Interpretable exemplar-based shape classification using constrained sparse linear models

Gunnar A. Sigurdsson, Zhen Yang, Trac D. Tran, and Jerry L. Prince

Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Many types of diseases manifest themselves as observable changes in the shape of the affected organs. Using shape classification, we can look for signs of disease and discover relationships between diseases. We formulate the problem of shape classification in a holistic framework that utilizes a lossless scalar field representation and a non-parametric classification based on sparse recovery. This framework generalizes over certain classes of unseen shapes while using the full information of the shape, bypassing feature extraction. The output of the method is the class whose combination of exemplars most closely approximates the shape, and furthermore, the algorithm returns the most similar exemplars along with their similarity to the shape, which makes the result simple to interpret. Our results show that the method offers accurate classification between three cerebellar diseases and controls in a database of cerebellar ataxia patients. For reproducible comparison, promising results are presented on publicly available 2D datasets, including the ETH-80 dataset where the method achieves 88.4% classification accuracy.

Keywords: shape classification, interpretable classifiers, sparse recovery, morphology, signed distance functions

1. INTRODUCTION

A patient suffering from spinocerebellar ataxia will exhibit atrophy in regions of the cerebellum. The ability to understand brain disease from shape gives the possibility of, for example, diagnosing specific subtypes of ataxia and guide treatment, particularly if there is no genetic diagnosis. To look for these changes in cerebellar shape, a variety of approaches from the shape analysis literature could be applied. Most methods first describe the shapes using extracted features and then apply known classification methods to those features.[1,2] Although feature selection typically reduces the dimension of the resulting classification stage—and therefore may save computation time—it is always associated with loss of information and may encourage training errors and lack of generalization to new shapes not included in training data.

Certain scalar fields, such as signed distance functions (SDFs)[3–5] or solutions to the Poisson equation,[1] are lossless representations of shapes. For classification of such high dimensional representations, specific classification methods must be used, such as the method of Wright *et al.*[6] which is a generalization of nearest neighbor and nearest subspace, that finds a sparse linear combination of a dictionary and uses it for classification. Sparse dictionaries of shapes have proven useful to model complex shape variations.[2] While using a space of scalar fields that is closed under certain linear operations is appealing (Poisson equation and convex combinations, for example), we found that linear approximations of the space of signed distance functions gave the best performance, and furthermore, admitted useful properties.

2. SHAPE REPRESENTATIONS

We represent the continuous binary shape A as a mapping from d-dimensional space to a binary value, that is, $A\colon \mathbb{R}^d \to [0,1]$. It is convenient to treat A like a set; in particular, we say that $p \in A$ if $A(p) = 1$. The distance function associated with A is

$$D_A(p) = \inf_{x \in A} d(x,p)\,, \tag{1}$$

Further author information: (Send correspondence to G.A.S.)
G.A.S.: E-mail: gunnar@jhu.edu
J.L.P.: E-mail: prince@jhu.com

where $d(\cdot,\cdot)$ is the Euclidean distance metric. The signed distance function (SDF) is defined as[7]

$$\phi_A(p) = \begin{cases} D_A(p) & \text{if } p \in A^c \\ -D_{A^c}(p) & \text{if } p \in A \end{cases}, \tag{2}$$

where A^c denotes the set complement, i.e., $A^c = \{p \mid A(p) = 0\}$. Given ϕ_A, we have that $\partial A = \{p \mid \phi_A(p) = 0\}$ and $A = \{p \mid \phi_A(p) \leq 0\}$. In practice SDFs are commonly truncated (e.g. $\phi(p) = 30$ if $\phi(p) > 30$).

3. SHAPE CLASSIFICATION BY SPARSE RECOVERY

Our goal is to estimate the class c of an observed shape from among the classes in a set $\mathcal{C}$. For example, $\mathcal{C}$ might contain the classes *healthy* and *diseased* for the task of classifying the health of an organ by shape. We use discretized signed distance functions which are created by stacking all pixels/voxels into large vectors. We form a dictionary of aligned training shapes and find a sparse representation of a given test shape using shapes (atoms) from the dictionary. The dictionary may be learned, but care must be taken that the dictionary contains actual examplars for interpretability, as shown in Section 4. The shape is then classified by finding the object class that minimizes the residual between the test shape and approximate shape using only the atoms associated with that class. Stated mathematically, we start with a dictionary of shapes,

$$\Phi = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K], \tag{3}$$

where each $\boldsymbol{\phi}_i \in \mathbb{R}^N$ is a (vectorized) SDF of a training shape, N is the total number of pixels/voxels, and K is the number of dictionary elements. Let $\mathbf{y} \in \mathbb{R}^N$ be a test shape, and $\mathbf{x} \in \mathbb{R}^K$ be a vector indicating the contribution of each atom in our dictionary when representing the test shape. Assuming the space of elements is closed with respect to linear combinations, we want $\mathbf{y} = \Phi\mathbf{x}$, where $\mathbf{x}$ is of sparsity s ($\|\mathbf{x}\|_0 = s$). Drawing from Wright *et al.*[6] we can write this problem ideally as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \Phi\mathbf{x} = \mathbf{y} \tag{4}$$

$$\hat{c} = \arg\min_{c \in \mathcal{C}} \|\mathbf{y} - \Phi\delta_c(\hat{\mathbf{x}})\|_2, \tag{5}$$

where $\delta_c(\mathbf{x})$ is defined to be x_i when the i-th column of Φ—i.e., i-th atom—belongs to class c and is zero otherwise. Ideally, the solution $\hat{\mathbf{x}}$ will be sparse and the non-zero entries will all correspond to atoms in the same class. In any case, $\hat{\mathbf{c}}$ will be selected to be the class from which the reconstructed measurements best fit the observations. It is possible to relate boundary displacements between shapes to the squared difference of their SDFs. By looking at small perturbations dx for each point on the boundary $x \in \partial X$ for some shape X (yielding a new shape X') and the resulting changes in the *influence zone* of x, $\mathcal{Z}_x \equiv \{p \mid \forall \xi \in \partial X,\ |d(p,x)| \leq |d(p,\xi)|\}$. Assuming the operations are defined:

$$\sum_{x \in \partial X} \int_{\mathcal{Z}_x} w_x(p) \|dx\|^2 \mathrm{d}p \approx \int \left(\phi_{X'}(p) - \phi_X(p)\right)^2 \mathrm{d}p \approx \|\boldsymbol{\phi}_{X'} - \boldsymbol{\phi}_X\|^2, \tag{6}$$

where $w_x(p)$ depends on the direction of dx relative to p. That is, by minimizing the squared error the classification method is indirectly minimizing the boundary distance between the shapes, where the parts of the boundary are weighted by their importance. (Corners and other prominent features of the shape have larger influence zones and thus higher weight.)

To avoid "inside-out" shapes in the analysis ($-\phi_A(p) = \phi_{A^c}(p)$) we draw ideas from the non-negative orthogonal matching pursuit (NNOMP) of Bruckstein *et al.*[8] to solve Equation (4) with the additional constraint that $\mathbf{x} \geq 0$. Furthermore, we observe that $k\phi \notin \mathcal{S},\ k \neq \pm 1$, and assuming objects to be approximately aligned at the origin ($\nabla\phi_i(p) \approx p/\|p\|$) we have the approximation $\phi_i(p) \approx m + \|p\|$, where $m \in \mathbb{R}$. Looking at the error ϵ between a linear approximation and the SDF $\phi(p)$

$$\epsilon = \|\phi(p) - (h + \sum k_i\phi_i(p))\|_2, \tag{7}$$

and using the approximation for ϕ, we have a necessary condition for a minimum, $\sum k_i = 1$ and $h + \sum k_i m_i = m$. From this observation we constrain the linear combinations to convex combinations, by replacing the inversion step in NNOMP with

$$\hat{\mathbf{x}} = \arg\min_{x} \|\Phi\mathbf{x} - \mathbf{y}\|_2 \text{ s.t. } \|\mathbf{x}\|_1 = 1 \text{ and } \mathbf{x} \geq 0, \tag{8}$$

This can be solved by the quadratic program

$$\hat{\mathbf{x}} = \arg\min_{x} \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{f}^T\mathbf{x} \text{ s.t. } \mathbf{1}^T\mathbf{x} = 1,\ \mathbf{x} \geq 0\,, \tag{9}$$

where $H = 2\Phi_S^T\Phi_S$, $\mathbf{f}^T = -2\mathbf{y}^T\Phi_S$. This shape classification method is presented in Algorithm 1. MATLAB implementation and examples are provided on the authors' website. (`http://iacl.ece.jhu.edu/gunnar`)

The sparse recovery step maintains all of the properties of NNOMP outlined in[8] since it is a special case. Our modification makes the optimization both faster, since this uses a quadratic program, and less prone to overfitting, since it avoids modelling infeasible shapes.

Algorithm 1 Shape Classification using Sparse Convex Combinations

Require: $\mathbf{y}$ (Test shape), $\Phi = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K]$ (Dictionary), s (Sparsity), $\mathcal{C}$ (Classes)
  $\Phi = \Phi \cup \{\mathbf{1}, -\mathbf{1}\}$
  $\mathbf{y}_r^0 = \mathbf{y}$, $S^0 = \emptyset$
  for n=1 **to** s **do**
    $i_{\max} = \arg\max_{i \subset [N]\setminus S} \langle \boldsymbol{\phi}_i, \mathbf{y}_r^{n-1} \rangle / \|\boldsymbol{\phi}_i\|_2$
    $S^n = S^{n-1} \cup \{i_{\max}\}$
    $\hat{\mathbf{x}} = \arg\min_x \|\Phi_{S^n}\mathbf{x} - \mathbf{y}\|_2$ s.t. $\|\mathbf{x}\|_1 = 1$ and $\mathbf{x} \geq 0$
    $\mathbf{y}_r^n = \mathbf{y} - \Phi_{S^n}\hat{\mathbf{x}}$
  end for
  return $\hat{c} = \arg\min_{c \in \mathcal{C}} \|\mathbf{y} - \Phi\delta_c(\hat{\mathbf{x}})\|_2$ (Class), Φ_{S^n} (Similar shapes), $\hat{\mathbf{x}}_{S^n}$ (Similarity weights)

If dimensionality reduction is required, common methods (PCA, random projections) may not work because of having to store a large projection matrix in memory. Fortunately, we noticed that taking a random sub-set of the data performed well in our experiments, which is equivalent to multiplying the data vector in $\mathbb{R}^N$ by a sub-matrix of the identity matrix in $\mathbb{R}^{N\times N}$. This is because the SDF has some redundancy (see previous discussion on influence zones), and choosing a random sub-set can be thought of as sampling from these influence zones.

4. OPERATIONS ON SIGNED DISTANCE FUNCTIONS

Because our dictionary elements are SDFs and the space is approximated by convex combinations to produce sparse approximations to a given shape, it is important to understand how such linear combinations yield new shapes and what properties this approximation possesses.

4.1 Resizing (erosions and dilations)

Subtracting a constant h from an SDF—i.e., $\phi'(p) = \phi_A(p) - h$—yields a similar shape whose size is either larger ($h > 0$) or smaller ($h < 0$). In fact, the new shape is either the dilation or erosion of the original by a circular structuring element of size h.[9] For sufficiently small h, the function $\phi'(p)$ is an SDF. Interestingly, SDFs have the property that a linear combination of resized SDFs satisfies:

$$\phi(p) = \sum_i \phi_i(p) + h_i = h + \sum_i \phi_i(p)\,, \tag{10}$$

where $h = \sum h_i$. That is, the individual resizing operations accumulate to a single resizing operation. Therefore, by adding two constants to the dictionary (one positive and one negative), the framework will adjust for the size of the shape.

4.2 Blending (weighted averages)

Consider the non-negative linear combination of two SDFs: $\phi'(p) = k_1\phi_A(p) + k_2\phi_B(p)$, where $k_1, k_2 \geq 0$. Then by the definition of an SDF the zero level set of ϕ' must satisfy

$$\{p \mid \phi'(p) = 0\} = \{p \mid k_1\phi_A(p) = -k_2\phi_B(p)\} \ , \tag{11}$$

which implies that the zero level set of ϕ' (the shape boundary) is a fraction $k_1/(k_1 + k_2)$ of distance from B's boundary to A's boundary. Through selection of k_1 and k_2, a new shape A' can be produced that is a blend of A and B. For example, by choosing $k_1 \gg k_2$ the resulting shape A' is much more similar to A than B. The case $k_1, k_2 = 1/2$ has been used to define the average shape.[3]

Finally, it can be shown by direct substitution that the h level set of the shape ϕ' is

$$\{p \mid \phi'(p) = h\} = \left\{p \;\middle|\; k_1\left(\phi_A(p) - \frac{h}{k_1 + k_2}\right) = -k_2\left(\phi_B(p) - \frac{h}{k_1 + k_2}\right)\right\} . \tag{12}$$

If $k_1 + k_2 = 1$, then

$$\{p \mid \phi'(p) = h\} = \{p \mid k_1\left(\phi_A(p) - h\right) = -k_2\left(\phi_B(p) - h\right)\} \ , \tag{13}$$

which reveals that the h level set of the convex combination of ϕ_A and ϕ_B is the convex combination of the h level sets of ϕ_A and ϕ_B. This is a desirable property because it yields level set functions ϕ' that are approximately equal to level set functions (resized versions) of the blended shapes. This intuition breaks down for large h, where large negative h yields the empty set for small shapes; but it remains valid over large regions of shape support, especially near the boundaries of the shapes themselves. Given this observation, we will impose a key restriction in our sparse reconstruction of SDFs: they are to be restricted to convex combinations.

5. APPLICATIONS

5.1 Interpretability

A typical output from the method applied on cerebellar shapes is presented in Figure 1. An interpretation of the output is that the similarity weights and the classification, are based on the resized exemplars whose weighted average boundary most closely approximates the shape—resulting in an intuitive interpretation of the final result.

Figure 1: Elements selected from the dictionary by the method to approximate the shape on the left. The similarity weights are represented as the width of the border.

5.2 Vehicle Shapes

The first results come from a vehicle video segmentation dataset.[10] The dataset contains four types of automobiles. What makes this a hard dataset, is that the car segmentations include shadow effects and there is large intra-class size difference. Since the shapes were not aligned initially we performed simple iterative closest point registration to align the shapes before calculation. The sparsity was set to $s = 15$ (approximately the size of the smallest class). Results are presented in Figure 2. We used leave-one-out cross-validation. The method is compared against classification accuracy by shape descriptors such as curvature ($\kappa + \mathrm{SVM}$), Fourier descriptor ($\mathrm{FD} + \mathrm{SVM}$), Zernike moments ($\mathrm{ZM} + \mathrm{SVM}$), where support vector machine (SVM) was used for classification, finally we compare against the method from Thakoor *et al.*[10] (HMM_E+WtL).

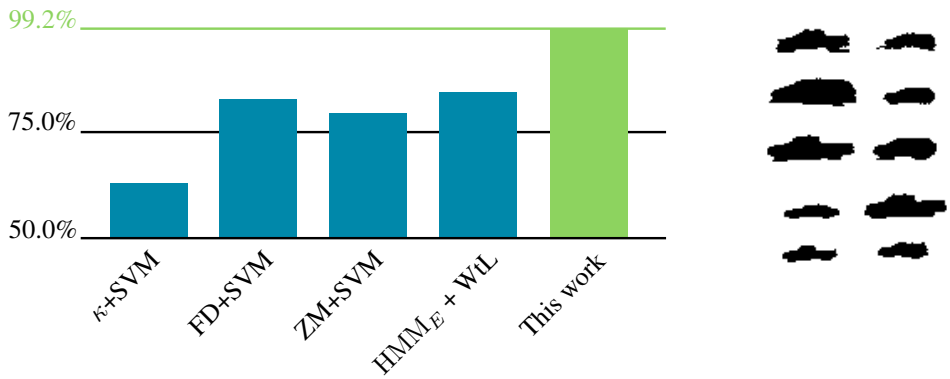

Figure 2: Classification accuracy for the vehicle video segmentation dataset along with example shapes.[10]

5.3 The ETH-80 dataset

The ETH-80 dataset[11] is a dataset composed of 80 objects from 8 categories. Each object is represented by 41 images from different viewpoints. The validation of an algorithm is intended to be based on a leave-one-object-out cross-validation. The sparsity parameter was set to $s = 15$, and for dimensionality reduction we used a random 10^4 dimensional sub-set ($\mathrm{RS}10^4$) (down from 7×10^6). The following algorithms are presented in Figure 3: PCA on binary masks,[11] shape context approach using dynamic programming[11] (SC + DP), Inner-Distance Shape Context with dynamic programming[12] (IDSC + DP), Height Functions,[13] as well as Symbolic Representation.[14] All of these methods use only the shape of the objects. It is clear that the proposed framework offers interpretability and performance that is at par with the state of the art.

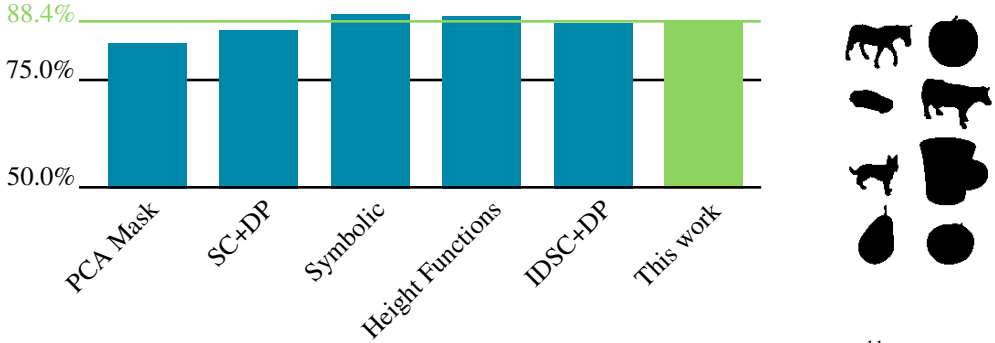

Figure 3: Accuracy for the ETH-80 dataset along with example shapes.[11]

5.4 3D shape classification

To demonstrate the applicability to classification of 3D deformable shapes, the method was applied on a dataset consisting of automatically segmented 3D human cerebellum masks (T1-weighted MRI). An example 3D cerebellum shape along with axial slices from the four classes are presented in Figure 4. The cerebellum masks were aligned by rigid transformations and the brain stem was not included in the mask. The dataset consists of 48 controls, 12 spinocerebellar ataxia type 2 patients (SCA2), 14 spinocerebellar ataxia type 6 patients (SCA6), and 19 ataxia-telangiectasia patients (AT). The confusion matrix obtained by leave-one-out cross-validation is presented in Table 1. Since there is no accepted way of performing 3D shape classification, traditional classification baseline techniques were applied to the dataset using both the binary representation (Binary) and the signed distance transform (SDF). In Figure 4 classification accuracy is presented for k-Nearest Neighbor (kNN), Linear Discriminant Analysis (LDA), Random Forest Classifier (RF), and direct application of Wright *et al.*[6] on binary shapes (Binary + OMP). Furthermore, principal component analysis (PCA) was used for dimensionality reduction in some cases. These results demonstrate that the proposed method is at par with the best baseline method, and importantly, is much easier to interpret than for example the LDA decision boundary in the principal component space.

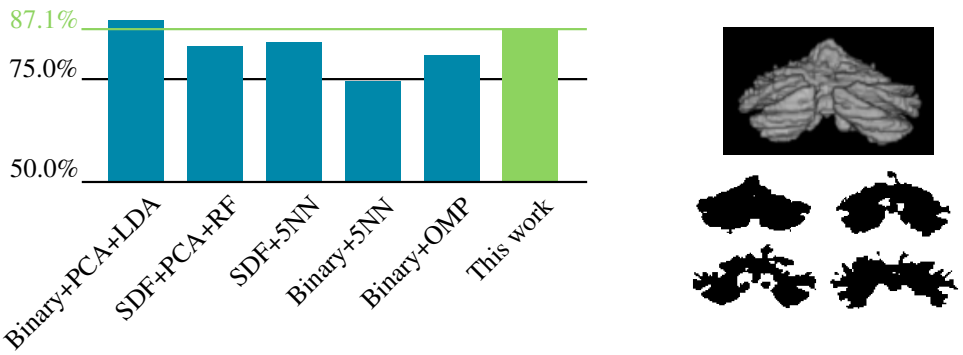

Figure 4: Classification accuracy for the cerebellum dataset along with example shapes.

Table 1: Confusion matrix for the cerebellar disease classification task.

		Prediction			
		Controls	SCA2	SCA6	AT
Truth	Controls	100%	0%	0%	0%
	SCA2	8.3%	83.3%	0%	8.3%
	SCA6	14.3%	0%	71.4%	14.3%
	AT	21.1%	0%	10.5%	68.4%

6. SUMMARY AND CONCLUSION

We proposed a dictionary-based shape classification method by combining sparse recovery with signed distance functions in a holistic framework. The method uses the full information of the shape, which simplifies analysis and feature selection. By exploiting properties of the shape representation, a variation of the popular Orthogonal Matching Pursuit was proposed, which offers fast and robust results in practice. We demonstrated the performance of the method in both two and three dimensional classification tasks, and showed that it achieves state of the art performance. Most importantly, the presented method allows for resizing and blending of arbitrary shapes from the dictionary to approximate new shapes. The weight of each shape directly represent its contribution to the final shape, which helps analysis and improves interpretability. Further research might expand this work to generalize over other possible classes of shapes, such as their unions.

Acknowledgements

This work was supported in part by NIH/NINDS 2R01NS056307-06A1. The authors would like to thank Yuanming Suo and John Goutsias, Johns Hopkins University, for invaluable suggestions.

REFERENCES

[1] Gorelick, L., Galun, M., Sharon, E., Basri, R., and Brandt, A., "Shape representation and classification using the poisson equation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(12), 1991–2005 (2006).

[2] Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D. N., and Zhou, X. S., "Towards robust and effective shape modeling: Sparse shape composition," *Medical image analysis* **16**(1), 265–277 (2012).

[3] Leventon, M. E., Grimson, W. E. L., and Faugeras, O., "Statistical shape influence in geodesic active contours," in [*Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*], **1**, 316–323, IEEE (2000).

[4] Golland, P., Grimson, W. E. L., Shenton, M. E., and Kikinis, R., "Small sample size learning for shape analysis of anatomical structures," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*], 72–82, Springer (2000).

[5] Tsai, A., Wells, W. M., Warfield, S. K., and Willsky, A. S., "An em algorithm for shape classification based on level sets," *Medical Image Analysis* **9**(5), 491–502 (2005).

[6] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y., "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **31**(2), 210–227 (2009).

[7] Ye, Q.-Z., "The signed euclidean distance transform and its applications," in [*Pattern Recognition, 1988., 9th International Conference on*], 495–499, IEEE (1988).

[8] Bruckstein, A. M., Elad, M., and Zibulevsky, M., "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *Information Theory, IEEE Transactions on* **54**(11), 4813–4820 (2008).

[9] Goutsias, J. and Batman, S., "Morphological methods for biomedical image analysis," *Handbook of medical imaging* **2**, 175–272 (2000).

[10] Thakoor, N., Gao, J., and Jung, S., "Hidden markov model-based weighted likelihood discriminant for 2-d shape classification," *Image Processing, IEEE Transactions on* **16**(11), 2707–2719 (2007).

[11] Leibe, B. and Schiele, B., "Analyzing appearance and contour based methods for object categorization," in [*Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*], **2**, II–409, IEEE (2003).

[12] Ling, H. and Jacobs, D. W., "Shape classification using the inner-distance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **29**(2), 286–299 (2007).

[13] Wang, J., Bai, X., You, X., Liu, W., and Latecki, L. J., "Shape matching and classification using height functions," *Pattern Recognition Letters* **33**(2), 134–143 (2012).

[14] Daliri, M. R. and Torre, V., "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognition* **41**(5), 1782–1798 (2008).