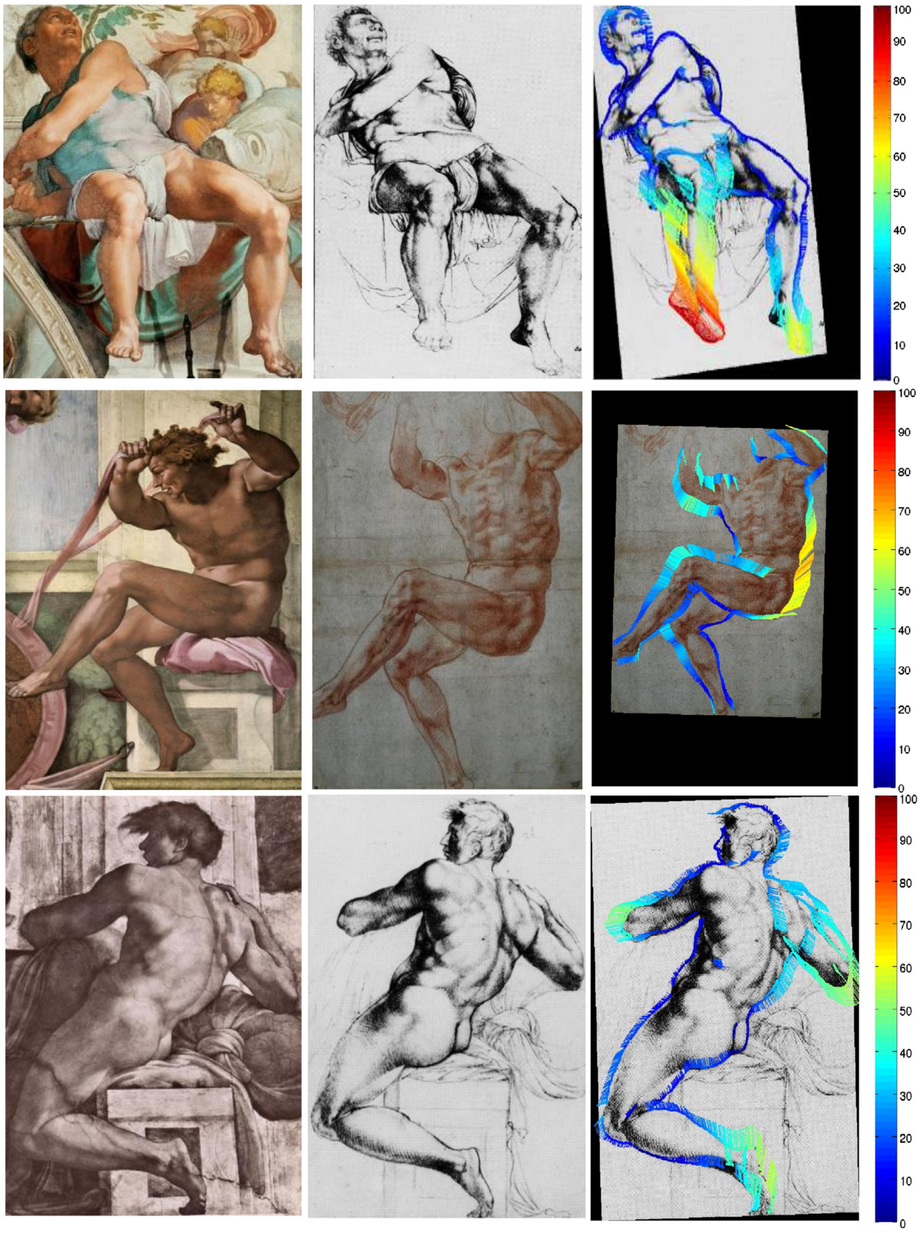

Fig. 5. Parts of Michelangelo's ceiling fresco in the Sistine Chapel (1508–1512). First column: original fresco images. Second column: Sketches made after Michelangelo's own preparatory drawings or by Dutch artists after the original. Third column: Error between rigid registered images. The color of the arrows corresponds to the magnitude of the registration error.

and the complexity required by the overall deformation model. We have tested our method in controlled scenarios as well as with real historical images. Based on ground-truth correspondences between images from the Codex Manesse and their 18th century reproductions, we have observed an improvement over the state-of-the-art in both registration and matching quality. Furthermore, our algorithm outperformed a manual solution of the problem showing the benefit of this method for art historians. Finally, an important experimental finding was the discovery that the drawings of two of the Ignudi and the Prophet Jonah in the ceiling fresco of the Sistine Chapel featured different deformations. These deformations corresponded either to semantical entities of the shape (e.g., the arms in Fig. 4(c)) or indicated slight modifications in the relative position of extremities (e.g., Fig. 4(a) and (b)) by the artist.

References

[1] A.C. Berg, T.L. Berg, J. Malik, Shape matching and object recognition using low distortion correspondence, CVPR, 2005, pp. 26–33.

[2] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2001) 1222–1239.

[3] B. Carque, Zur manuellen Reproduktion der Miniaturen, Die Codex Manese und die Endeckung der Liebe, 2010.

[4] Y.S. Chang, D.G. Stork, Warping realist art to ensure consitent perspective: a new software tool for art investigations, Human vision and electronic, imaging, 2012.

[5] H. Chui, A. Rangarajan, A new point matching algorithm for non-rigid registration, Comput. Vis. Image Underst. 89 (2–3) (2003) 114–141.

[6] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P.-Y. Bondiau, N. Ayache, G. Malandain, An efficient locally affine framework for the smooth registration of anatomical structures, Elsevier Science, 2008.

[7] N. da Silva, J. Costeira, Subspace segmentation with outliers: a grassmannian approach to the maximum consesus subspace, CVPR, 2008.

[8] A. Delong, A. Osokin, H. Isack, Y. Boykov, et al., Fast Approximate Energy Minimization with Label Costs, CVPR, 2010.

[9] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, D. Kriegman, Clustering appearances of objects under varying illumination conditions, CVPR, 2003.

[10] L. Hongsheng, J. Huang, S. Zhang, X. Huang, Optimal object matching via convexification and composition, ICCV, 2011.

[11] I. Kanatani, Motion segmentation by subspace separation and model selection, ICCV, 2001.

[12] N. Komodakis, N. Paragios, G. Tziritas, Clustering via LP-based stabilities, NIPS, 2009.

[13] T. Lange, V. Roth, M.L. Braun, J.M. Buhmann, Stability-based validation of clustering solutions, Neural Comput. 16 (6) (2004) 1299–1323.

[14] N. Lazic, I. Givoni, B. Frey, FLoSS: facility location for subspace segmentation, ICCV, 2009.

[15] M. Leordeanu, M. Herbert, R. Sukthankar, An integer projected fixed point method for graph matching and MAP inference, NIPS, Springer, 2009.

[16] M. Maire, P. Arbelaez, C. Fowlkes, J. Malik, Using contours to detect and localize junctions in natural images, CVPR, 2008.

[17] A. Monroy, P. Bell, B. Ommer, Shaping art with art: morphological analysis for investigating artistic reproductions, ECCV, , VISART, 2012.

[18] A. Monroy, B. Carque, B. Ommer, Reconstructing the drawing process of reproductions from medieval images, ICIP, 2011.

[19] A. Myronenko, X. Song, Point set registration: coherent point drift, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010).

[20] C. Rao, et al., Linear models: least squares and alternatives, Springer Series in Statistics, 1999.

[21] D. Sýkora, J. Dingliana, S. Collins, As-rigid-as-possible image registration for hand-drawn cartoon, NPAR, 2009.

[22] Y. Usami, D.G. Stork, J. Fujiki, H. Hino, S. Akaho, N. Murata, Improved methods for dewarping images in convex mirrors in fine art: applications to van Eyck and Parmigianino, Computer vision and Image analysis of art II, 2011.

[23] U. von Luxburg, Clustering stability: an overview, Found. Trends Mach. Learn. 2 (3) (2010).

[24] R. Vidal, Y. Ma, S. Sastry, Generalized principal component analysis (GPCA), IEEE Trans. Pattern Anal. Mach. Intell. 27 (12) (2005) 1945–1959.

[25] J. Wang, E. Adelson, Representing moving images with layers, IEEE Trans. Image Process. 3 (5) (1994).

[26] J. Yanv, M. Pollefeys, A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate, ECCV, 2006.