COMPUTER GRAPHICS TECHNICAL REPORTS

CG-2005/02

Correspondence Generation and Matching of 3D Shape Subparts

Marcin Novotni
Universität Bonn. marcin@cs.uni-bonn.de

Patrick Degener
Universität Bonn. degener@cs.uni-bonn.de

Reinhard Klein
Universität Bonn. rk@cs.uni-bonn.de

Institut für Informatik II
Universität Bonn
D-53117 Bonn, Germany

ISSN 1610-8892

The task of setting up correspondences between 3D shapes is a prerequisite for numerous computer graphics applications, which has been performed manually up to now. We present an automatic method for characterizing 3D shape subparts and establishing correspondences between them. The corner stones of our algorithm are: (i) finding stable *salient points* that are representative for certain parts of shapes along with generating the associated local shape descriptors; and (ii) finding matching subsets of salient points so that the local shape dissimilarity and the deformation magnitude due to the resulting mapping is minimal. The salient points are found as minima of the scale space Laplacian-of-Gaussian applied to the characteristic object function, the rotation invariant local shape descriptors are a variant of Spherical Harmonic Descriptors. The matching is performed by finding sets of corresponding salient point pairs minimizing the 3D bending energy computed from the thin-plate spline interpolation of these discrete point mappings.

Unlike the graph based methods, the resulting correspondences are completely independent of object topology but nevertheless stable. Possible applications include (sub-)shape similarity estimation, finding object subparts in a larger scene, generation of dense matching e.g. for morphing or statistical shape analysis.

1. Introduction

In this paper we tackle the correspondence problem for partially similar 3D shapes. Our goal is twofold: First, we aim at solving the partial matching problem, i.e. detect partial similarities between two or more shapes. Second, we need to establish a partial correspondence between subparts that are found to match.

The automatic inference of correspondences between two given 3D shapes — the correspondence problem — is a very challenging and at the same time relevant problem in shape matching as a large number of applications in computer graphics, among which are morphing, remeshing, surface editing and transfer of colors, normals, or texture between surfaces require the specification of homologous points. Furthermore, homologous points which are also often called landmarks form the fundamental basis for a statistical shape analysis [DM98]. While recently substantial progress could be achieved in the extrapolation of a small set of homologous points on the surface to dense correspondence maps [SAPH04, KS04], the manual specification of even a small number of homologous points per object quickly becomes cumbersome and time-consuming task especially if correspondences for more than two objects are needed which is usually the case in statistical shape analysis applications.

Our approach to partial matching adopts and extends some ideas from 2D image registration: For each object to be matched a number of salient points are selected from the surrounding space. A local shape descriptor centered at these salient points is subsequently evaluated and matched between two or more objects. We propose a robust matching algorithm that uses non-rigid thin-plate-spline (TPS) registration to identify potentially homologous subparts as sets of matching salient point pairs that can be mapped with low deformation measured in terms of TPS-bending energy. Based only on local shape information encoded by shape descriptors and on the relative geometrical configuration of salient points, the proposed method reliably detects

hb

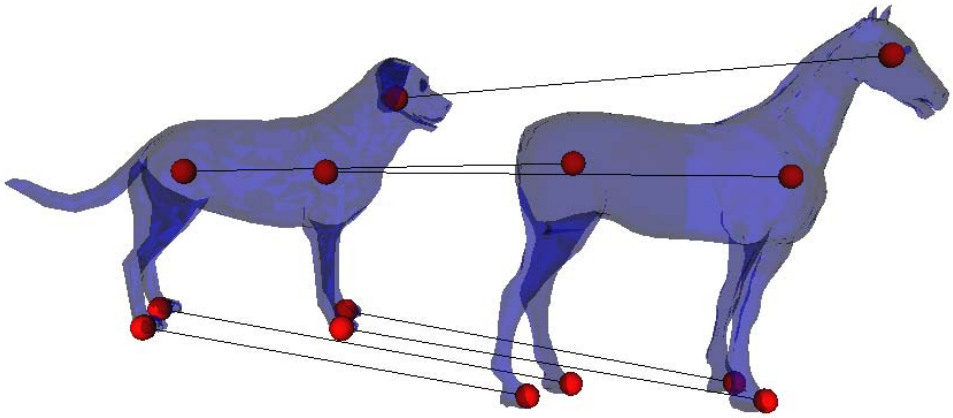

Figure 1: Corresponding salient points. As a first step, our method extracts *salient points* which are subsequently matched based on the shape of their vicinity and the thin plate spline based deformable matching model. An important feature of our algorithm is that it allows subpart matching, as well.

similarity between sub-shapes even if two shapes are dissimilar as a whole and is therefore robust with respect to insets and deformation.

2. Related Work

While there is an abundance of work on general object recognition and matching in both 2D and 3D, in the following overview we concentrate on correspondence based methods.

In the domain of 2D images a number of successful approaches have been presented in the recent years which tackle the problem of object detection in cluttered scenes. These methods extract so called *salient points* in the images and try to find correspondences between them. An excellent overview over the state-of-the-art on salient point detectors and local image descriptors can be found in [MTS*05] and [MS05], respectively. Generally, the object matching in these methods is invariant under affine transformations, a more complex deformation model is not considered. A more flexible model is utilized in [CJ04], where the authors introduce semi local geometric constraints to provide for a degree of robustness against deformations. On another thread of research, some effective approaches for deformable matching of point sets have been developed. In [CR03] a thin-plate spline (TPS) based deformable model has been elaborated for point sets along with a matching technique based on deterministic annealing. A similar technique presented in [BMP02] uses *shape contexts* to prune the point correspondences and subsequently uses TPS to model the deformations. Both methods rely on relatively dense sampling of the object features like contours resulting in several hundred points to be matched. Another common characteristic of the methods is that they are adopted to global object matching

not allowing for partial recognition and matching.

In 3D the landmark based deformable correspondence matching has been widely used in medical applications, see e.g. [FRS01]. Here the methodology to detect the landmarks and to match them is very mature, however, it is usually application specific, e.g. brain MRI or surface matching, etc. There are several techniques for extracting graphs from shapes, e.g. medial axis transform or 3D shock scaffold [LK01] or Reeb graphs [HSKK01], etc. While these techniques have the potential to partially match deformed articulated 3D shapes, they are strongly dependent on the topology of the objects. The matching of weighted point sets [TV03] relies on computing the earth mover's distance of two previously aligned attributed point sets extracted from the objects to be matched. In [FKS*04] a weighted chamfer distance is computed between the aligned objects, where the authors allow to define a weighting of query object surface, thereby facilitating the assignment of variable importance to different regions of the object. However, this method relies on the correct global pre-alignment of the objects and thus is not very versatile. A 3D generalization of the shape contexts was presented in [KPNK03] for the purpose of shape similarity estimation. In [FHK*04] the 3D shape contexts were adopted to the task of recognizing car geometries is scanned cluttered scenes. The idea here was to limit the support of the local shape contexts which made the approach suitable for part detection. However, a major limitation of this method is that it did not provide for scale invariance. Moreover, its versatility is limited by the lack of geometric constraints in the matching methods. The approach relies solely on the concept of representative shape contexts introduced in [MBM01] and compares the local shape descriptors in isolation.

3. General Setup and Notation

Our objective in this paper is to find the possibly multiple instances of a query object (the model) in a given scene (the data). We assume in the following that both model and scene are closed, i.e. they partition the space into interior and exterior areas and denote their interior by O and O' respectively. For both shapes we select a set of 3D salient points at positions $\mathbf{x}_i \in \mathbb{R}^3$, $i = 1,\ldots,N$ equipped with attributes: a radius (or scale) $r_i \in \mathbb{R}$, a unit length direction vector $\mathbf{p}_i \in \mathbb{R}^3$ and a local shape descriptor $\mu_i \in \mathbb{R}^M$. We will distinguish the attributed point sets in the model and the data by using primed and unprimed variables respectively. The unprimed point set X consisting of tuples $X = \{(\mathbf{x}_i, r_i, \mathbf{p}_i, \mu_i) | i = 1,\ldots,N\}$ stands for the model to be matched against the primed data set $X' = \{(\mathbf{x}'_i, r'_i, \mathbf{p}'_i, \mu'_i) | i = 1,\ldots,N'\}$. Put in other way, we want to find instances of the unprimed model in the primed data.

4. Salient Point Detection

The detection of salient points in a shape is a crucial ingredient of our approach. The choice of the "right" salient point detector is highly application dependent and special types of salient points may be designed for narrow domain problems. However, this choice is particularly challenging for general shapes. In spite of extensive literature on salient points for images, there are almost zero research results on this for 3D shapes. We therefore describe in this section the

generalization of a scale-space based image salient point detector by Lindeberg [Lin98] to 3D shapes.

4.1. Robust 2D Image Salient Points

Detecting salient feature points has a long tradition in image processing and computer vision. Salient points as e.g. Harris corners [HS88] are commonly used for image registration, image matching, object recognition and motion analysis. Focusing on image registration Mikolajczyk and Schmid evaluated in [MS05] the robustness of several salient points detectors and local image descriptors with respect to scale, illumination and viewpoint changes. The SIFT method [Low04] by Lowe was found to perform best. SIFT features are found as local extrema of the difference of Gaussian filter which approximates the scale normalized Laplacian of Gaussian (LoG) applied to an intensity image $f : \mathbb{R}^2 \to \mathbb{R}$:

$$L_{norm}(x,\sigma) := \sigma^2 \Delta G(x,\sigma) * f(x) \tag{1}$$

where G denotes the two dimensional Gaussian

$$G(x,\sigma) := \frac{1}{2\pi\sigma^2} e^{-\frac{\|x\|^2}{2\sigma^2}}$$

Computing salient points as local extrema of L_{norm} was initially proposed by Lindeberg [Lin98]. He showed that local extrema of L_{norm} correspond to Gaussian blob like structures in the image. Note that the extrema of L_{norm} are computed over both x and σ, i.e. we are looking for points $(\mathbf{x}_i, \sigma_i)$ that (locally) maximize or minimize L_{norm}. Salient points detected in this way have the important property that they come with a built-in scale σ_i. Lindeberg showed in [Lin93] that this scale is in fact equal to the radius of the detected Gaussian blob. This inherent scale of the salient points is essential for our approach as it allows us to choose local descriptors with an appropriate region-of-influence for each detected feature.

Another important feature of Lindebergs blob detection method is its robustness with respect to noise: Since the Laplacian and the convolution commute in equation 1 L_{norm} can be regarded as a smoothing step and subsequent differential operator. Introducing an additional smoothing step is common practice to regularize the (ill-posed) estimation of differentials from sampled data and results in robust differential operators. Therefore, the detected salient points are inherently robust with respect to noise or small deformations (see figure 2). Certainly features detected at larger scales are less affected by noise or high frequency deformation than small scale features.

4.2. Generalization to 3D Shapes

Lindebergs blob features generalize to three dimensional images $f : \mathbb{R}^3 \to \mathbb{R}$ in a straight forward way. As we are interested in features of three dimensional shapes we chose the following simple approach: A given closed shape S is converted into an image by computing its characteristic function χ_S, i.e. the binary function that equals 0 outside the shape and 1 in its interior. Salient feature points of shapes can then be simply computed as the extrema of L_{norm} on this three dimensional image.

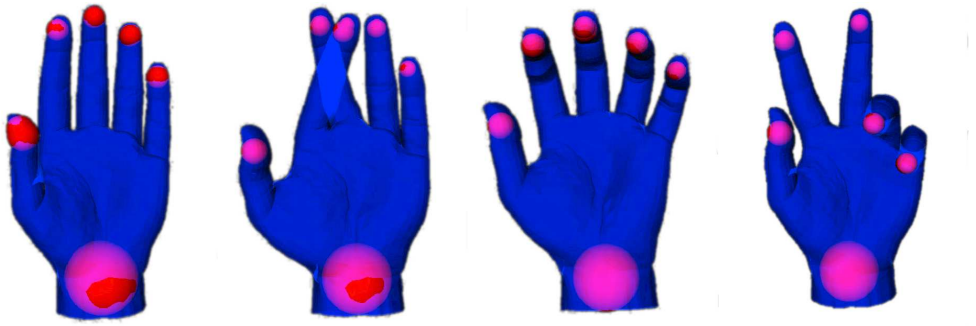

Figure 2: Salient points detected on a human hand in different poses. For each salient point a sphere with the detected radius is drawn within the transparent hand. Note that the detected points are very robust with respect to deformations

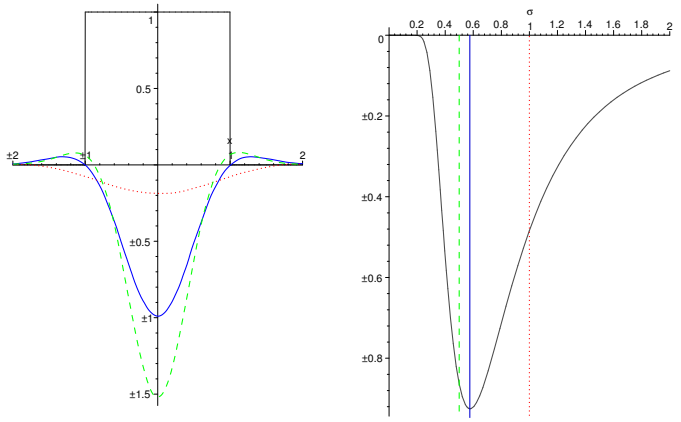

Figure 3: Left: Characteristic function of 3D unit ball and normalized Laplacian of Gaussian along X-axis. Right: Response of the 3D normalized LoG filter in the center of the 3D unit ball

Defined as the convolution of an image with a flipped Mexican hat shaped function (see figure 3), L_{norm} attains a high value at the center of a circular outside regions delimited by an inside region, i.e. high values indicate a high probability for an "outer blob". Vice versa, it attains a low value at the center of a circular inside region delimited by an outside region. Maxima of L_{norm} therefore correspond to outer blobs while minima correspond to inner blobs. As outer blobs are detected between different parts of the actual shape they are often less robust than inner blobs in particular when matching articulated shapes like gestures of a human hand as shown in figure 2. To obtain a small set of highly salient features we therefore chose to select only local minima of L_{norm} as salient points. Note however, that although minima correspond to inside blobs they might actually lie outside the shape. For example in the center of an annulus both an outside and inside blob are detected at different scales.

As mentioned above the scale normalization of L_{norm} was chosen so that it attains a unique local minimum at scale σ_0 for a 2D-Gaussian with $\sigma = \sigma_0$. However, a simple calculation shows that this is no longer true in three dimensions: for a 3D Gaussian a unique local minimum is

attained in the center at scale $\sigma = \sqrt{\frac{2}{3}}\sigma_0$. Furthermore, as we are dealing with binary characteristic functions, we are more interested in detecting true binary balls rather than Gaussian blobs. In order to select the correct scale we consider 3D ball B_r of radius r. Analogous to [Lin93], we observe that the spatial minimum of L_{norm} is assumed in the center of the ball and evaluates to

$$\begin{aligned} L_{norm}(0,\sigma) &= \sigma^2 G(x,\sigma) * \chi_{B_r} \\ &= \sigma^2 \int_{B_r} G(x,\sigma) \\ &= -\sqrt{\frac{2}{\pi}}(\frac{r}{\sigma})^3 e^{-\frac{r^2}{2\sigma^2}} \end{aligned}$$

Differentiating and solving $\frac{\partial}{\partial\sigma} L_{norm} = 0$ we find that the minimum is assumed for $\sigma = \frac{r}{\sqrt{3}}$ (see also figure 3). We therefore apply an additional scale adjustment assigning each salient point a radius of $\mathbf{r}_i = \sqrt{3}\sigma_i$ where σ_i denotes the scale at which the actual minimum of L_{norm} is found.

4.3. Computation

To compute Lindebergs blob-features in 2D intensity images efficiently an algorithm was proposed by Lowe [Low04] which is based on the following observation: Using the fact that the Gaussian satisfies

$$\frac{\partial}{\partial\sigma} G = \sigma \Delta G$$

the normalized LoG of an image f can be rewritten as

$$\begin{aligned} L_{norm}(x,\sigma) &= \sigma^2 \Delta G * f \\ &= \sigma(\frac{\partial}{\partial\sigma} G) * f \\ &\approx \sigma(\frac{G(x,k\sigma) - G(x,\sigma)}{k\sigma - \sigma}) * f \\ &= \frac{1}{k}(G(x,k\sigma) * f - G(x,\sigma) * f) \end{aligned}$$

where k denotes a fixed scale step factor. Therefore, the scale normalized Laplacian of Gaussian can be approximated by a simple differences between two Gaussian filtered versions of the initial image f. Lowe proposed to build an image pyramid consisting of Gaussian filtered versions of the input image by incrementally convolving with Gaussians, so that the resulting filtered images are separated by a constant factor k in scale. From this image pyramid a sequence of 2D difference images is computed and local extrema are detected by comparing each difference value with its neighbors both in spatial domain and scale dimension.

The above described algorithm can be generalized to three dimensional images in a straight forward way. However, as already discussed only minima in the difference images (instead of all local extrema) are selected by comparing each difference value to its neighbors.

Finally, we assign each detected salient point a unit-length local direction vector $\mathbf{p}_i \in \mathbb{R}^3$. As this direction is later used to orient local descriptors around each salient point $\mathbf{x}_i \in \mathbb{R}^3$ and

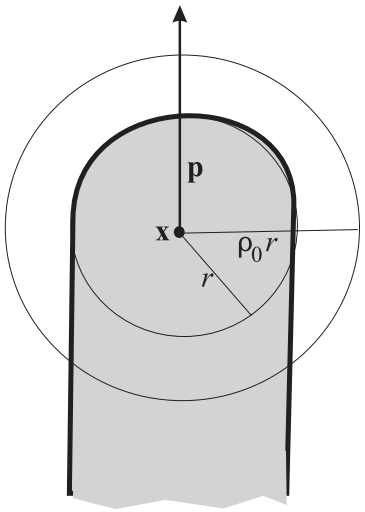

Figure 4: The local direction **p** points in the direction maximally outside of the object with respect to the location **x**.

to compute matching scores for a set of matches, it should also be descriptive and robust with respect to small deformations. To this extent we select for each salient point the direction that points in direction "as much outside" the shape as possible, cf. Fig. 4. More specifically, we consider for a salient point $\mathbf{x}_i$ with associated radius $r_i \in \mathbb{R}$ the integral of squared distances

$$D(x) := \int_{x' \in B(\rho_0 r_i, \mathbf{x}_i) \cap O} \|x - x'\|^2 \tag{2}$$

where $B(\rho_0 r_i, \mathbf{x}_i)$ denotes a ball with radius $\rho_0 r_i$ centered at $\mathbf{x}_i$ and $O \subset \mathbb{R}^3$ the object. We choose the direction of $\mathbf{p}_i$ as the gradient $\nabla D|_{x=\mathbf{x}_i}$, i.e. as the direction in which the distances to $B(\rho_0 r_i, \mathbf{x}_i) \cap S$ grow fastest. It is easy to see that $\mathbf{p}_i$ is given by

$$\mathbf{p}_i = \frac{\mathbf{x}_i - Mean(B(\rho_0 r_i, \mathbf{x}_i) \cap O)}{\|\mathbf{x}_i - Mean(B(\rho_0 r_i, \mathbf{x}_i) \cap O)\|} \tag{3}$$

There are certainly cases where this simple choice of $\mathbf{p}_i$ is not robust, e.g. for an isolated spherical shape. Therefore, we simply prune salient points for which the distance $\|\mathbf{x}_i - Mean(B(\rho_0 r_i, \mathbf{x}_i) \cap O)\|$ is below a threshold value.. In our implementation we set $\rho_0 = 1.5$.

5. Local Shape Descriptors

The purpose of local shape descriptors is to incorporate local geometry information in the vicinity of the salient points into the process of establishing the correspondences the salient point sets X and X'. Beside the descriptiveness with respect to the shape there is a number of requirements on these descriptors that are implicated by the very setting described in Section 1. The shape information incorporated into the descriptor should be localized around the position of

salient point detection within a certain radius. Furthermore, it should be invariant under certain transformations of the global shape like scaling and rotation.

Before we show how the above requirements are fulfilled, we describe the Spherical Harmonics (SH) and the SH based shape descriptors used e.g. in [FMK*03] which are the basis for further discussion.

5.1. Spherical Harmonics

We provide a brief overview of spherical harmonics with focus on basic formulations and invariance properties that are relevant with respect to our application; for more details we refer to [VMK88]. A spherical harmonic $Y_l^m(\vartheta,\varphi)$ is a scalar valued complex function in spherical coordinates where $\vartheta \in [0,\pi)$ denotes the colatitudinal and $\varphi \in [0,2\pi)$ the longitudinal coordinate. It is characterized by the parameters l and m, which take values $l = 0,1,\ldots$ and $m = -l,-l+1,\ldots,l-2,l-1,l$. Consequently, for each l there exist $2l+1$ functions with different m's. In explicit form, the spherical harmonics are expressed as a product of the Fourier exponential basis function $e^{im\varphi}$ with the associated Legendre polynomial P_l^m which is multiplied by a normalizing constant factor:

$$Y_l^m(\vartheta,\varphi) = e^{im\varphi} P_l^m(\cos\vartheta)\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}. \tag{4}$$

The behavior of SH under rotation is of particular interest with respect to our application. Let ϑ,φ and ϑ',φ' denote the position in the original and final coordinate systems, respectively, transformed according to an arbitrary rotation $\mathbf{R} \in SO(3)$. If the $2l+1$ spherical harmonics for the order l are collected into vectors $\mathbf{Y}_l(\vartheta,\varphi) = (Y_l^l(\vartheta,\varphi), Y_l^{l-1}(\vartheta,\varphi),\ldots,Y_l^{-l}(\vartheta,\varphi))^t$, the transformation of spherical harmonics can be expressed as

$$\mathbf{Y}_l(\vartheta',\varphi') = \mathbf{o}_l(\mathbf{R})\mathbf{Y}_l(\vartheta,\varphi) \tag{5}$$

where $\mathbf{o}_l(\mathbf{R})$ is a unitary matrix, i.e. it does not change the norms of vectors. Consequently, after projecting a function f defined on the unit sphere S^2 onto the functions of the vector $\mathbf{Y}_l$, we obtain invariant features μ_l of f by computing the norms of the so computed vectors:

$$\begin{aligned} \mu_l &= (\textstyle\sum_{m=-l}^{l} \langle f(\vartheta,\varphi), Y_l^m(\vartheta,\varphi)\rangle_{S^2}^2)^{1/2} \\ &= (\textstyle\sum_{m=-l}^{l} \langle f(\vartheta',\varphi'), Y_l^m(\vartheta,\varphi)\rangle_{S^2}^2)^{1/2} \end{aligned}$$

This equation is the basis for discarding phase information for SH descriptors in [FMK*03] and 3D Zernike descriptors in [NK04]. However, throwing away the phase also implies information loss. In order to alleviate this problem we define a variant of the features that are invariant under rotation around a specific axis as opposed to a general rotation. It is well known that the amplitudes of Fourier coefficients are invariant to rotation, i.e. for a function $g(\varphi)$ with $\varphi \in [0,2\pi)$ we have

$$|\langle g(\varphi), e^{im\varphi}\rangle_{S^1}| = |\langle g(\varphi+\varphi_0), e^{im\varphi}\rangle_{S^1}|.$$

Getting back to spherical harmonics, suppose we have a rotation around the z-axis of the coordinate system, or equivalently, we change only the longitudinal coordinate φ of a function $f(\vartheta,\varphi)$.

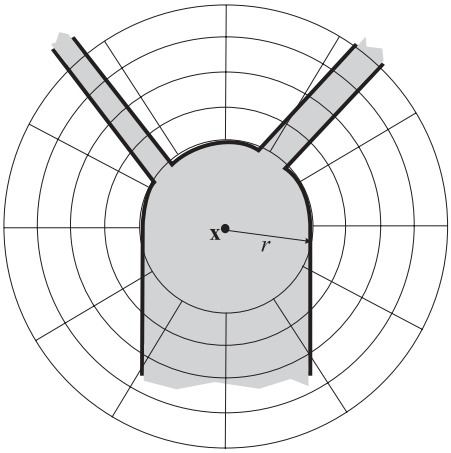

Figure 5: 2D sketch of the support region for local descriptor computation. r denote the local radius, $\mathbf{x}_i$ is the position of the salient point and the outermost circle corresponds to R^{max}.

According to Eqn. 4, the dependency on φ is expressed as a Fourier exponential function, therefore, the absolute value of the SH coefficients will not change under such rotation:

$$|\langle f(\vartheta,\varphi), Y_l^m(\vartheta,\varphi)\rangle_{S^2}| = |\langle f(\vartheta,\varphi+\varphi_0), Y_l^m(\vartheta,\varphi)\rangle_{S^2}|. \tag{6}$$

We note that the harmonic 3D shape context descriptors presented in [FHK*04] are based on the above observation.

5.2. Local SH Descriptors

The basis for the construction of our local shape descriptors are the SH descriptors from [FMK*03]. The three dimensional space around the origin is sampled as concentric shells, where the shells are defined by n equal radial intervals with radius $r \in [0, R_{max}]$. Subsequently, the shells are discretized into equiangular bins along ϑ and φ, and define a binary spherical function that equals 1 if there is an object point in such a bin and 0 otherwise. The object representation consists of a spherical harmonic decomposition for each shell. We modify these descriptors to improve descriptiveness, get scale and rotation invariant local shape descriptors.

We recall that for the ith salient point we have its position $\mathbf{x}_i \in \mathbb{R}^3$, the associated unit length principal direction vector $\mathbf{p}_i \in \mathbb{R}^3$ and the radius of detection $r_i \in \mathbb{R}$. In a nutshell, $\mathbf{x}_i$ is used to obtain translation invariance, $\mathbf{p}_i$ to orient the coordinate system in order to utilize Eqn. 6, and r_i is used to scale the shell radii which results in scale invariance. The concentric shells are centered at $\mathbf{x}_i$ with radius range between R_i^{min} and R_i^{max}. As described in Section 4, the scale of salient point detection corresponds to the size of detected structure (the radius of fitted spherical blob), and therefore, there is very little discriminating information within a radius of $\mathbf{r}_i$ around $\mathbf{x}_i$. Hence, we set $R_i^{min} = r_i$ and $R_i^{max} = \rho \cdot r_i$, with $\rho \geq 1$. The situation is sketched as a 2D example in Fig. 5. Finally, we compute the shape descriptors that are invariant under rotation around the principal direction $\mathbf{p}_i$. To achieve this, we compute the transformation $\mathbf{R}$ rotating

the z-axis onto $\mathbf{p}_i$ and determine the matrix $\mathbf{o}_l(\mathbf{R})$ from Eqn. 5. Subsequently, we compute the projections of discrete spherical functions for each shell onto the SH-basis, rotate the coefficients according to Eqn. 5 and compute the absolute values of the resulting coefficients to obtain the axial rotation invariance, as described by Eqn. 6. We denote the local descriptors computed in this way as $\mu_i \in \mathbb{R}^M$.

We experimentally found the following setting to work well: ρ is set to 4, we used an SH basis with order $l \leq 8$ and 8 concentric shells resulting in $M = 45 \cdot 8 = 360$ dimensional vectors as local shape descriptors. Note that we included only coefficients with $m \geq 0$ into the descriptors; this is due to the fact that for real functions the projections onto Y_l^m and Y_l^{-m} are related by a negation and complex conjugate, thus making the corresponding absolute values equal.

6. Salient Point Set Matching

With salient point sets X and X' in the model and data at hand the task of matching consists in a selection of "correct" matches from the set off all possible correspondences $C = X \times X'$. However, as the salient point attributes $\mathbf{x}_i$,r_i,$\mathbf{p}_i$ and μ_i capture only local shape properties considering each correspondence $c \in C$ by itself is not likely to yield a correct set of correspondences. The salient points at the fingertips in figure 2 for example carry nearly identical attributes so that a correct match (forefinger maps to forefinger) can hardly be distinguished from a false match (forefinger maps to middle finger).

We therefore pursue a different strategy: Instead of single correspondences we consider sets of correspondences $S \subset C$. For each set S we compute the thin-plate-spline deformation $f_S : \mathbb{R}^3 \to \mathbb{R}^3$ (described in section 6.1) that maps the salient points in the model onto the corresponding points in the data. The correspondence map f_S is then used to estimate the likelihood that the correspondences $c \in S$ are correct, i.e. are pairs of homologous points.

More precisely, we define a set of correspondences $S \subset C$ as correct if it satisfies the following constraints:

1. The shape descriptors match for all correspondences in S, i.e. we have

$$d_d(S) := \max_{(i,j)\in S} \|\mu_i - \mu'_j\| < t_d \tag{7}$$

2. A set of correct correspondences S results in correspondence map f_S that exhibits a low amount of space "bending". The notion of bending is stated more precisely in section 6.1 and measured in terms of the TPS bending energy $\Delta(f_S)$, i.e.

$$\Delta(f_S) < t_\Delta \tag{8}$$

   The intuition behind that constraint is as follows: Given that a set of salient points P in the model are homologous to a set of points P' in the data only little space bending is needed beyond moving, scaling and rotating to align P with P'.

3. If the correspondences in S are correct f_S does not only map the salient points of S in the model onto their corresponding points in the data, but also all other points in the model.

We measure the extent to which this assumption is fulfilled by the chamfer distance d_c between the deformed model $f(O)$ and the data O' (described in section 6.2.3). Therefore we enforce

$$d_c(S,O,O') < t_c \tag{9}$$

As described above there are possibly multiple instances of the model in the data. Instead of a single correct set S we are rather looking for a family of correct sets $\mathscr{S} = (S_i)_{i=1..n}$. Each S_i must be functional regarded as a binary relation, i.e. each salient point in the model must map to at most one point in the data since otherwise the bending energy is infinitely high. Furthermore, we are certainly only interested in maximal sets of correct matchings i.e. for each S_i and $c \in C$ we have that $S_i \cup \{c\}$ is not correct. We can now cast the matching problem as follows: Given the two sets X and X' of salient points compute a family $\mathscr{S}$ of maximal correct sets and an enumeration of $\mathscr{S}$ such that $|S_i| \geq |S_j|$ for $i < j$.

Clearly, due to the size of the power set $|\mathscr{P}(C)| = 2^{|C|}$ that has to be searched for correct sets S_i an exhaustive search is impractical. After giving a description of the utilized TPS model for deformation we therefore present in subsection 6.2 a matching algorithm along with a set of heuristics to approximate the ideal solution to the above stated matching problem.

6.1. Thin Plate Splines and Their Use

A crucial ingredient of our matching algorithm is the usage of thin-plate splines (TPS) as a mapping between the point sets. There are several method utilizing TPS which have found wide usage since the seminal paper of Bookstein [Boo89]. TPS have been used in medical analysis, as well as in 2D object recognition [BMP02] or even registration 3D point clouds [BR04]. The novelty of our approach lies in the utilization of the total curvature of the mapping as a measure of compatibility between sets of correspondences which is the building block of our algorithm to find the sets of desired correspondences.

The goal is to compute a non-rigid transformation $f : \mathbb{R}^3 \to \mathbb{R}^3$ that maps the salient points $\mathbf{x}_i$ onto their corresponding pairs $\mathbf{x}'_i$ with $i = 1,\ldots,N$. As the mapping is non-rigid, there are an infinite number of transformations mapping these points onto each other. Consequently, further constraints have to be introduced. A natural demand is to ask the mapping to be smooth, which is fulfilled by the definition in terms of thin plate splines which minimize the total curvature:

$$I_{f^*} = \int_{\mathbb{R}^3} \Delta^2 f^*(\mathbf{x}) d\mathbf{x},$$

where $* \in \{x,y,z\}$, meaning that the total curvature of each coefficient of the vector valued function f is minimized. The solution function, or TPS interpolant has the form

$$f(\mathbf{x}) = \begin{bmatrix} f^x(\mathbf{x}) \\ f^y(\mathbf{x}) \\ f^z(\mathbf{x}) \end{bmatrix} = \mathbf{A}\mathbf{x} + \mathbf{t} + \sum_{i=1}^{N} \begin{bmatrix} w_i^x \\ w_i^y \\ w_i^z \end{bmatrix} \phi(\|\mathbf{x} - \mathbf{x}_i\|)$$

where $\phi(r) = |r|$. As it can be seen, the TPS interpolation can easily be divided into an affine part characterized by the 3×3 matrix $\mathbf{A}$ along with translation $\mathbf{t}$ and the non-affine part computed as

the products of warping coefficients w_i^* with the TPS kernel function ϕ. The unknowns of the above equation can be determined by solving the following system of linear equations:

$$\begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{O}_1 \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \mathbf{t} \\ \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \\ \mathbf{O}_2 \end{bmatrix}$$

where $\mathbf{K}_{ij} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $\mathbf{O}_1$ and $\mathbf{O}_2$ are 4×4 and 4×3 matrices of zeros, respectively. The rows of the $N \times 4$ matrix $\mathbf{P}$ are of the form $(1, x_i, y_i, z_i)$, $\mathbf{W}$ is a $N \times 3$ matrix with rows (w_i^x, w_i^y, w_i^z), and $\mathbf{V}$ is a matrix of target point coordinates with rows (x_i', y_i', z_i'). The last 4 rows of the system are needed to guarantee the square integrability of the second order derivatives, see e.g. [DM98] for more details. Let $\mathbf{L}$ denote the $N+4 \times N+4$ matrix of the above system and $\mathbf{L}_N^{-1}$ the upper $N \times N$ block of $\mathbf{L}^{-1}$. It can be shown that for the sum of total curvatures for each coefficient of f the following proportionality relation applies:

$$\Delta(f) = \sum_{*\in\{x,y,z\}} I_{f^*} \propto trace[\mathbf{V}^t \mathbf{L}_N^{-1} \mathbf{V}] = trace[\mathbf{W}^t \mathbf{K} \mathbf{V}].$$

Thus, we are able to quantify the amount of bending needed to warp the 3D space to transform the corresponding points onto each other by a maximally smooth mapping. This quantity provides a measure on the structure change within the point set, i.e. how far the points have moved with respect to each other; roughly said, the larger the movement, the more bending it causes.

So far, we have only considered the geometrical positions $\mathbf{x}_i$ of points to be mapped. Therefore, the question remains open how to incorporate other salient point information into the interpolation like the radius r_i and local direction $\mathbf{p}_i$. To facilitate this, further constraints can be introduced on the TPS map: beside the one to one mapping of point positions, the derivatives of the map should respect the scale and direction information. As both of these attributes obey the desired invariance under global scaling and rotation, a simple way to approximate the derivative constraints is to use them to generate new points which will be used as further constraints in the computation of the TPS mapping similar in spirit to normal constrains in [TO99]. The new points are generated for each salient point by shifting the original points in the direction of the principal direction by a length corresponding to the radius of the salient point:

$$\mathbf{y}_i = \mathbf{x}_i + r_i \cdot \mathbf{p}_i.$$

An important implication of this trick is that we are able to deal with the geometrical constraints including scale and local orientation together with the position of the salient point simply by utilizing TPS.

6.2. Matching algorithm

In the following we present an algorithm that aims at approximating the solution to the matching problem as formulated in the beginning of this section. Our algorithm proceeds in four steps: In the first step we limit our search space by pruning all subsets $S \in \mathscr{P}(C)$ that contain correspondences failing to satisfy equation 7. Although 3 correspondences are sufficient to uniquely determine a TPS deformation we found that sets of only three correspondences are unlikely to

define good matchings. We therefore restrict our search to sets of at least four correspondences. The second step creates a set of such seed quadruples $Q_i \subset C$ which are augmented to maximal correspondence sets S_i in a third step. In the last step incorrect steps are pruned and the remaining maximal sets are sorted. The individual steps are now described in detail:

6.2.1. Pruning and initial quadruples of correspondences

In a first straightforward step we prune the set of all possible correspondences C, such that all the remaining correspondences obey the constraint of Eqn. 7. We denote the resulting set of correspondences $\hat{C}$; in the following only the elements of $\hat{C}$ are considered. From $\hat{C}$ we then select correct seed quadruples Q_i. As the order of set $\hat{C}$ may still be large after the pruning, a naïve verification of correctness for all possible $O(|\hat{C}|^4)$ quadruples often proves to be impractical. We therefore consider here only quadruples, for that the shapes of mapped tetrahedrons and the relative local radii of corresponding salient points are similar. In addition, as the direction vectors $\mathbf{p}$ incorporate important local shape information, we require that they are consistent, as well. The exact selection scheme is described in Appendix A.

6.2.2. Maximal correspondence sets

With a set of seed quadruples Q_i at hand we try to grow maximal correct correspondence sets. However, even after the pruning in the previous steps the number of quadruples might still be to high to grow all of them. We therefore resort to an heuristic that selects the most promising quadruples by assigning a score which should be high for the quadruples that are likely to include correct correspondences (see Appendix B for the details on the cost function). This score is computed as a combination of local descriptor distances, condition number of the affine transformation which maps the point positions onto each other and the local direction consistency of the corresponding salient points. It is clear that more similar local shapes and consistent local directions imply a higher likelihood that the correspondences of the quadruple are correct.

The initial quadruples are now considered for growing in order of increasing cost function up to a given bound on the cost function and/or the total number of initial tuples. Again, a complete test for correctness during the growing is computationally too expensive. We therefore use the following heuristic to compute sets that are likely to be correct:

1. For each correspondences $c_j \in \hat{C}$ we compute a compatibility with respect to an initial quadruple Q_i by adding it as fifth correspondence to the initial quadruple $Q_i^j = Q_i \cup \{c_j\}$ and computing the deformation magnitude $\Delta(Q_i^j)$. A correspondence c_i with $\Delta(Q_i^j) < t_\Delta$ is considered compatible.

2. All compatible correspondences are candidates for an augmentation of Q_i. However, a simple augmentation of Q_i by all compatible c_j is not appropriate: In many cases an augmentation of Q_i by two compatible correspondences results in a intolerable high TPS deformation. Therefore, we define an undirected graph $\mathscr{G}_i(V_i, E_i)$ with the vertices corresponding to the c_i with $\Delta(Q_i^j) < t_\Delta$. Two vertices v_p and v_q are connected by an edge if $\Delta(Q_i^{pq}) < 2 \cdot t_\Delta$, where $Q_i^{pq} := Q_i \cup \{c_p\} \cup \{c_q\}$ is constructed by adding the two correspondences to the initial quadruple.

3. Finally, we compute maximal cliques (complete subgraphs) in $\mathscr{G}_i$ as candidates for a maximal correct sets S_i

6.2.3. Final correspondence sets

At this point we have computed a number of candidates for maximal correct sets. In this last step we simply prune incorrect sets S_i from the family $\mathscr{S}$. We use the TPS mapping $f_{S_i} : \mathbb{R}^3 \to \mathbb{R}^3$ to map the model O, of which we are looking for instances in the data O', and measure the distance between them. The formulation is as follows:

$$d_c(f_{S^j}(O), O') = \int_O \inf_{\mathbf{x}' \in O'} \{\| f_{S^j}(\mathbf{x}) - \mathbf{x}' \|\} d\mathbf{x}.$$

Thus, d_c is the chamfer distance of the mapped O and O'. To be efficient, we use the discrete distance transform $DT(O')$ of O', and the integration is in fact a summation over the interior voxels of O which get mapped into $DT(O')$ and the corresponding distance values are simply accumulated. With this last step we ensure that the computed correspondences and the resulting mapping really registers the model with the data – for parts not set into direct correspondence by means of salient points.

7. Results

Before we present some results generated using our approach, some comments on the realization are in order. We used voxel grids of 128^3 to compute the salient point positions. As it becomes clear from the previous sections, like in most of non-trivial pattern recognition problems, there is a number of parameters involved. We already provided a suggested value for each threshold value, we now give some hints on their impact. The challenge is to set the thresholds, so that the search space for the algorithm is limited enough, however, if set incorrectly, no correspondences may be found. Essentially, there are two parameters that have large influence on the performance: the tolerance on the descriptor distance t_d and on the total curvature based correspondence compatibility measure t_Δ. For the examples presented in this paper these were set to $t_d = 250$ and $t_\Delta = 25$. It may be desirable to set the tolerances higher for some applications, though. In such cases longer runtimes have to be taken into account. For the examples presented, the runtime was up to 5 seconds, which was higher for more complex objects involving matching of larger sets of salient points.

We focus on the following applications of our approach: (i) matching of whole objects, where substantial deformation may be present and (ii) matching of sub-shapes and finding the symmetries. Fig. 6 demonstrates the ability of our technique to match whole objects. Notably, the method finds sets of correct correspondences in spite of considerable complexity of the objects or large differences in the global shape. As it can be seen also in Fig. 1, our method is versatile with respect to object classes.

The algorithm presented in Section 6.2 aims at finding the best matching subsets of the model and the data. Therefore we are able to find subparts of a given object in a target scene. This ability of our method is demonstrated in Fig. 7. As it can be seen, symmetries and self-similarities of the object are recovered, as well. Thanks to our TPS based flexible matching, we are able

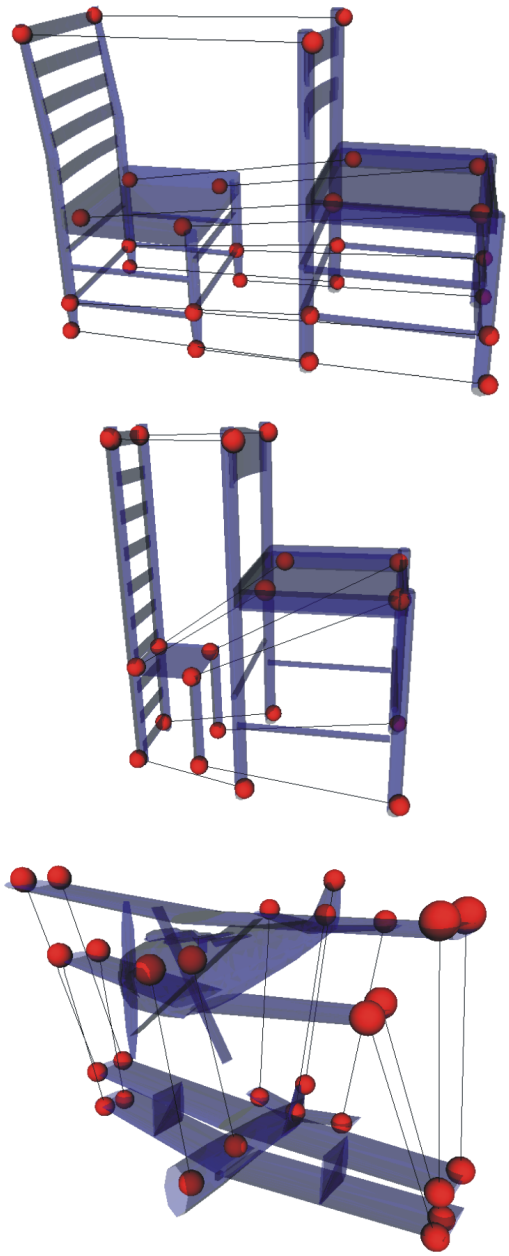

Figure 6: Whole object matches. In the left side example it can be seen that as many salient point matches are found, as possible. In the middle, the matches are correct, although there is a considerable difference in the global shapes of the objects. The example on the right side demonstrates the suitability of our approach for more complex objects, as well.

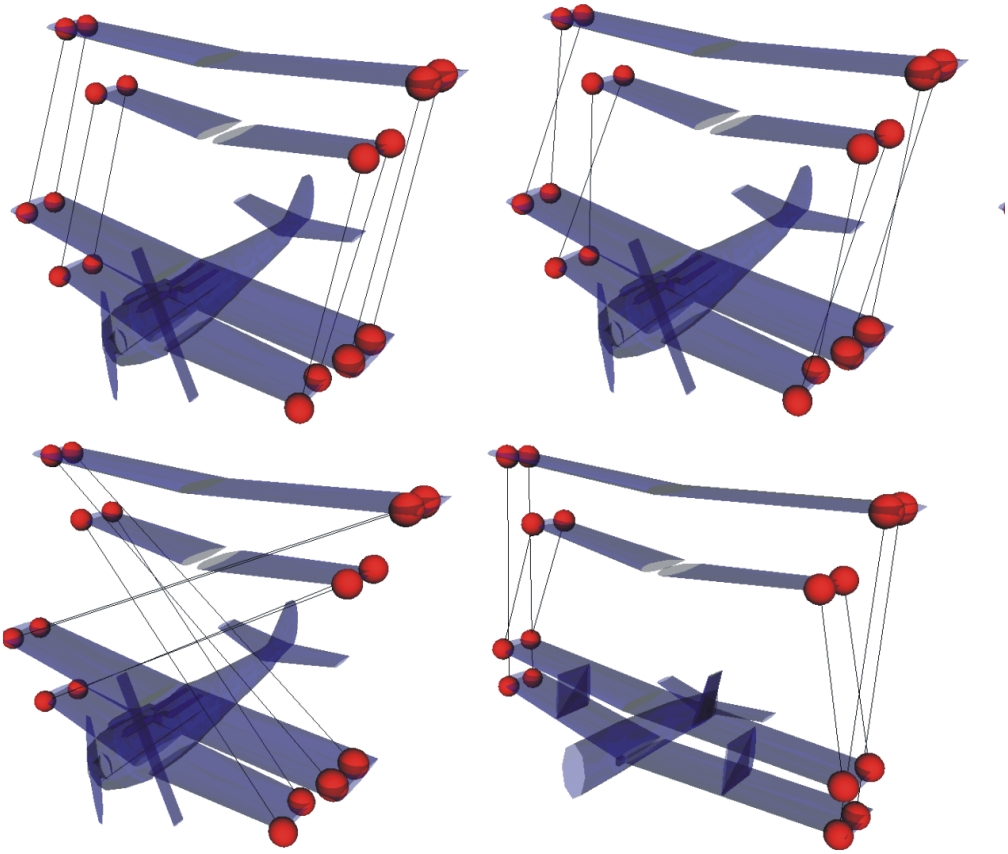

Figure 7: Partial matching. The separate wings of a biplane are detected in the full model; note that symmetries are recovered, as well. In the rightmost image a match against different biplane is shown having considerably different wings: the wing planes are closer and their geometry is similar as opposite to the above model, where the lower wing is shorter.

to find the query object even under considerable deformation, as demonstrated by the rightmost image.

8. Conclusion

We presented a method for correspondence based partial matching of 3D shapes. Our approach relies on extracting a set of stable salient points and subsequently finding sets of correspondences in the objects to be matched. The utilized model for matching incorporates local shape information and allows for recovering partial matches even under substantial deformation.

We believe that the technique presented in this paper has the potential to serve as a starting point to adoption and extension of the presented matching approach to suit a number of applications. These include generation of dense correspondences between shapes for e.g. statistical shape analysis or morphing, retrieval of parts of 3D shapes and shape completion algorithms

that exploit self-similarities in a shape to fill in missing parts or holes.

A. Computation of correspondence quadruples

Let us consider all the possible pairs of elements in $\hat{C}$. For a pair $c_p, c_q \in \hat{C}$, we have two points from X and X' each equipped with positions $\mathbf{x}_p$, $\mathbf{x}_q$ and $\mathbf{x}'_p$, $\mathbf{x}'_q$ and local radii r_p, r_q and r'_p, r'_q. As discussed in Section 4, the local radii r reflect the size of the local structure around the salient points which should be scaled in a similar manner as the distance between the salient points under the mapping which registers the model with the data. Hence, it is reasonable to assume that for similar shapes the proportions of distances between the positions and the local radii should be about the same:

$$\frac{\|\mathbf{x}_p - \mathbf{x}_q\|}{\|\mathbf{x}'_p - \mathbf{x}'_q\|} \approx \frac{r_p}{r'_p} \approx \frac{r_q}{r'_q}.$$

We use the above observation as a basis for constructing an undirected graph $\mathscr{G}(V,E)$ with vertices assigned to correspondences $V = \{v_i | i = 1 \dots |\hat{C}|\}$. We compute a compatibility measure for each pair of correspondences based on the above assumption:

$$w_{ij} = \max\left(\left| \frac{\|\mathbf{x}_p - \mathbf{x}_q\|}{r_q} - \frac{\|\mathbf{x}'_p - \mathbf{x}'_q\|}{r'_q} \right|, \left| \frac{\|\mathbf{x}_p - \mathbf{x}_q\|}{r_p} - \frac{\|\mathbf{x}'_p - \mathbf{x}'_q\|}{r'_p} \right| \right)$$

If $w_{ij} < t_w$ for a prescribed threshold t_w (set to 0.4), the vertices v_i and v_j are connected by an edge. From the graph defined in such a way we generate the quadruples by computing cliques (complete subgraphs) of size 4.

We further prune the correspondence quadruples by enforcing the local direction consistency. To achieve this, for a quadruple $Q \subset \hat{C}$ we first compute the affine mapping $\mathbf{A}_Q$ which maps the point positions according to the correspondences, and apply this mapping to the local directions $\mathbf{p}$ for all 4 salient points of the model. If the maximal angle $d_a(Q)$ between the directions of the mapped model and those at the data points is larger then a threshold t_a then the quadruple is discarded. In our experiments we set the maximal allowable angle difference $t_a = 0.3$.

B. Correspondence quadruple score

There are many conceivable score functions measuring the likelihood that a quadruple Q^j contains correct correspondences. In our experiments we found the following choice to perform well:

$$\varepsilon(Q^j) = \exp\left(\frac{d_d^2(Q^j)}{2\sigma_d^2}\right) \cdot \left(1 - \frac{d_a(Q^j)}{t_a}\right) \cdot I_{1/k(\mathbf{A}_{Q^j}) > t_k}$$

Here d_d denotes the maximal descriptor distance of the quadruple, cf. Eqn. 7, we empirically found the value of $\sigma_d = 130$ to work well. $d_a(Q^j)$ and the threshold t_a are defined and set in Appendix A. $k(\mathbf{A}_{Q^j})$ denotes the condition number of the affine transformation matrix computed as in Appendix A, we set $t_k = 0.15$; the binary valued function I is 1 if and only if the condition in the index is true.

References

[BMP02] BELONGIE S., MALIK J., PUZICHA J.: Shape matching and object recognition using shape contexts. *IEEE Trans. on PAMI 24*, 4 (2002), 509–522.

[Boo89] BOOKSTEIN F. L.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on PAMI 11*, 6 (1989), 567–585.

[BR04] BROWN B., RUSINKIEWICZ S.: Non-rigid range-scan alignment using thin-plate splines. In *Symposium on 3D Data Processing, Visualization, and Transmission* (September 2004).

[CJ04] CARNEIRO G., JEPSON A.: Flexible spatial models for grouping local image features. In *CVPR04* (2004), pp. II: 747–754.

[CR03] CHUI H., RANGARAJAN A.: A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst. 89*, 2-3 (2003), 114–141.

[DM98] DRYDEN I., MARDIA K.: *Statistical Shape Analysis*. John Wiley & Sons, New York, 1998.

[FHK*04] FROME A., HUBER D., KOLLURI R., BULOW T., MALIK J.: Recognizing objects in range data using regional point descriptors. In *Proceedings of ECCV* (May 2004).

[FKS*04] FUNKHOUSER T., KAZHDAN M., SHILANE P., MIN P., KIEFER W., TAL A., RUSINKIEWICZ S., DOBKIN D.: Modeling by example. *ACM Trans. Graph. 23*, 3 (2004), 652–663.

[FMK*03] FUNKHOUSER T., MIN P., KAZHDAN M., CHEN J., HALDERMAN A., DOBKIN D., JACOBS D.: A search engine for 3d models. *ACM Trans. Graph. 22*, 1 (2003), 83–105.

[FRS01] FORNEFETT M., ROHR K., STIEHL H.: Radial basis functions with compact support for elastic registration of medical images. *IVC 19*, 1-2 (January 2001), 87–96.

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detector. In *Proceedings of the Fourth Alvey Vision Conference* (1988), The University of Sheffield Printing Unit, pp. 147–152.

[HSKK01] HILAGA M., SHINAGAWA Y., KOHMURA T., KUNII T. L.: Topology matching for fully automatic similarity estimation of 3D shapes. In *Proceedings of ACM SIGGRAPH* (2001).

[KPNK03] KÖRTGEN M., PARK G.-J., NOVOTNI M., KLEIN R.: 3d shape matching with 3d shape contexts. In *The 7th Central European Seminar on Computer Graphics* (April 2003).

[KS04] KRAEVOY V., SHEFFER A.: Cross-parameterization and compatible remeshing of 3d models. *ACM Trans. Graph. 23*, 3 (2004), 861–869.

[Lin93] LINDEBERG T.: On scale selection for differential operators. In *ISRN KTH* (1993).

[Lin98] LINDEBERG T.: Feature detection with automatic scale selection. *Int. J. Comput. Vision 30*, 2 (1998), 77–116.

[LK01] LEYMARIE F., KIMIA B.: The shock scaffold for representing 3d shape. In *Proc. of 4th International Workshop on Visual Form (IWVF4)* (2001).

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60*, 2 (2004), 91–110.

[MBM01] MORI G., BELONGIE S., MALIK H.: Shape contexts enable efficient retrieval of similar shapes. In *CVPR 1* (2001), pp. 723–730.

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors, 2005. To appear in IEEE Trans. on PAMI.

[MTS*05] MIKOLAJCZYK K., TUYTELAARS T., SCHMID C., ZISSERMAN A., MATAS J., SCHAFFALITZKY F., KADIR T., GOOL L. V.: A comparison of affine region detectors, 2005. To appear in International Journal of Computer Vision.

[NK04] NOVOTNI M., KLEIN R.: Shape retrieval using 3d zernike descriptors. *Computer Aided Design 36*, 11 (2004), 1047–1062.

[SAPH04] SCHREINER J., ASIRVATHAM A., PRAUN E., HOPPE H.: Inter-surface mapping. *ACM Trans. Graph. 23*, 3 (2004), 870–877.

[TO99] TURK G., O'BRIEN J. F.: Shape transformation using variational implicit functions. In *SIGGRAPH '99* (1999), pp. 335–342.

[TV03] TANGELDER J. W., VELTKAMP R. C.: Polyhedral model retrieval using weighted point sets. *International Journal of Image and Graphics 3*, 1 (2003), 1–21.

[VMK88] VARSHALOVICH D., MOSKALEV A., KHERSONSKII V.: *Quantum Theory of Angular Momentum*. World Scientific, 1988.