A Morphometric Approach to Reception Analysis of Premodern Art

Antonio Monroy[1], Peter Bell[1 2], Björn Ommer[1]

[1]HCI & IWR, [2]Institute of European Art History
University of Heidelberg

Abstract:
The focus of the present paper is to introduce a computer-based methodology for measuring and analyzing the variability between compositions of complete scenes and individual objects of similar medieval manuscripts. On the level of scene compositions, we present a methodology to describe global deformations by a novel piecewise linear registration model that adapts the complexity of each component according to the shape deformation in the underlying region. Moreover, the assignment of regions in a scene or object to different model components induces a clustering of the scene which in turn helps to visualize the structure and geometry of the deformation introduced in the reproduction by the artist. Our algorithm simultaneously infers the correspondences between original and reproduction, finds groups in the image which share the same transformation, and finally estimates the transformation of those groups. Information about the transformation through the illustrations of medieval manuscripts is adjuvant for scholars to reconstruct historic context and semantic changes and is crucial for stylistic interpretation.

1. Introduction

Cultural heritage is not only made up of innovative new art but a significant part are reproductions of existing work and their variations. Therefore it is crucial to evaluate the quality of the reproductions of art as well as their stylistic and semantic changes. Especially the manuscript culture of the Middle Ages flourished through manual reproductions. A prominent example is the frequently reproduced codex of Eike von Repgow's Sachsenspiegel composed ca. 1220-1235 in eastern Saxony. It constitutes an outstanding piece of medieval cultural history. Eike von Repgow's text is one of the oldest prose works written in German and the earliest German vernacular law book and thus an important monument in the history of German law. Only four illustrated redactions of the text remain, these Codices picturati from the 14th century are named after their present location in Heidelberg (H), Dresden (D), Wolfenbüttel (W), and Oldenburg (O). The focus of the present paper is to introduce a computer-based methodology for measuring and analyzing the variability between compositions of complete scenes and individual objects of these different manuscripts. On the level of scene compositions, we present a methodology to describe global deformations between scenes. Therefore, deformations are represented by a novel piecewise linear registration model that adapts the complexity of each component according to the shape deformation in the underlying region. Moreover, the assignment of regions in a scene or object to different model components induces a clustering of the scene which in turn helps to visualize the structure and geometry of the deformation introduced in the reproduction by the artist. The main challenge consists of simultaneously solving three tasks. Firstly, the correspondences between original and reproduction have to be inferred. Secondly, groups in the image which share the same transformation need to be found and finally, the transformations of those groups needs to be estimated. We propose a mathematical optimization approach capable of solving the three tasks simultaneously.

We are going to show that both the Wolfenbüttel and the Dresden codices feature important variations. These changes occur on the level of scene arrangements as well as on single individuals like the magistrate. Our analysis shows that the arrangement of persons is not static but that it changes dynamically according to the artistic and historical context in which each codex was reproduced. Furthermore, we show how a statistical approach can be used to help art historians to better analyze the deformations between objects.

2. Related Work

In [1] 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. We address these issues in the present contribution. The present paper formulates a single optimization problem where affine transformations are estimated and points are grouped within the same procedure.

In [2] 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 [1] 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 [2] we also assumed for simplicity to have fixed 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 goodness 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 [3] 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, accurate estimation of small and continuous deviations in transformations cannot be estimated with this approach. In [4], a regularized energy function was minimized with Graph-Cuts ([5]) 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, the authors of [6] presented a 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. [7] 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 [7], [8]. This latter task commonly assumes that the data points lie on several distinct *linear subspaces* [9], [8], [10], [11], [12]. 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 which define a linear subspace but three tasks need to be solved jointly: the correspondence between both shapes, the groups in the image which share the same transformation, and the estimation of the transformations of those groups.

In the field of computer graphics Sýkora et al. [13] 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, the authors of [13] are not able to handle deformations which 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 creates distractors that need to be handled, whereas the method of [13] is only applied to cartoons without any clutter. Another interesting related work is [14], 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, the authors of [15] 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 which are related to estimating transformations between artworks are [16], [17]. While [16] tries to ensure consistent perspective in art images, [17] aims to dewarp image reflections shown in convex mirrors within very specific paintings. Common non-linear registration algorithms like [18] or [19] are also not suited to the purpose of the present task. Whereas [18] uses a Thin Plate Spline (TPS) to model the transformation, [19] estimates a displacement vector for each point in the shape. In both cases, these models introduce artifacts in the registration as observed in [2], which is undesirable for art comparison.

3. Methodology for Scene Analysis

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. Thus, the shape of the original artwork is referred to with the matrix $X \in \mathbb{R}^{m \times 3}$ and with $Y \in \mathbb{R}^{n \times 3}$ the reproduced shape.

3.1. Problem statement

The main challenge consists of simultaneously solving three tasks. Firstly, the correspondences between both shapes have to be inferred. Secondly, the groups in the image which share the same transformation need to be found and finally, the transformations of those groups and thus the overall deformation model needs to be estimated. The missing groups correspond to image regions which 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:

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

This paper studies an alternating approach for solving the aforementioned problem by first constructing a large superset of affine transformations

$$T_{\text{pool}} := \{T^\nu \,|\, T^\nu \in \mathbb{R}^{3 \times 3}, \, \nu = 1, \cdots, l\}, \tag{2}$$

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_{pool} (we assume to have an estimate of matrix C). 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 problem we intend to solve can be formulatted as follows:

$$\min_{M,W,C,T^\nu} \underbrace{\sum_{\nu=1}^{l} w_\nu \left(\sum_{i=1}^{n} M_{\nu i} r_{\nu i} \right)}_{=:E_{lin}(W,M,C,T^\nu)} + E_{quad} \tag{3}$$

$$s.t. \sum_{\nu=1}^{l} w_\nu = k, \tag{4}$$

$$n * w_\nu - \sum_{i=1}^{n} M_{\nu i} \geq 0 \ (\forall \nu = 1, \cdots, l) \tag{5}$$

$$w_\nu \in \{0,1\} \tag{6}$$

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

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

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

Here the binary vector $w_\nu = 1$ indicates that the ν-th element of the set T_{pool} is being used and otherwise $w_\nu = 0$. Whereas the constraint (4) guarantees to obtain the desired number of transformations k, the constraint (5) avoids the assignment of points to inactive transformations $w_\nu = 0$. This becomes clearer by remarking that constraint (5) is fulfilled whenever the logical constraint $w_\nu = 0 \Rightarrow \sum_{i=1}^{n} M_{\nu i} = 0$ is met.

3.1.1. Finding correspondences

The quadratic term E_{quad} we inserted in 3 measures the distortion between pairs of points between the shapes and is used to improve the estimation of matrix C. This measure is defined as

$$\begin{aligned} d(y_i, y_j; x_a, x_b) := & \gamma d_a(y_i, y_j; x_a, x_b) \\ & + (1-\gamma) d_l(y_i, y_j; x_a, x_b), \end{aligned} \tag{10}$$

$$d_a(y_i, y_j; x_a, x_b) := \left(\frac{\alpha_d}{|s_{ij}|} + \beta_d \left| \arcsin\left(\frac{\hat{s}_{ab} \times s_{ij}}{|\hat{s}_{ab}||s_{ij}|} \right) \right| \right), \tag{11}$$

$$d_l(y_i, y_j; x_a, x_b) := \frac{||s_{ij}||\hat{s}_{ab}||}{(|s_{ijf}| + \sigma_d)}; \tag{12}$$

$$s_{ij} := y_i - y_j, \ \ \hat{s}_{ab} := x_a - x_b. \tag{13}$$

In order 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 $w_\nu = 1$) we need to solve the problem (3) can be alternatively formulated as

$$\min_{z} \quad \sum_{\nu=1}^{k} z^T D^\nu z; \; s.t. \; Az = 1, \; z \in \{0,1\}. \tag{14}$$

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 equation (3), the many-to-one constraints of matrix C are expressed through the matrix A. We solve for each group independently using the IPFP algorithm. As starting solution both shapes X and Y are registered using a single global transformation.

3.1.2. 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). This is a hard task due to the quadratic non-linear term E_{quad} in Eq. (3). Therefore, in praxis we focus only on the minimization of the linear term. During the first iteration, all elements of matrix M are set to one and the transformations to build r are taken from T_{pool}. We adopt an alternate procedure to minimize the linear term:

- 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 $i = 1, \cdots, n$) 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 [1], [2], where the transformations were only approximated using the Levenberg Marquardt algorithm.

3.2. 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 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

$$\hat{d}_{\text{mmd}}(M_i, M_j) = \min_{\pi} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[M_i(i) \neq \pi(M_j(i))]}, \tag{15}$$

where the minimum is taken over all permutations π of the k labels. In other words, $\hat{d}_{\text{mmd}}(M_i, M_j)$ measures the percentage of points which changed the assignment (up to a permutation). In order to avoid a bias when the number of clusters k is increased, $\hat{d}_{\text{mmd}}$ is normalized similar to [21] with the median $r(n)$ of pairwise distances between random labelings. 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. Therefore,

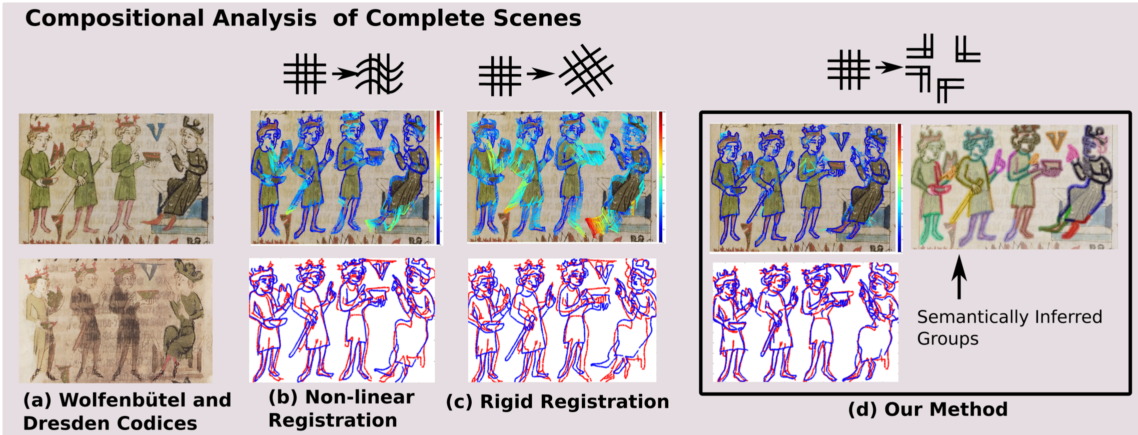

Fig. 1. Results of a compositional analysis of complete scenes.

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:

$$\text{instab}(k) := \sum_{i}^{\text{nbins}} h(i) * c_h(i), \tag{16}$$

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, correspond to unstable clustering solutions for a certain value k. Therefore, the ideal most stable number of affine transformations required for registration is defined as:

$$k_{\text{opt}} := \min_{k} \text{instab}(k). \tag{17}$$

4. Results

We evaluate our approach on five scenes of D and W and show reuslts in Fig . 1 (a)-(d). We observe that whereas a single global rigid transformation (c) is not able to describe the highly complex scene deformation, the global non-linear state-of-the-art registration method (b) is also not able to cope with the global deformation since its complexity is controlled by a single global parameter, thus not allowing for the required local flexibility. In contrast to this, our method (d) is not only capable of improving the registration quality of complex scenes, but it is also able to reveal the structure of how the scene was transformed between the manuscripts. We have observed that the artist approached the reproduction by independently reproducing small parts that correspond to semantical entities. These different parts are visualized in (d), where each color indicates that a region in the image was transformed using certain linear transformation. Furthermore, we show how a statistical approach can be used to help art historians to better understand the deformations between objects visualized in Fig. 2 (e)-(h). The relative deformation in the arrangement of the figures is shown in (e), where the i-th row of the depicted matrix encodes the relative transformation between individuals in the scene if the i-th person is used to align both images. Using principal component analysis we were able to infer and visualize (f) that arm gestures and leg positions are responsible for most of the deformations between the magistrates of a single manuscript. Furthermore, we also discover that the movements of these parts follow a certain pattern. For instance, the movement of body parts within the feet are highly correlated and thus constitute a single entity. However, the pointing finger of both hands moves differently to

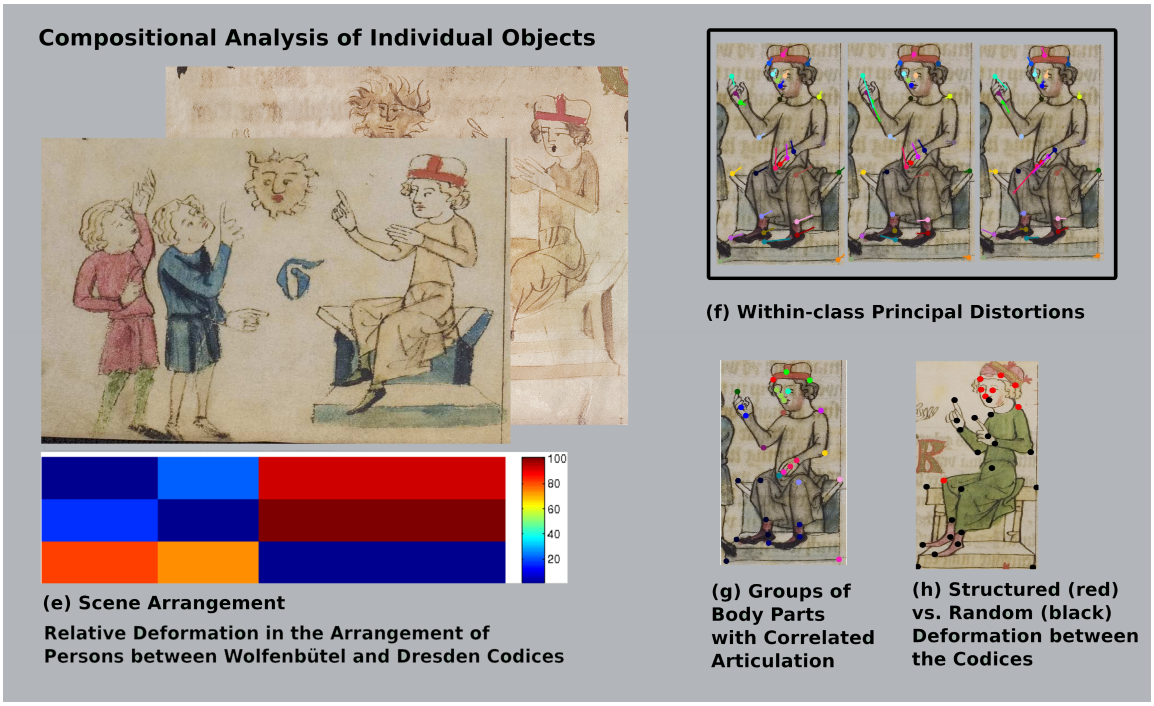

Fig. 2. Analysis of individual objects.

the rest of the hand. Finally, we were also able to discover by means of a Kolmogorov- Smirnov hypothesis test with a significance level of a = 0.05, which parts of the shape consistently feature structured deformations between the codices. For the codices H and D, the head region of the magistrates is systematically deformed between both manuscripts. Therefore we can observe a difference between the two artists in the notion of human proportion. Although one is copying from the other, both have their own ideas about the appearance and posture of the human figures. Bodyparts correalate in their own logic (g). Specially the relationship between body and head is treated independently in D and W. We visualize this finding in (h).

5. Conclusion

An important contribution of the present paper is to show that both the Wolfenbüttel and the Dresden codices feature important variations both at the level of scene arrangements and single individuals. For instance, we observed that in judiciary scenes, the relative distance between the magistrate and the rest of the individuals substantially varies between the codices. Our analysis shows that the arrangement of persons is not static but that it changes dynamically according to the artistic and historical context in which each codex was reproduced.

References

[1] A. Monroy, B. Carque, and B. Ommer, “Reconstructing the Drawing Process of Reproductions from Medieval Images,” in *ICIP*, 2011.

[2] A. Monroy, P. Bell, and B. Ommer, “Shaping Art with Art: Morphological Analysis for Investigating Artistic Reproductions,” *ECCV (VISART)*, 2012.

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

[4] A. Delong, A. Osokin, H. Isack, Y. Boykov *et al.*, “Fast Approximate Energy Minimization with Label Costs,” in *CVPR*, 2010.

[5] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization Via Graph Cuts,” *PAMI:*, no. 29, pp. 1222–1239, 2001.

[6] N. Komodakis, N. Paragios, and G. Tziritas, "Clustering via LP-based Stabilities," in *NIPS:*, 2009.
[7] N. Lazic, I. Givoni, and B. Frey, "FLoSS: Facility Location for Subspace Segmentation," *ICCV:*, 2009.
[8] J. Yanv and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," *ECCV:*, 2006.
[9] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," *CVPR:*, 2003.
[10] N. da Silva and J. Costeira, "Subspace segmentation with outliers: A grassmannian approach to the maximum consesus subspace," *CVPR:*, 2008.
[11] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (gpca)," *PAMI:*, vol. 27, no. 12, pp. 1945–1959, 2005.
[12] I. Kanatani, "Motion segmentation by subspace separation and model selection," *ICCV:*, 2001.
[13] D. Sýkora, J. Dingliana, and S. Collins, "AS-rigid-as-possible image registration for hand-drawn cartoon," in *NPAR*, 2009.
[14] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P.-Y. Bondiau, N. Ayache, and G. Malandain, "An Efficient Locally Affine Framework for the Smooth Registration of Anatomical Structures," *Elsevier Science*, 2008.
[15] L. Hongsheng, J. Huang, S. Zhang, and X. Huang, "Optimal Object Matching via Convexification and Composition," in *ICCV*, 2011.
[16] Y. S. Chang and D. G. Stork, "Warping realist art to ensure consitent perspective: A new software tool for art investigations," in *Human vision and electronic imaging*, 2012.
[17] Y. Usami, D. G. Stork, J. Fujiki, H. Hino, S. Akaho, and N. Murata, "Improved methods for dewarping images in convex mirrors in fine art: Applications to van Eyck and Parmigianino," in *Computer vision and Image analysis of art II*, 2011.
[18] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," vol. 89, no. 2-3, pp. 114–141, 2003.
[19] A. Myronenko and X. Song, "Point Set Registration:Coherent Point Drift," *PAMI*, vol. 32, no. 12, 2010.
[20] C. Rao *et al.*, *Linear Models: Least Squares and Alternatives*, ser. Springer Series in Statistics, 1999.
[21] T. Lange, V. Roth, M. L. Braun, and J. M. Buhmann, "Stability-Based Validation of Clustering Solutions," vol. 16, no. 6, pp. 1299–1323, 2004.