Entropy-Based Particle Systems for Shape Correspondence

Joshua Cates, Miriah Meyer, P. Thomas Fletcher and Ross Whitaker

Scientific Computing and Imaging Institute
University of Utah
Salt Lake City, Utah

Abstract. This paper presents a new method for constructing statistical representations of ensembles of similar shapes. The proposed method relies on an optimal distribution of a large set of surface point correspondences, rather than the manipulation of a specific surface parameterization. The optimal configuration is defined as one in which the entropy or information content of each shape is balanced against the entropy of the ensemble of shapes. The correspondences are modeled as sets of dynamic particles that are manipulated using a gradient descent on the entropies of the shapes and the ensemble, but constrained to lie on a set of implicit surfaces. The proposed, particle-based method for finding correspondences requires very little preprocessing of data or parameter tuning, and therefore makes the problem of shape analysis more practical for a wider range of problems. This paper presents the formulation and several synthetic and real shape examples in two and three dimensions.

1 Introduction

Computing statistics on sets of shapes requires quantification of shape differences, which is a fundamentally difficult problem. A widely-used strategy for computing shape differences is to compare the positions of corresponding points among sets of shapes. Medical or biological shapes, however, are typically derived from the interfaces between organs or tissue types. Such surfaces are usually defined implicitly in the form of *segmented volumes*, rather than from explicit parameterizations or surface point samples. Thus, no defined, a priori relationship between points across surfaces exists. Correspondences must therefore be inferred from the shapes themselves, giving rise to the difficult, yet very important, *correspondence problem*.

Until recently, correspondences for shape statistics were established manually by choosing small sets of anatomically significant landmarks on organs or regions of interest, which would then serve as the basis for shape analysis. The demand for more detailed analyses on ever larger populations of subjects has rendered this approach unsatisfactory. Recently, Davies et al. [1] present methods for automatically establishing relatively dense sets of correspondences based on the information content of the set of points needed to describe an ensemble of similar shapes. These methods, however, rely on mappings between

fixed parameterizations, and because most shapes in medicine or biology are not derived parametrically, this reliance on a parameterization presents some significant drawbacks. Automatic selection of correspondences for nonparametric, point-based shape models has been explored in the context of surface registration [2], but such methods are not sufficient for shape analysis because they are typically concerned only with pairwise correspondences, and not correspondences across populations of points. Furthermore, these methods also assume a fixed set of samples on surfaces, whereas, in the context of this paper, we are given implicit surfaces (volume segmentations) and dynamically resample them as part of the correspondence selection process.

This paper presents a new method for extracting dense sets of correspondences that describe ensembles of similar shapes. The method is nonparametric and borrows technology from the computer graphics literature for representing surfaces as discrete point sets. The proposed method iteratively modifies a system of dynamic particles so that they follow trajectories across the surfaces to find positions that optimize the information content of the system. This strategy is motivated by a recognition of the inherent tradeoff between geometric accuracy and statistical simplicity. Our assertion is that each unit of complexity, or information, across the ensemble should be balanced against a unit of information on the surface. This approach provides a natural equivalence of information content and reduces the dependency on ad-hoc regularization strategies and free parameters. Since the points are not tied to a specific parameterization, the method operates directly on volumetric data, extends easily to higher dimensions or arbitrary shapes, and provides a more homogeneous geometric sampling as well as more compact statistical representations. The method draws a clear distinction between the objective function and the minimization process, and thus can more readily incorporate additional information such as adaptive surface sampling and high-order geometric information.

2 Related Work

The strategy of finding of parameterizations that minimize information content across an ensemble was first proposed by Kotcheff and Taylor [3]. They represent each two-dimensional contour as a set of N samples taken at equal intervals from a parameterization. Each shape is treated as a point sample in a $2N$-dimensional space, with associated covariance Σ and cost function, $\sum_k \log(\lambda_k + \alpha)$, where λ_k are the eigenvalues of Σ, and α is a regularization parameter that prevents the very *thinnest* modes (smallest eigenvalues) from dominating the process. This is the same as minimizing $\log|\Sigma + \alpha I|$, where I is the identity matrix, and $|\cdot|$ denotes the matrix determinant.

Davies et al. [1] propose *minimum description length* (MDL) as a cost function. In that work they use arguments regarding quantization to limit the effects of thin modes and to determine the optimal number of components that should influence the process. They propose a piecewise linear reparameterization and a hierarchical minimization scheme. Monotonicity in the reparameterizations en-

sures that those composite mappings are diffeomorphic. Davies et al. [4] propose a 3D extension to the MDL method, which relies on spherical parameterizations and subdivisions of an octahedral base shape, and smoothed updates, represented as Cauchy kernels. The parameterization must be obtained through another process such as [5], which relaxes a spherical parameterization onto an input mesh. The overall procedure is empirically satisfactory, but requires significant data preprocessing, including a sequence of optimizations—first to establish the parameterization and then on the correspondences—each of which entails a set of free parameters or inputs in addition to the segmented volumes. A significant concern with the basic MDL formulation is that the optimal solution is often one in which the correspondences all collapse to points where all the shapes in the ensemble happen to be near (e.g., crossings of many shapes). Several solutions have been proposed [4, 6], but they entail free parameters and assumptions about the quality of the initial parameterizations.

The MDL formulation is mathematically related to the min-log $|\Sigma + \alpha I|$ approach, as noted by Thodberg [6]. Styner et al. [7] describe an empirical study that shows ensemble-based statistics improve correspondences relative to pure geometric regularization, and that MDL performance is virtually the same as that of min-log $|\Sigma + \alpha I|$. This last observation is consistent with the well-known result from information theory that MDL is, in general, equivalent to minimum entropy [8].

Another body of relevant work is the recent trend in computer graphics towards representing surfaces as scattered collections of points. The advantage of so-called *point-set surfaces* is that they do not require a specific parameterization and do not impose topological limations; surfaces can be locally reconstructed or subdivided as needed [9]. A related technology in the graphics literature is the work on particle systems, which can be used to manipulate or sample [10] implicit surfaces. A particle system manipulates large sets of particles constrained to a surface using a gradient descent on radial energies that typically fall off with distance. The proposed method uses a set of interacting particle systems, one for each shape in the ensemble, to produce optimal sets of surface correspondences.

3 Methods

3.1 Entropy-Based Surface Sampling

We treat a surface as a subset of $\Re^d$, where $d = 2$ or $d = 3$ depending whether we are processing curves in the plane or surfaces in a volume, respectively. The method we describe here is limited to smooth, closed manifolds of codimension one, and we will refer to such manifolds as *surfaces*. We sample a surface $\mathcal{S} \subset \Re^d$ using a discrete set of N points that are considered random variables $Z = (X_1, X_2, \ldots, X_N)$ drawn from a probability density function (PDF), $p(X)$. We denote a realization of this PDF with lower case, and thus we have $z = (x_1, x_2, \ldots, x_N)$, where $z \in \mathcal{S}^N$. The probability of a realization x is $p(X = x)$, which we denote simply as $p(x)$.

The amount of information contained in such as random sampling is, in the limit, the differential entropy of the PDF, which is $H[X] = -\int_S p(x) \log p(x) dx = -E\{\log p(X)\}$, where $E\{\cdot\}$ is the expectation. When we have a sufficient number of samples from p, we can approximate the expectation by the sample mean [8], which gives $H[X] \approx -(1/N) \sum_i \log p(x_i)$. We must also estimate $p(x_i)$. Density functions on surfaces can be quite complex, and so we use a nonparametric, Parzen windowing estimation of this density using the samples themselves. Thus we have

$$p(x_i) \approx \frac{1}{N(N-1)} \sum_{\substack{j=1 \\ j \neq i}}^{N} G(x_i - x_j, \sigma) \tag{1}$$

where $G(x_i - x_j, \sigma)$ is a d-dimensional, isotropic Gaussian with standard deviation σ. The cost function C, is therefore an approximation of (negative) entropy: $H[X] \approx -C(x_1, \ldots, x_N) = \sum_i \log \frac{1}{N(N-1)} \sum_{j \neq i} G(x_i - x_j, \sigma)$,

In this paper, we use a gradient-descent optimization strategy to manipulate particle positions. The optimization problem is given by:

$$\hat{z} = \arg\min_z E(z) \text{ s.t. } x_1, \ldots, x_N \in \mathcal{S}. \tag{2}$$

The negated gradient of E is

$$-\frac{\partial E}{\partial x_i} = \frac{1}{\sigma 2} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j) G(x_i - x_j, \sigma)}{\sum_{\substack{j=1 \\ j \neq i}}^{N} G(x_i - x_j, \sigma)} = \sigma^{-2} \sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - x_j) w_{ij}, \tag{3}$$

where $\sum_j w_{ij} = 1$. Thus to minimize C, the samples (which we will call *particles*) must move away from each other, and we have a set of particles moving under a repulsive force and constrained to lie on the surface. The motion of each particle is away from all of the other particles, but the forces are weighted by a Gaussian function of inter-particle distance. Interactions are therefore local for sufficiently small σ. We use a Jacobi update with forward differences, and thus each particle moves with a *time* parameter and positional update $x_i \leftarrow x_i - \gamma \frac{\partial C}{\partial x_i}$, where γ is a time step and $\gamma < \sigma 2$ for stability.

The surface constraint is specified by the zero set of a scalar function $F(x)$. This constraint is maintained, as described in several papers [10], by first projecting the gradient of the cost function onto the tangent plane of the surface (as prescribed by the method of Lagrange multipliers), and then by iterative reprojection of the particle onto the nearest root of F by the method of Newton-Raphson. Another aspect of this particle formulation is that it computes Euclidean distance between particles (in the ambient space), rather than the geodesic distance on the surface. Thus, we assume sufficiently dense samples so that nearby particles lie in the tangent planes of the zero sets of F. This is an important consideration; in cases where this assumption is not valid, such as highly convoluted surfaces, the distribution of particles may be affected by neighbors that are outside of the true manifold neighborhood. The question of particle interactions with more general distance measures remains for future work.

Finally, we must choose a σ for each particle, which we do automatically, before the positional update, using the same optimality criterion described above. The contribution to C of the ith particle is simply the probability of that particle position, and optimizing that quantity with respect to σ gives a maximum likelihood estimate of σ for the current particle configuration. We use Newton-Raphson to find σ such that $\partial p(x_i, \sigma)/\partial\sigma = 0$, which typically converges to machine precision in several iterations.

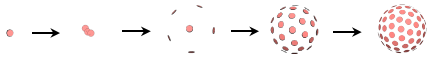

Fig. 1. A system of 100 particles (right) produced by successive splitting of a single particle (left).

There are a few important numerical considerations. We must truncate the Gaussian kernels, and so we use $G(x, \sigma) = 0$ for $|x| > 3\sigma$. This means that each particle has a finite radius of influence, and we can use a spatial binning structure to reduce the computational burden associated with particle interactions. If σ for a particle is too small, a particle will not interact with its neighbors at all, and we cannot compute updates of σ or position. In this latter case, we update the kernel size using $\sigma \leftarrow 2\sigma$, until σ is large enough for the particle to interact with its neighbors. Another numerical consideration is that the system must include bounds $\sigma_{\min}$ and $\sigma_{\max}$ to account for anomalies such as bad initial conditions, too few particles, etc. These are not critical parameters, so as long as they are set to include the minimum and maximum resolutions the system operates reliably.

The mechanism described in this section is, therefore, a self contained, self tuning system of particles that distribute themselves using repulsive forces to achieve optimal distributions. For this paper we initialize the system with a single particle that finds the nearest zero of F, which then splits (producing a new, nearby particle) at regular intervals until a specific number of particles are produced and they reach a steady state. Figure 1 shows this process for a sphere.

3.2 The Entropy of The Ensemble

An ensemble $\mathcal{E}$ is a collection of M surfaces, each with their own set of particles, i.e. $\mathcal{E} = z1, \ldots, z^M$. The ordering of the particles on each shape implies a correspondence among shapes, and thus we have a matrix of particle positions $P = x_j^k$, with point samples along the rows and shapes across the columns. We model $z^k \in \Re^{Nd}$ as an instance of a random variable Z, and propose to minimize the combined ensemble and shape cost function

$$Q = H(Z) - \sum_k H(P^k), \tag{4}$$

which favors a compact ensemble representation, balanced against a uniform distribution of particles on each surface. The different entropies are commensurate so there is no need for ad-hoc weighting of the two function terms.

For this discussion we assume that the complexity of each shape is greater than the number of examples, and so we would normally choose $N > M$. Given

the low number of examples relative to the dimensionality of the space, we must impose some conditions in order to perform the density estimation. For this work we assume a normal distribution and model $p(Z)$ parametrically using a Gaussian with covariance Σ. The entropy is then given by

$$H(Z) \approx \frac{1}{2} \log |\Sigma| = \frac{1}{2} \sum_{j=1}^{Nd} \log \lambda_j, \tag{5}$$

where $\lambda_1, ..., \lambda_{Nd}$ are the eigenvalues of Σ.

In practice, Σ will not have full rank, in which case the entropy is not finite. We must therefore regularize the problem with the addition of a diagonal matrix αI to introduce a lower bound on the eigenvalues. We estimate the covariance from the data, letting Y denote the matrix of points minus the sample mean for the ensemble, which gives $\Sigma = (1/(M-1))YY^T$. Because $N > M$, we perform the computations on the dual space (dimension M), knowing that the determinant is the same up to a constant factor of α. Thus, we have the cost function G associated with the ensemble entropy:

$$\log |\Sigma| \approx G(P) = \log \left| \frac{1}{M-1} Y^T Y, \right| \text{ and } -\frac{\partial G}{\partial P} = Y(Y^T Y + \alpha I)^{-1}. \tag{6}$$

We now see that α is a regularization on the inverse of $Y^T Y$ to account for the possibility of a diminishing determinant. The negative gradient $-\partial G/\partial P$ gives a vector of updates for the entire system, which is recomputed once per system update. This term is added to the shape-based updates described in the previous section to give the update of each particle:

$$x_j^k \leftarrow \gamma \left[-\partial G/\partial x_j^k + \partial E^k/\partial x_j^k \right]. \tag{7}$$

The stability of this update places an additional restriction on the time steps, requiring γ to be less than the reciprocal of the maximum eigenvalue of $(Y^T Y + \alpha I)^{-1}$, which is bounded by α. Thus, we have $\gamma < \alpha$, and note that α has the practical effect of preventing the system from slowing too much as it tries to reduce the thinnest dimensions of the ensemble distribution. This also suggests an annealing approach for computational efficiency (which we have used in this paper) in which α starts off somewhat large (e.g., the size of the shapes) and is incrementally reduced as the system iterates.

The choice of a Gaussian model for $p(Z = z)$ is not critical for the proposed method. The framework easily incorporates either nonparametric, or alternate parametric models. In this case, the Gaussian model allows us to make direct comparisons with the MDL method, which contains the same assumptions. Research into alternative models for Z is outside the scope of this paper and remains of interest for future work.

The method outlined above assumes a population of surfaces that are in alignment with one another. For medical image datasets, this is often not the case, and some surface registration technique must be applied as a part of the

algorithm for finding correspondences. Goodall [11], for example, suggests the point-based Procrustes method. For the results given in the following section, we assume the surface data is in alignment and leave the analysis of the stability and interplay between the proposed method and any surface registration techniques for future work. Preliminary results do suggest, however, that Procrustes alignment may be effectively applied at intervals in the proposed correspondence optimization.

4 Results and Conclusions

We begin with two experiments on closed curves in a 2D plane and a comparison with the 2D open-source Matlab MDL implementation given by Thodberg [6]. In the first experiment, we used the proposed, particle method to optimize 100 particles per shape on 24 *box-bump* shapes, similar to those described in [6]. Each shape was constructed from a set of point samples using cubic b-splines with the same rectangle of control, but with a bump added at a random location along the top of its curve. Distance transforms from these shapes were constructed using the fast-marching algorithm [12], which forms implicit contours suitable for input to the proposed algorithm. MDL correspondences were computed using 128 nodes and *mode 2* of the Matlab software, with all other parameters set to their defaults (see [6] for details). Both methods identified a single dominant mode of variation, but with different degrees of leakage into orthogonal modes. MDL lost 0.34% of the total variation from the single mode, while the proposed method lost only 0.0015%. Figure 2 illustrates the mean and three standard deviations of the first mode of the two different models. Shapes from the particle method remain more faithful to those described by the original training set, even out to three standard deviations where the MDL description breaks down. A striking observation from this experiment is how the relatively small amount of variation left in the minor modes of the MDL case produce such a significant effect on the results of shape deformations.

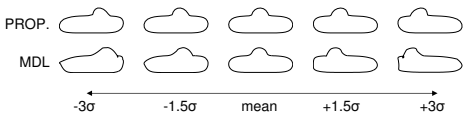

Fig. 2. A comparison of the mean and three standard deviations of the *box-bump* experiment.

The second experiment was conducted on the set of 18 hand shape contours described in [1], again applying both the particle method and MDL using the same parameters as described above. As with the box-bump data, distance transforms were generated from the spline-based contour models for input to the correspondence algorithm. In this case, we also compared results with a set of *ideal*, manually selected correspondences, which introduce anatomical knowledge of the digits. Figure 3 compares the three resulting models in the top three modes of variation to ± 3 standard deviations. A detailed analysis of the principle components showed that the proposed particle method and the manually

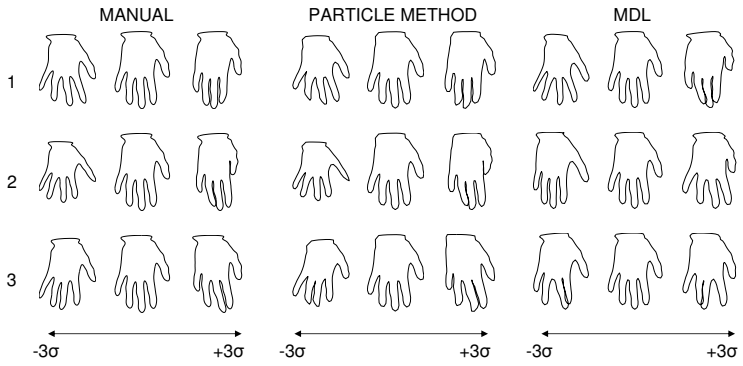

Fig. 3. The mean and three standard deviations of the top three modes of the hand models.

selected points both produce very similar models, while MDL differed significantly, particularly in first three modes.

Existing 3D MDL implementations rely on spherical parameterizations, and are therefore only capable of analyzing shapes topologically equivalent to a sphere. The particle-based method does not have this limitation. We applied the proposed method to a set of randomly chosen tori, drawn from a 2D distribution that is parameterized by the small radius r and the large radius R. Samples were chosen from a distribution with mean $r = 1, R = 2$ and $\sigma_r = 0.15, \sigma_R = 0.30$, with a rejection policy that excluded invalid tori (e.g., $r > R$). Figure 4 shows the particle system distribution across two of the torus shapes in the sample set with 250 correspondences. An analysis of the correspondences showed that the particle system discovered the two pure modes of variation, with only 0.08% leakage into smaller modes.

We applied the proposed method to a set of 20, volumetric hippocampus segmentations chosen at random from a larger data set described in Styner, et al. [13]. Using the fast-marching algorithm, this time in 3D, we generated distance transforms from the boundaries of these segmentations for input to the method. Fig. 4 shows the particle system distributed across two of the shapes after optimizing 300 particles per shape. We used the modes from the resulting Gaussian model to construct a set of surface reconstructions for the three largest principle components. These modes are illustrated in Fig. 5 to three standard deviations, with percentage of total variation of 38.78%, 26.31%, and 12.29% for the modes, respectively. The surface meshes shown were generated by the tight cocone algorithm for sur-

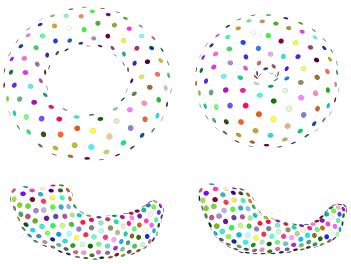

Fig. 4. Particle correspondences in two tori (left) and two hippocampus (right) shapes. Corresponding inter-shape particles have matching colors.

face reconstruction from point clouds [14], using the implementation provided by the authors of that work.

Because the proposed method is completely generalizable to higher dimensions, we were able to perform both the 2D and 3D experiments using the same *C++* software implementation, which is templated by dimension. All experiments were run on a 2Ghz processor with run times of approximately 20 minutes for 2D cases and 45 minutes the 3D cases. In each case, the numerical parameter $\sigma_{\min}$ was set to machine precision and $\sigma_{\max}$ was set to the size of the domain. For the annealing parameter α, we used a starting value roughly equal to the diameter of an average shape and reduced it to machine precision over several hundred iterations. The results presented in this section are typical of reliably similar results produced during several experimental runs, suggesting that the proposed method is fairly robust to the initialization.

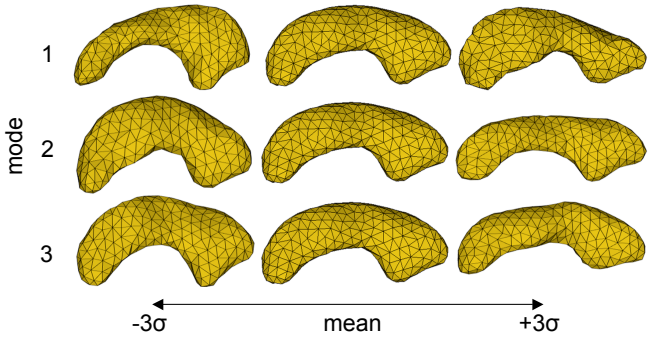

Fig. 5. The mean and three standard deviations of the top three modes of the hippocampus model.

The proposed nonparametric method for shape correspondences produces results that compare favorably with the state of the art. The method works directly on volumes, requires very little parameter tuning, and generalizes easily. The particle-based shape correspondence offers a practical solution for a wider range of shape analysis problems relative to the work in the literature. Although the energy functions push the system toward consistent configurations of particles, the approach does not guarantee diffeomorphic mappings between shapes. The system might be extended to include neighborhood configurations as a part of the model, but this remains a topic of further investigation.

Acknowledgments.

The authors wish to thank Hans Henrik Thodberg for the use of his open source MDL Matlab code.

This work was funded by the Center for Integrative Biomedical Computing, National Institutes of Health (NIH) NCRR Project 2-P41-RR12553-07.

This work is also part of the National Alliance for Medical Image Computing (NAMIC), funded by the NIH through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`.

References

1. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imaging **21**(5) (2002) 525–537
2. Audette, M., Ferrie, F., Peters, T.: An algorithmic overview of surface registration techniques for medical imaging. Medical Image Analysis **4** (2000) 201–217
3. Kotcheff, A., Taylor, C.: Automatic Construction of Eigenshape Models by Direct Optimization. Medical Image Analysis **2** (1998) 303–314
4. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: 3d statistical shape models using direct optimisation of description length. In: ECCV (3). (2002) 3–20
5. Brechbuhler, C., Gerig, G., Kubler, O.: Surface parameterization and shape description. Visualisation in Biomedical Computing (1992) 80–89
6. Thodberg, H.H.: Minimum description length shape and appearance models. In: IPMI. (2003) 51–62
7. Styner, M., Rajamani, K.T., Nolte, L.P., Zsemlye, G., Székely, G., Taylor, C.J., Davies, R.H.: Evaluation of 3d correspondence methods for model building. In: IPMI. (2003) 63–75
8. Cover, T., Thomas, J.: Elements of Information Theory. Wiley and Sons (1991)
9. Boissonnat, J.D., Oudot, S.: Provably good sampling and meshing of surfaces. Graphical Models **67** (2005) 405–451
10. Meyer, M.D., Georgel, P., Whitaker, R.T.: Robust particle systems for curvature dependent sampling of implicit surfaces. In: In Proceedings of the International Conference on Shape Modeling and Applications. (2005) 124–133
11. Goodall, C.: Procrustes methods in the statistical analysis of shape. J. R. Statistical Society B **53** (1991) 285–339
12. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press (1996)
13. Styner, M., Lieberman, J., Gerig, G.: Boundary and medial shape analysis of the hippocampus in schizophrenia. In: MICCAI. (2003) 464–471
14. Dey, T.K., Goswami, S.: Tight cocone: a water-tight surface reconstructor. In: SM '03: Proceedings of the eighth ACM symposium on Solid modeling and applications, New York, NY, USA, ACM Press (2003) 127–134