

# Probabilistic Mixtures of Differential Profiles for Shape Recognition

Lei Ding      Mikhail Belkin

Department of Computer Science and Engineering  
The Ohio State University  
Columbus, OH 43210 USA

{dinglei,mbelkin} @ cse.ohio-state.edu

## Abstract

*We present a new shape descriptor, the differential profile, for shape representation and recognition, which is derived from differential and geometric quantities evaluated at points on a shape. We model differential profiles from a class of shapes as a finite mixture of Gaussians, and use an expectation-maximization (EM) procedure for class-wise model learning. Experiments on handwritten digit and 3D object datasets give promising results.*

## 1 Introduction

Shape representation and recognition are fundamental problems in computer vision. Over the years, numerous approaches have been proposed. They have achieved success in a broad range of applications from character recognition to object categorization to medical image analysis, etc. However, statistical work [2] often assumes that point correspondences are known, which is not always the case in practice. Computing such correspondences between two shapes can incur a heavy computational cost in optimization (e.g., the assignment problem in [1]). We also note that a method for shape analysis should, similarly to the human visual system, be robust to various common transformations (translation, scaling and rotation as needed) and require few examples for achieving accurate recognition. To address these problems, we develop a new shape descriptor based on smoothing the shape by convolving it with a two-dimensional Gaussian function and using the differential and geometric properties of the resulting convolution interpreted as a surface.

The problem of shape representation is to use a suitable vector representation for local shape features at a point or context features of a point in relation to other points on the same shape. A 2D shape  $P$  is defined as a

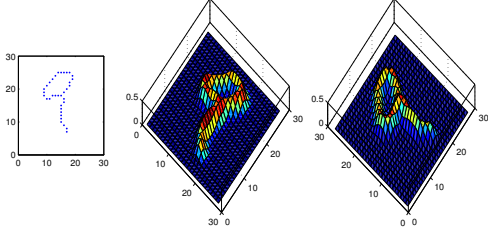
discrete planar set of points  $P = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ . Various methods have been proposed for extracting features from shapes, commonly known as shape representation techniques. Shape context [1] is a popular recent descriptor: the shape context at a reference point is computed from the set of remaining  $n - 1$  points; that is, at each point  $\mathbf{p}_i$  we compute a spatial histogram  $h_i$  of the relative coordinates of all other points in a shape. Formally,  $h_i(k) = \#\{\mathbf{q} \neq \mathbf{p}_i | (\mathbf{q} - \mathbf{p}_i) \in \text{bin}(k)\}$ . The histogram  $h_i$  is defined as the shape context of  $\mathbf{p}_i$ . Typically, bins that are uniform in a *log-polar* space are used, which makes shape context descriptors more sensitive to positions of points near the reference point than to those of points farther away.

After two shapes are represented as appropriate features, similarity between them can be quantified as a real number. This problem is posed as comparing two sets of feature vectors, and various distance measures can be used: Hausdorff distance [3], Earth Mover's Distance [10], etc. In this paper, however, we do not use any pairwise shape distance measure. We choose to model the set of shape descriptors from one class as a mixture of Gaussians. The advantages are clear: first, one gains an understanding of important shape structures from the learned models; second, learning mixture models is more efficient on a large dataset than pairwise distance computation, which is consuming even with a simple distance measure such as Hausdorff distance.

## 2 Shape Characterizations

### 2.1 A Continuous Shape Model

A planar shape may be hard to analyze with differential tools since the pixel representation of a shape is not continuous and shapes often contain more than one simple curve. A simple way to deal with these difficulties is to transform a shape into a surface (Figure 1) in a three-dimensional space. After this transformation, it is natu-



**Figure 1. A planar shape transformed into a surface (viewed from two aspects) by Gaussian smoothing.**

ral to compute derivatives of the corresponding function and the geometric quantities characterizing the surface. This is the intuition behind the proposed methodology.

Mathematically a shape  $I$  can be written as a function:  $I(x, y) = \sum_i \delta(x - x_i, y - y_i)$ , where  $x_i$ 's and  $y_i$ 's are the coordinates for points composing the shape, and  $\delta(\cdot, \cdot)$  is Dirac's delta function.  $I(x, y)$  is a function whose domain is a finite subset of  $R^2$ , but it is not differentiable. In order to render it differentiable, we can smooth it with, for example, a 2D Gaussian function:  $\Gamma(x, y) = I * g(x, y)$ . Therefore, for any shape  $I(x, y)$  we can convert it into  $\Gamma(x, y)$ , which also defines a surface  $(x, y, \Gamma(x, y))$  in a three-dimensional space under the Monge-patch representation and is rich in analyzable geometric quantities.

## 2.2 Basic Quantities

Derivatives of a continuous intensity function have been studied in [4, 6] and applied to problems such as image retrieval [11], but are less common for shape-based applications. According to the Taylor's expansion, we can write, for instance, around the point  $(0, 0)$ :

$$\Gamma(x, y) = \Gamma + \Gamma_x x + \Gamma_y y + \frac{1}{2}(\Gamma_{xx} x^2 + 2\Gamma_{xy} xy + \Gamma_{yy} y^2) + \frac{1}{6}(\Gamma_{xxx} x^3 + 3\Gamma_{xxy} x^2 y + 3\Gamma_{xyy} x y^2 + \Gamma_{yyy} y^3) + \dots$$

The coefficients  $\Gamma$ ,  $\Gamma_x$ ,  $\Gamma_y$  and higher-order terms define how the function behaves in a small neighborhood around a given point. The following identity is used:

$$\Gamma_{x^m y^n}(x, y) = \partial_{x^m} \partial_{y^n} (I * g) = I * \partial_{x^m} \partial_{y^n} g(x, y).$$

We will use the following Gaussian function:  $g(x, y) = \frac{1}{2\pi} \exp(-(\frac{x^2}{2\sigma_1^2} + \frac{y^2}{2\sigma_2^2}))$ . Partial derivatives of the Gaussian function  $g(x, y)$  satisfy the recurrent equations:

$$g_x(x, y) = -xg(x, y)/\sigma_1^2, \quad (1)$$

$$g_{xy}(x, y) = -xg_y(x, y)/\sigma_1^2 = -yg_x(x, y)/\sigma_2^2. \quad (2)$$

Therefore all coefficients in the expansion can easily be expressed in terms of the Gaussian functions and point locations. The quantities  $\psi_{x^m y^n}(x, y) = \sigma_1^m \sigma_2^n \Gamma_{x^m y^n}(x, y)$  are called *scale-normalized* derivatives, as they are not dependent on the scale parameters.

Next, we introduce quantities in the two fundamental forms describing the geometry of the surface [5]. We define the surface  $\mathbf{r} = \mathbf{r}(x, y) = (x, y, I * g(x, y))$ . Quantities in the first fundamental form are:  $E = \mathbf{r}_x \cdot \mathbf{r}_x$ ,  $F = \mathbf{r}_x \cdot \mathbf{r}_y$ ,  $G = \mathbf{r}_y \cdot \mathbf{r}_y$ . The unit normal vector  $\mathbf{n}$  at a point on the surface is length normalized  $\mathbf{r}_x \times \mathbf{r}_y$ . Then, quantities in the second fundamental form are:  $L = \mathbf{r}_{xx} \cdot \mathbf{n}$ ,  $M = \mathbf{r}_{xy} \cdot \mathbf{n}$ ,  $N = \mathbf{r}_{yy} \cdot \mathbf{n}$ .

## 2.3 Invariants

Koenderink et al. [4] and Florack et al. [6] theoretically established differential invariants of a gray-scale image as an intensity function. The polynomial invariants (besides the function value  $\Gamma$  itself) under the  $SO(2)$  group of transformation involving only up to second order derivatives are listed as:

$$\nu_1 = \Gamma_x^2 + \Gamma_y^2, \quad (3)$$

$$\nu_2 = \Gamma_x^2 \Gamma_{xx} + \Gamma_y^2 \Gamma_{yy} + 2\Gamma_x \Gamma_y \Gamma_{xy}, \quad (4)$$

$$\nu_3 = \Gamma_{xx} + \Gamma_{yy}, \quad (5)$$

$$\nu_4 = \Gamma_{xx}^2 + \Gamma_{yy}^2 + 2\Gamma_{xy}^2. \quad (6)$$

We note that  $\nu_1$  is the squared norm of the gradient, while  $\nu_3$  is the Laplacian of  $\Gamma$ .

Besides, from the six quantities in the fundamental forms, we can compute Gaussian curvature  $K$  and mean curvature  $H$  at each point on the surface.

$$K = \frac{LN - M^2}{EG - F^2}, \quad (7)$$

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)}. \quad (8)$$

The quantities are also invariant under  $SO(2)$  and describe the basic geometry of the surface at each point. According to Gauss's *Theorema Egregium*,  $K$  is invariant under all isometric transformations of the surface.

## 3 Differential Profile

### 3.1 Definition

We consider the the problem of shape description at each point  $(x, y)$  of a 2D shape as constructing a feature vector in a higher-dimensional space, to which we wish to transform a coordinate pair  $(x, y)$  to represent the point in the whole shape. This representation should have a combination of invariance for the same class of shapes under view changes and/or slight nonlinear

transformation as well as discrimination with respect to different classes of shapes. We define the *differential profile* at each point  $(x, y)$  on a shape to be the following feature vector  $DP(x, y)$ ,

$$\underbrace{(\psi, \psi_x, \psi_y, \psi_{xx}, \psi_{xy}, \psi_{yy}, E, F, G, L, M, N, \nu_1, \nu_2, \nu_3, \nu_4, K, H)}_{\text{basic differential \& surface quantities}} \underbrace{\hspace{10em}}_{\text{differential \& surface invariants}}$$

Hence, a shape can be represented using the differential profile as a bag of feature vectors. We note that although this representation describes mainly the local properties of a shape at a point and its vicinity, it does take into account locations of all points on the shape as it is. However, close points have higher influence due to the shape of the Gaussian function.

### 3.2 Invariance Analysis

Invariance to transformations largely defines the empirical robustness of a shape representation. Given that the shape  $I(x, y) = \sum_i \delta(x - x_i, y - y_i)$ , we have the following for the scale-normalized derivatives:

$$\psi_x(x, y) = \sum_i -\frac{x - x_i}{\sigma_1} g(x - x_i, y - y_i), \quad (9)$$

$$\psi_{xy}(x, y) = \sum_i \frac{(x - x_i)(y - y_i)}{\sigma_1 \sigma_2} g(x - x_i, y - y_i). \quad (10)$$

Hence, these features in the differential profile give a smoothed sum of a point’s normalized coordinate relative to all points in a shape and various combinations. By properly chosen Gaussian bandwidths (e.g.,  $\sigma_1$  and  $\sigma_2$  being a constant fraction of the length and width of a shape respectively), these will lead to both translation and scale invariant description. In reality, if all shapes in a dataset have roughly the same size (this can also be achieved by normalizing shapes), then the two bandwidths become constants. For instance, we can use a unit bandwidth for each.

Therefore, the differential profile descriptor at each point has two components: its extrinsic part of basic quantities is scale and translation invariant, while its intrinsic part, including differential and surface invariants, is  $SO(2)$  invariant and Gaussian curvature, in particular, is isometrically invariant.

### 3.3 Generative Shape Class Models

When it comes to shape comparison, relative positions of features contribute to the formation of a shape and hence to shape recognition. However, modeling this



Figure 2. Left: objects in ETH-80. Right: an object, its silhouette and its contour.

positional information incurs a much higher computational cost, which is not always justified. Therefore we choose a simple *bag-of-features* model for overall shape representation and a generative Gaussian mixture model for the set of feature descriptors from each class.

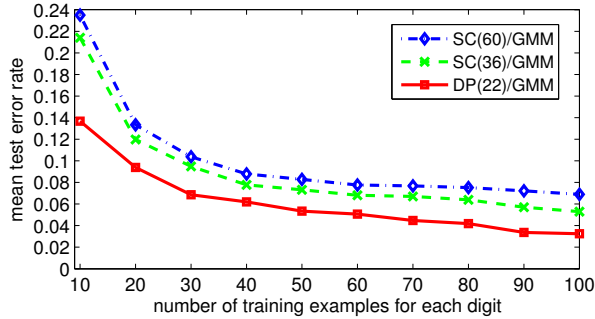
Specifically, we model shape descriptors ( $\rho \in \mathbb{R}^d$ ) from one shape class as a finite mixture of Gaussian components:  $\rho \sim \sum_j \alpha_j \mathcal{N}(\mu_j, \Sigma_j)$ . A shape representation  $s_i = \{\rho_1, \rho_2, \dots, \rho_{n_i}\}$  consists of a variable collection of such features, where  $n_i$  denotes the number of features composing the shape. Computationally, each class model of features is individually learned by a standard EM procedure initialized with K-means estimates. In the test phase, features of one shape are plugged into all the learned class models, and the class with the highest average log-likelihood is reported.

Besides the advantage in efficiency of using probabilistic models over pairwise shape comparison, one other nice property is that the relative weights of features in a shape descriptor are implicitly determined by the covariance estimation of each component. For differential profile, it would otherwise be difficult to properly assign weights to the basic quantities and the invariants. Therefore, we consider Gaussian mixtures a natural choice for modeling shape descriptors.

## 4 Experiments

We use MNIST [7] dataset of handwritten digits (60000 training and 10000 test digits) and ETH-80 [8] dataset of 3D objects, having 8 categories of common objects. Each category contains 10 objects with 41 views per object, spaced equally over the viewing hemisphere. Features are extracted from points on the skeleton (for a digit) or on the contour (for an object).

The first experiment is on MNIST. We compare results of the differential profile descriptor (DP) with



**Figure 3. Test error rates on MNIST (dimensionality of descriptors shown in the legend).**

those of the shape context descriptor (SC). We use 10 to 100 random labeled examples for each digit from the training set and predict a label for each digit in the test set. For better discrimination, we also include third-order extrinsic derivatives in the differential profile descriptor. Multiple runs are performed in each case and means of test set error rates are plotted in Figure 3. As shown, results from the proposed DP are considerably better than those from normalized shape context. In what follows we have a comparison with the results in a most recent paper [9] that uses a learned hierarchy of invariant features for doing recognition: with 300 labeled examples, our mean error rate is 6.86%, while theirs is 7.18%; with 1000 labeled examples, our mean error rate is 3.25%, while theirs is 3.21%. The second experiment involves recognition of 3D objects from the ETH-80 dataset. The test mode is leave-one-out cross-validation, in which we train with 79 objects and test with the remaining one object to see if the correct class label can be assigned. Results are averaged over all 80 test objects, reported in the Table 1 and compared with results from [8] using the same training/testing procedure but with shape context as the descriptor.

## 5 Conclusion

We propose a new differential profile descriptor for shape representation and recognition. Features are modeled by Gaussian mixtures, which are efficiently learned via an EM procedure. In experiments, our descriptor performs better than the shape context descriptor with the same Gaussian mixture modeling as well as standard shape distance measures.

| objects | grSC [8]     | dySC [8]      | pSC    | pDP           |
|---------|--------------|---------------|--------|---------------|
| apple   | 22.93%       | 23.66%        | 21.49% | <b>21.25%</b> |
| pear    | 9.27%        | <b>8.29%</b>  | 12.20% | 9.66%         |
| tomato  | 29.27%       | 29.76%        | 30.06% | <b>23.51%</b> |
| cow     | 13.17%       | 13.66%        | 16.59% | <b>9.76%</b>  |
| dog     | 18.05%       | <b>17.07%</b> | 20.67% | 20.82%        |
| horse   | 15.37%       | 15.37%        | 19.54% | <b>13.23%</b> |
| cup     | <b>0.24%</b> | 0.98%         | 2.74%  | 2.32%         |
| car     | 0.49%        | <b>0.00%</b>  | 1.89%  | 2.38%         |
| average | 13.60%       | 13.60%        | 15.65% | <b>12.87%</b> |

**Table 1. Test error rates on ETH-80. Left two columns show results using shape context with greedy matching and dynamic programming for distance computing. Right two columns show results using Gaussian mixtures of SC and DP respectively.**

## References

- [1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE PAMI*, 24(4):509–522, 2002.
- [2] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley, Chichester, 1998.
- [3] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE PAMI*, 15(9):850–863, 1993.
- [4] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.
- [5] W. Kuhnel. *Differential Geometry*. American Mathematical Society, 2002.
- [6] L. M. J. Florack and B. M. ter Haar Romeny and J. J. Koenderink and M. A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4(2):171–187, 1994.
- [7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. of IEEE*, 86(11):2278–2324, 1998.
- [8] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, 2003.
- [9] M. Ranzato, F. Huang, Y. Boureau, and Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *CVPR*, 2007.
- [10] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *ICCV*, 1998.
- [11] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE PAMI*, 19(5):530–534, 1997.