# Correspondences between Salient Points on 3D Shapes

Raoul Wessel[1], Marcin Novotni[2], Reinhard Klein[3]

Computer Graphics Group
University of Bonn, Germany
Email: {wesselr[1], rk[3]}@cs.uni-bonn.de
novotni.marcin@evosoft.hu[2]

## Abstract

Establishing correspondences between salient points on 3D shapes results in functions mapping similar parts of 3D objects. In this paper, we present a new method for establishing binary correspondences between salient points on 3D shapes. Our algorithm is independent of the shape representation and the object topology and does not require any prepositioning of the objects. Our method first detects stable salient points that are representative for certain parts of the shape. For each of these salient points it then computes an associated local shape descriptor. We introduce a matching energy between the salient points of two shapes which depends on the similarity of the descriptors and the spatial relationship of the salient points. An iterative optimization scheme determines a correspondence mapping between the salient points which minimizes this energy. The resulting binary correspondences between the salient points can be used for applications like 3D shape retrieval based on similarity estimation, classification of 3D objects, editing, and statistical shape analysis. It is especially useful as an initialization method for approaches relying on prior knowledge about corresponding points like cross parameterization or morphing.

## 1 Introduction

The automatic inference of correspondences between small sets of salient points of 3D shapes is a crucial ingredient for numerous computer graphics applications. While recently substantial progress could be achieved in the extrapolation from a small set of corresponding points on the surface to dense correspondence maps [SAPH04, KS04], the manual specification of even a small number of corresponding points quickly becomes a cumbersome and time-consuming task especially if correspondences for more than two objects are needed. Examples for applications relying on correspondences between 3D shapes are:

- **3D shape retrieval and classification**: Both of these methods require a similarity measure between two shapes. Given corresponding salient points, comparing the local geometry around them delivers this measure.
- **Statistical shape analysis**: Common statistical tools like principal component analysis require an alignment of so-called *landmarks* which can be obtained by the corresponding salient points.
- **Dense matchings**: Methods like cross parameterization or morphing require knowledge about at least some corresponding points in both shapes. Up to now, these correspondences are typically defined manually.

Approaches to establish correspondences between salient points on shapes face three major problems:

1. **Detection of robust salient points**: Determine those points, which are significant for a certain region of the object. Robustness towards noise, i.e. small changes in the structure of the object and towards different scales must be guaranteed.
2. **Distance measure between salient points**: Characteristics of the detected salient points must be captured. A measure allowing for the comparison of two salient points according to these characteristics has to be introduced.
3. **Establishing correspondences**: Using the distance measure, corresponding points need to be found.

Typically, the detection of salient points and the distance measure are closely connected. Most approaches to this task define the distance of two

salient points by the similarity of associated local descriptors. The performance of various local descriptors in 2D images was evaluated in [MS03]. The SIFT method [Low04] and the SIFT-based GLOH method [MS03] were found to perform best. Both approaches rely on scale invariant salient point detectors providing the information on which scale the salient points were detected. This scale is a crucial ingredient for the construction of any descriptor relying on local geometry. It provides information about the size of the detected geometric feature and therefore determines the size of the area around the salient point that must be captured by the local descriptor. The SIFT method detects salient points as local extrema of the scale-space difference-of-Gaussian function. These extrema refer to so-called blob-like structures [Lin98]. In our framework, we generalize this successful method to three dimensions in a straightforward way.

Instead of using the gradient-based descriptors, which are designed and used for 2D applications in [Low04], we modify the *Spherical Harmonics* (SH) shape descriptors [FMK*03] for our purpose. Like the SIFT features in 2D, these descriptors also rely on a scale determined by the salient point detector and provide the ability to be invariant under rotation, scaling and translation.

Given the salient points and a distance measure between them, the problem of establishing the correspondences between the salient points remains. We therefore consider the approach of Chui [CR03], in which a point matching algorithm for non-rigid registration is developed. The method is based on a stochastic optimization algorithm that minimizes a matching energy function based on the spatial relationship of both point sets. It determines a non-rigid transformation between objects as well as correspondences between points. A great advantage of this method is that no prior knowledge about the object orientation is needed. Although providing excellent results for 2D and 3D point sets, it cannot be applied directly to the above defined problem as it relies on a dense sampling of the object. Given only a few salient points representing the object, there are too few constraints preventing the algorithm from choosing a transformation that destroys the structure of the object, such that the results become rather poor.

We therefore adopt a different strategy incorporating knowledge about the similarity of the salient points and their spatial relationship. We modify the matching energy function used in [CR03] extending it by the distance measure provided by the SH descriptors. This additional information allows us to use the stochastic optimization algorithm even if we are only dealing with very few salient points.

## 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. They extract salient points in the images and determine correspondences between them. An overview of the state-of-the-art on salient point detectors and local image descriptors can be found in [MTS*05] and [MS05]. The object matching procedure in these methods is invariant under affine transformations, a more complex deformation model is not considered. A more flexible model is used 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 matching of point sets have been developed, amongst which the iterative closest point (ICP) [BM92] algorithm is probably the most well known. There, an iterative optimization scheme is used to compute binary correspondences at each iteration. Deformable matchings are examined in [CR03]. This matching technique uses deterministic annealing to compute a thin-plate spline (TPS) [Boo89] interpolation and the resulting correspondences between two point sets. 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 samplings of the object features like contours resulting in several hundred points to be matched. In [GMGP05], local integral volume descriptors computed on surface points deliver a coarse alignment of two shapes. This alignment is used as an initialization for ICP.

A variational approach to surface matching of objects given in a parametric representation is presented in [LDRS05]. It provides remarkable results

by applying well-established matching approaches from image processing to the domain of surfaces. However, a coarse alignment of the parameter domains is a prerequisite for good results.

Partial matching of two shapes is addressed in [MGGP06] where shape signatures that are invariant under rigid transformations are computed from surface patches. They are used to estimate partial similarity between two shapes. Although providing good results in several applications like automatic scan alignment this method cannot deal with differently scaled objects. In [GCO06], local surface descriptors for triangle meshes are introduced. A voting scheme is used to determine self-similarities, alignments and partial matchings.

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. Other techniques are based on matching graphs, e.g. medial axis transform or 3D shock scaffold [LK01] or Reeb graphs [HSKK01] extracted from the shapes. While they 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 different regions of objects can be assigned variable importance using weights for query object surfaces. However, this method relies on correct global pre-alignments of the objects. 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 in scanned cluttered scenes. The idea was to limit the support of the local shape contexts making the approach suitable for part detection. However, a major limitation of this method is that it does 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 Salient Point Detection

The detection of salient points in a shape is a crucial ingredient of our approach. We first describe how the 2D method introduced by Lindeberg [Lin98] generalizes to three dimensions in a straightforward way. We then show how the salient points can efficiently be computed using the algorithm described in [Low04].

## 3.1 Robust 3D Salient Points

According to [Lin98], 2D salient points 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 $f : \mathbb{R}^2 \to \mathbb{R}$:

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

where $G$ denotes the two dimensional Gaussian $G(x,\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{\|x\|^2}{2\sigma^2})$, where $x$ denotes the pixel in the intensity image and $*$ denotes the convolution operator.

Lindeberg 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 $\sigma$, 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 $\sigma_i$. Lindeberg showed in [Lin93] that this scale is equal to the radius of the detected Gaussian blob.

The blob features generalize to three dimensional images $f : \mathbb{R}^3 \to \mathbb{R}$ straightforwardly. For a given closed shape $S$, a distance field of a fixed size is computed. This distance field is converted into a voxel grid. All voxels further apart from the shape than the diameter of one voxel are set to 0. All voxels with a negative distance larger than the diameter of one voxel, i.e. those voxels lying completely inside $S$ are set to 1. The remaining voxels are set to values between 0 an 1, depending on how close to the shape they are. By that, the salient feature points can simply be computed as the extrema of $L_{norm}$ on this voxel grid. Note that this smooth voxelization is crucial for the detection of salient points. Assigning only binary values to the voxels leads to artifacts resulting in a lot of small scale salient points on the shape surface.

Defined as the convolution of an image with a flipped Mexican hat shaped function (see Figure 1),

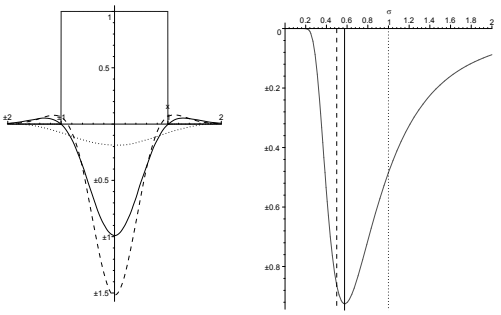


Figure 1: *Left: Characteristic function of 3D unit ball and normalized LoG along X-axis. Right: Response of the 3D normalized LoG filter in the center of the 3D unit ball*

$L_{norm}$ attains a high value at the center of a circular outside region 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. To obtain a small set of highly salient features we therefore first select only local minima of $L_{norm}$. If a point attains more than only one minimum in the scale space, we assign it the scale where the global minimum is achieved. Note however, that although minima correspond to inside blobs they might actually lie outside the shape.

The scale normalization of $L_{norm}$ was chosen such that it attains a unique local minimum at scale $\sigma_0$ for a 2D-Gaussian with $\sigma = \sigma_0$. However, a 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 $\sigma = \sqrt{2/3}\sigma_0$. 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 the 3D ball $B_r$ of radius $r$. Analogously to [Lin93], we observe that the spatial minimum of $L_{norm}$ is attained in the center of the ball and evaluates to

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

Differentiating and solving $\frac{\partial}{\partial\sigma} L_{norm} = 0$, we find that the minimum is attained for $\sigma = \frac{r}{\sqrt{3}}$ (see also Figure 1). We therefore apply an additional scale adjustment assigning each salient point a radius of $\mathbf{r}_i = \sqrt{3}\sigma_i$ where $\sigma_i$ denotes the scale at which the actual minimum of $L_{norm}$ is found.

### 3.2 Computation

To efficiently compute blob features in 2D intensity images Lowe [Low04] proposed an algorithm 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

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

where $k$ denotes a fixed scale step factor. Therefore, the scale normalized LoG 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, such that the resulting filtered images are separated by a constant factor $k$ in scale. Each octave of scale space (i.e. doubling of $\sigma$) is subdivided into $s$ intervals, such that $k = 2^{1/s}$. Additionally, a prior smoothing step is applied to the first level of each octave. From the resulting 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 algorithm can be generalized to three dimensional straightforwardly. However, as already discussed only minima are selected by comparing each difference value to its neighbors.

We evaluated the stability of the detected salient points on a set of 32 objects taken from the Princeton Shape Benchmark (PSB) [SMKF04]. It is a database containing 1814 3D objects clustered into several classes. The objects were chosen from a diverse range, including animals, humans, tools, cars and buildings. The focus of our experiments was to examine the influence of the voxelization on the repeatability of the salient points. We therefore com-

| $\sigma_p$ | 1.0 | 1.5 | 2.0 |
|---|---|---|---|
| Repeatability % | 79.72 | 86.65 | 87.73 |
| #Salient points | 442 | 313 | 266 |

Table 1: *Salient point detection: Evaluation results*

puted 16 rotations around an arbitrary axis for each of the 32 objects. The smooth voxelization was conducted using a grid of size $128 \times 128 \times 128$. The salient points were first computed on the original model. According to the rotations applied, we could precisely predict where the salient points in the rotated versions should be located. Like in the evaluation for 2D salient points in [Low04], we changed the kernel width $\sigma_p$ for the prior smoothing and the number $s$ of scales sampled per octave. Two salient points in the original and in the rotated object with scales $\sigma_o$ and $\sigma_r$ are considered to match if they are detected within a radius of $\sigma_o$ and if $\sigma_r$ is within a factor of $\sqrt{3}$ of $\sigma_o$.

Table 1 shows the repeatability for different prior smoothings using 4 intervals per octave. The lower row contains the total number of salient points found in the 32 original objects. The repeatability is comparable to that in [Low04] for 2D images.

## 4 Local Shape Descriptors

The purpose of a local shape descriptor is to characterize the surrounding geometry of the salient point it is associated to. This is done by providing a compact representation of the local geometry within a certain radius chosen with respect to the scale the associated salient point was detected on. In this section we describe how to compute SH based shape descriptors that provide invariance under rotation, translation and scaling.

### 4.1 Local SH Descriptors

The basis for the construction of our local shape descriptors are the global SH descriptors introduced in [FMK*03]. While in this work one global SH descriptor incorporating the complete object geometry is constructed, we instead compute one local SH descriptor for each detected salient point. Thereby, the scale $\sigma_i$ the salient point was detected on determines that part of the geometry around the point that is captured by the descriptor.

For the $i$th salient point we know its position $x_i \in \mathbb{R}^3$ and the scale of detection $\sigma_i$. In a nutshell, $x_i$ is used to obtain translation invariance and $\sigma_i$ is used to scale the shell radii which results in scale invariance. As described in Section 3, the scale of salient point detection corresponds to the size of the detected structure (the radius of the fitted spherical blob), and therefore, there is very little discriminating information within a radius of $\sigma_i$ around $x_i$. Hence, we construct $n$ concentric shells of equal radius intervals that are centered at $x_i$ with radius range between $\sigma_i$ and $\rho \cdot \sigma_i$, with $\rho \geq 1$. Subsequently, the shells are discretized into equiangular bins along the solid angle $\omega$. For the $k$-th shell, the spherical function $f_k(\omega)$ describes the number of object voxels per bin normalized by the bin size. The situation is sketched as a 2D example in Fig. 2. Figures 4(a)-4(e) show examples of salient points computed on different shapes. The scales $\sigma_i$ these points were detected on are represented by the colored spheres within the shapes. The larger transparent spheres represent the radius $\rho \cdot \sigma_i$ up to which the geometry around the salient point is captured by the descriptor.

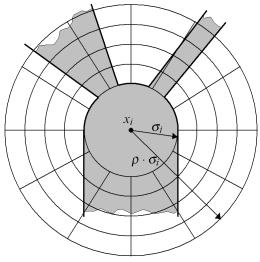


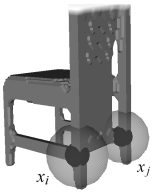

Figure 2: *2D sketch of the support region for local descriptor computation. $x_i$ is the position of the salient point. The shells take radii between $\sigma_i$ and $\rho \cdot \sigma_i$.*

The spherical function $f_k(\omega)$ can be decomposed into the SH-basis:

$$f_k(\omega) = \sum_{l=0}^{L} \sum_{|m|=\leq l} \hat{f}_m^{kl} Y_m^l(\omega), \tag{2}$$

where $L$ denotes the maximum bandwidth. To consider the behavior of SH under rotation, let $Y_l(\omega) = (Y_l^l(\omega), Y_l^{l-1}(\omega), \ldots, Y_l^{-l}(\omega))^t$ and let $R(\Theta) = R_z(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma)$ be a rotation parametrized by the zyz-Euler angles $\Theta = (\alpha, \beta, \gamma)$. Then the rotated SH can then be written as

$$Y_l(R(\Theta)\omega) = O_l(R(\Theta)) Y_l(\omega), \tag{3}$$




(a) Part of object m821 from PSB.

| Method | DPD | CMRD |
|---|---|---|
| Corr($d_{xi}$,$d_{xi}$) | 1.0 | 1.0 |
| Corr($d_{xj}$,$d_{xj}$) | 1.0 | 1.0 |
| Corr($d_{xi}$,$d_{xj}$) | 1.0 | 0.815 |

(b) Correlation of the associated descriptors.

Figure 3: *Example for object features that are axis-symmetric but not invariant under rotation. Figure 3(a) shows two of the detected salient points denoted by $x_i$ and $x_j$. The inner and outer sphere limit the volume for which $f(\omega)$ is built. Table 3(b) shows that discarding the phase information leads to a descriptor that cannot distinguish these features (DPD) while our descriptor including the correlation maximizing rotation (CMRD) is able to do so.*

where $O_l(R(\Theta))$ is a $L \times L$ matrix consisting of the Wigner-D functions $D^l_{ij}(R(\Theta))$ at the $ij$-th site. As $O_l(R(\Theta))$ is unitary, it does not change norms of vectors. By that, the norms of the projected $f_k(\omega)$

$$\begin{aligned} \mu^k_l &= (\sum_{m=-l}^{l} \langle f_k(\omega), Y^m_l(\omega)\rangle^2)^{1/2} \qquad (4) \\ &= (\sum_{m=-l}^{l} \langle f_k(R(\Theta)(\omega)), Y^m_l(\omega)\rangle^2)^{1/2} \end{aligned}$$

lead to the rotation invariant global shape descriptor in [FMK*03].

However, discarding the phase information implies a loss of information. In fact, the resulting descriptor is by its construction not able to distinguish axis-symmetric features that are variant under rotation (see Figure 3). We therefore adopt a different strategy: For each salient point $x_i$ we store a vector $d_{xi}$ of length $n \times (L^2 + 2L + 1)$, containing the SH coefficients $\hat{f}^{lk}_m$. Before comparing two descriptors $d_{xi}$ and $d_{xj}$, we estimate the underlying $R(\Theta)$ that is needed to orient the descriptors consistently.

The correlation of two discrete vector-valued functions $f$ and $g$ on the unit sphere $S^2$ which are rotated according to $R$ is defined as follows:

$$Corr_{f,g}(R(\Theta)) = \sum_{\omega \in S^2} \sum_{k=0}^{n-1} f_k(\omega)\overline{g_k(R(\Theta)\omega)}. \qquad (5)$$

Our goal is to find the rotation $R(\hat{\Theta})$ maximizing this correlation. Simply calculating the correlation for a dense set of points in the space of the rotation group $SO(3)$ and returning the point with the maximum correlation will be horribly slow. E.g., if we assume a desired accuracy of one degree in all three dimensions we actually have to calculate the correlation for $360 \times 180 \times 360$ rotations. Using the SH representation of the functions, we can adopt a more sophisticated strategy.

According to Equation 3, we can express $g_k(R(\Theta)))$ as follows:

$$g_k(R(\Theta)) = \sum_{l=0}^{L} \sum_{|q|\le l} \left( \sum_{|m|\le l} \hat{g}^l_m D^l_{qm}(R(\Theta)) \right) Y^l_q(\omega)$$

which by exploiting the orthogonality of the SH basis immediately leads to

$$Corr_{g,f}(R(\Theta)) = \sum_{l=0}^{L} \sum_{|q|\le l} \sum_{|m|\le l} \sum_{k=0}^{n-1} \hat{f}^{lk}_q \overline{\hat{g}^{lk}_m D^l_{qm}(\Theta)}. \qquad (6)$$

The direct evaluation of Equation 6 has complexity $O(L^3)$. Using the inverse FFT in [KR03] which evaluates the equation for a grid of $(2L+1)^3$ rotations, complexity reduces to $O(L^3 \log^2 L)$ compared to $O(L^6)$ for the direct evaluation.

The resulting maximal correlation is divided by the length of the vectors $\hat{f}$ and $\hat{g}$ containing the SH coefficients, which leads to a normalized correlation taking values in $[-1,1]$. In the following section, we will use this normalized correlation as a similarity measure between two salient points.

## 5 Matching Energy

In this section we will show how the computation of correspondences between salient points can be written as an energy minimization problem.

First we will show why choosing correspondences according to the distance of the associated SH descriptors is not sufficient. We therefore refer to Figure 4(d), showing an example of two similar shapes on which salient points and descriptors have been computed. As we are dealing with self-similarities caused by symmetries in this example, the emphasized descriptor in the left object has almost the same distance to every of the four descriptors in the right object. As we see, the use of only descriptor based distances would lead to non-unique and therefore non-binary correspondences.

To alleviate this problem, we enforce a constraint based on the spatial relationship of both salient point sets. As we want to determine correspondences between similar objects, it is most likely

that there exists a parametric transformation that is able to map one object onto the other, given correspondences between the points of both sets. We will therefore use an optimization scheme that iteratively estimates the correspondences and the supposed underlying transformation.

Let $\mathscr{X} = \{x_1, \cdots, x_n\}$ and $\mathscr{Y} = \{y_1, \cdots, y_m\}$ be two sets of salient points. The correspondences between salient points $x_i$ and $y_j$ we want to determine are denoted by $c_{ij}$, aggregated in a $n \times m$ correspondence matrix $C$. As we are dealing with binary correspondences, $c_{ij} \in \{0,1\}$, where $c_{ij} = 1$ means that both points are corresponding. Using the maximum correlation of descriptors $d_{xi}$ and $d_{yj}$, we define descriptor-based weights $w_{ij} \in [0,1]$ computed according to $w_{ij} = 1 - (1 + Corr(d_{xi}, d_{yj}))/2$.

We now consider the matching energy functional $E$ describing the energy required to map $\mathscr{X}$ onto $\mathscr{Y}$, given the correspondences $c_{ij}$, the descriptor based weights $w_{ij}$ and a transformation $\Phi$:

$$E(C,\Phi) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\Big(w_{ij} \cdot ||y_j - \Phi(x_i)||^2\Big). \quad (7)$$

Our goal is to minimize $E$ with respect to $c_{ij}$ and $\Phi$.

Depending on the type of transformation used it might be necessary to extend the above formulation of the matching energy by additional constraints. For more details we refer to [WNK06].

## 6 Optimization by deterministic annealing

For cost functions with rather complicated structure like Equation 7, a suitable optimization procedure has to address the problem of local minima. Deterministic annealing (DA) provides this ability by enforcing a certain amount of randomness to the solution with respect to the target random variables. This randomness is determined by the *temperature parameter* $T$ which is decreased after each optimization iteration according to the *annealing rate* $\delta$. This scheme results in an algorithm consisting of an inner and an outer loop. Thereby, the inner loop is very similar to the optimization scheme induced by the EM algorithm [DLR77]. Enforcing randomness to the solution, the correspondences $c_{ij}$ now take values in $[0,1]$. When $T$ reaches 0 at the end of the optimization, $c_{ij}$ becomes either 0 or 1. Algorithm 1 shows a brief sketch of the resulting algorithm applied to the problem of estimating optimal correspondences.

**Algorithm 1** DA for correspondence estimation

```
Initialize T = T_init, Φ = Φ_init, c_ij = 1/n.
while T > 0 do
  while Matching Energy E changes do
    E-Step: Compute optimal correspondences
            c_ij with respect to a given Φ
    M-Step: Compute energy minimizing Φ
  end while
  Update T according to T_new = δ · T_old
end while
```

For a more detailed introduction to deterministic annealing we refer to [Ros98]. Technical aspects of the application to the discussed correspondence estimation problem can be found in [WNK06].

## 7 Experiments and Results

To evaluate our correspondence estimation framework, we used a selection of similar objects which can be found in the PSB. In our experimental setting, we first determined a smooth voxelization for each shape. The grid was of size $128 \times 128 \times 128$.

The salient points were computed according to the approach described in section 3. We used 4 intervals per octave in the image pyramid. A kernel width $\sigma_p = 2.0$ was found to provide the most stable salient points (see Table 1). However, we encountered that with $\sigma_p = 2.0$ important object details are no longer detected by a salient point as they disappear during the smoothing step. We therefore used $\sigma_p = 1.5$ yielding a repeatability close to the optimum while still preserving important object details.

For each detected salient point, a local SH descriptor was computed according to the method described in Section 4. We experimentally found the following parameter setting to work well: $\rho$ is set to 2.5, we use a SH basis with maximum bandwidth $L = 7$ and 8 concentric shells resulting in 512 dimensional local shape descriptors.

The transformation $\Phi$ in the DA scheme was chosen to be a TPS. It is a mapping consisting of a rigid and a non-rigid transformation. For details concerning the optimization process, we refer to [WNK06].

Figures 4(a)-4(f) show results of our experiments. In 4(a)-4(e) the size of the inner spheres represents the radius $\sigma_i$ of the salient points. The outer

spheres have radius $2.5 \cdot \sigma_i$. The spheres limit the volume that is captured by the local descriptor. In Figure 4(f), only the positions of the salient points are shown for the sake of clearness. The points determined to correspond are connected by lines.

## 8 Conclusion

In this work we presented a method for detection of salient points on 3D shapes. The method works for arbitrary objects. The repeatability of the salient points is comparable to the 2D results for images presented in [Low04]. SHs are used to compute scale invariant descriptors around the salient points, delivering a compact representation of the surrounding geometry. Invariance under rotation is achieved by orienting two descriptors in a consistent way. The resulting correlation between the descriptors delivers a similarity measure. Correspondences between the salient points are established by an iterative optimization scheme incorporating the position of the salient points and the information provided by the associated SH descriptor. The algorithm does not need any prepositioning of the objects. It is kept general such that it allows for the use of other descriptors than our SH based ones.

Future work should include an implementation of an effective outlier management. By now, all detected salient points in the set to which $\Phi$ is applied are taken into account. Certain details which are not contained in the other point set may therefore be matched in an undesired way (see Figure 4(f)).

More sophisticated methods to measure the similarity of two descriptors could also be examined. As the SH coefficients correspond to certain frequencies, it could be useful to assign the coefficients weights before computing the optimal correlation.

## Acknowledgments

This work was partially funded by the German Research Foundation (DFG) under grant GZ 554975(1) Oldenburg BIB 48 OLof 01-02 *Probado*.

## References

[BM92] BESL P., MCKEY N.: A method for registration of 3-d shapes. *TPAMI 14*, 2 (1992), 239–256.

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

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

[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. *CVIU 89*, 2-3 (2003), 114–141.

[DLR77] DEMPSTER A. P., LAIRD N. M., RUBIN D. B.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society B. 39* (1977), 1–39.

[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 (2001), 87–96.

[GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Tr. Gr. 25*, 1 (2006), 130–150.

[GMGP05] GELFAND N., MITRA N. J., GUIBAS L. J., POTTMANN H.: Robust global registration. In *SGP* (2005), pp. 197–206.

[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).

[KR03] KOSTELEC P., ROCKMORE D.: *FFTs on the rotation group*. Tech. Rep. 03-11-060, Santa Fe Institute's Working Paper Series, 2003.

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

[LDRS05] LITKE N., DROSKE M., RUMPF M., SCHRÖDER P.: An image processing approach to surface matching. In *Symposium on Geometric Processing* (2005), pp. 207–216.

[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 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.

[MGGP06] MITRA N. J., GUIBAS L., GIESEN J., PAULY M.: Probabilistic fingerprints for shapes. In *SGP* (2006), pp. 121–130.

[MS03] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. In *CVPR* (2003), vol. 2, pp. 257–263.

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. *IEEE TPAMI 27*, 10 (2005), 1615–1630.

[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. *International Journal of Computer Vision 65*, 1/2 (2005), 43–72.

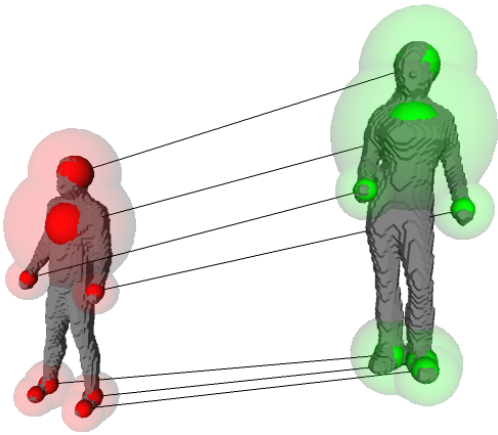
[Ros98] ROSE K.: Deterministic annealing for clustering, compression, classification, regression and related optimization problems. *Proceedings of the IEEE 86*, 11 (1998), 2210–2239.

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

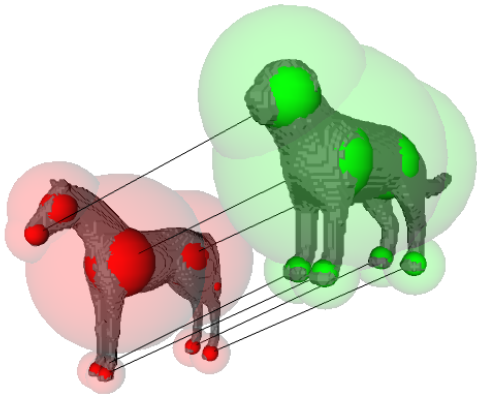
[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. In *Shape Modeling International* (June 2004).

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

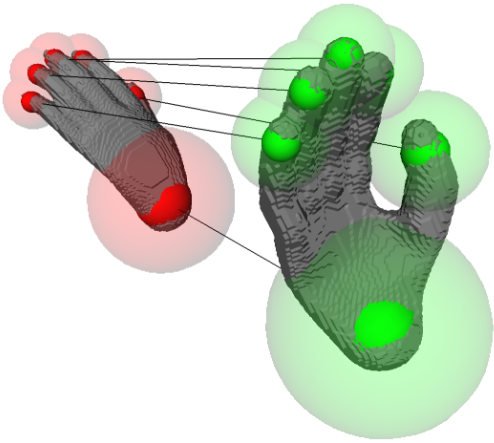
[WNK06] WESSEL R., NOVOTNI M., KLEIN R.: *Correspondences between Salient Points on 3D Shapes*. Tech. Rep. CG-2006-3, Universität Bonn, June 2006.



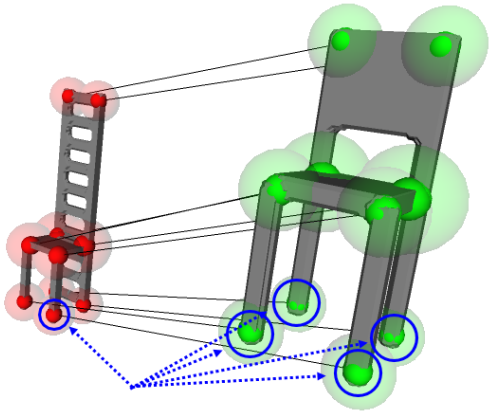
(a) A man and a woman, objects `m125` and `m179`.



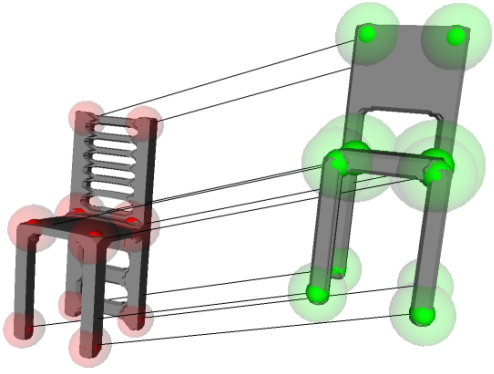
(b) A dog and a horse, objects `m88` and `m107`.



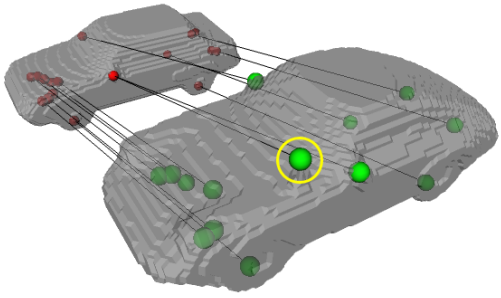
(c) A left hand and a right hand, objects `m325` and `m326`.



(d) Two Chairs, objects `m809` and `m807`.



(e) Objects `m809` and `m807`. The TPS transformation estimated in the optimization process was applied to `m809`.



(f) A sports car and a sedan, objects `m1524` and `m1552`.

Figure 4: *Correspondence results for several examples taken from the PSB. In 4(d), self-similarities cause similar distances between the descriptor in the left object and the four emphasized descriptors in the right object. In 4(f), the antenna highlighted in the sports car has no equivalent in the sedan and was therefore matched in an undesired way.*