Learning Shape Detector by Quantizing Curve Segments with Multiple Distance Metrics

Ping Luo[1,2], Liang Lin[1,2,⋆], and Hongyang Chao[1]

[1] School of Software, Sun Yat-Sen University, Guangzhou 510275, P.R. China
[2] Lotus Hill Research Institute, P.R. China
pluo.lhi@gmail.com, linliang@ieee.org, isschhy@mail.sysu.edu.cn

Abstract. In this paper, we propose a very efficient method to learn shape models using local curve segments with multiple types of distance metrics. Our learning approach includes two key steps: feature generation and model pursuit. In the first step, for each category, we first extract a massive number of local "prototype" curve segments from a few roughly aligned shape instances. Then we quantize these curve segments with three types of distance metrics corresponding to different shape deformations. In each metric space, the quantized curve segments are further grown (spanned) into a large number of ball-like manifolds, and each of them represents a equivalence class of shape variance. In the second step of shape model pursuit, using these manifolds as features, we propose a fast greedy learning algorithm based on the information projection principle. The algorithm is guided by a generative model, and stepwise selects the features that have maximum information gain. The advantage of the proposed method is identified on several public datasets and summarized as follows. (1) Our models consisting of local curve segments with multiple distance metrics are robust to the various shape deformations, and thus enable us to perform robust shape classification and detect shapes against background clutter. (2) The auto-generated curve-based features are very general and convenient, rather than designing specific features for each category.

1 Introduction

Although many shape descriptors have been proposed for distortion and deformation measurement, learning shape detector incorporating with multiple types of distance metrics has been rarely addressed in previous work. This paper presents a novel learning-based shape detector for detecting and matching shapes from cluttered edge maps.

In the following, we briefly review the previous work for (i) shape descriptors (or similarity measurements) and (ii) learning shape models.

(i) Many shape matching problems are posed as minimizing the distance measures of deformation and bending by searching corresponding points between two shapes. Most of these distance measures are mainly defined on the spaced landmarks of shape boundaries and designed to account for various shape transformation. For example, the procrustes distance [9] is very robust to the rigid affine transformation; the inner distance [13] and shock graph distance [21] capture the articulation transformation very well. Recently, to deal with more complex non-rigid shape deformations

⋆ Corresponding author.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part III, LNCS 6313, pp. 342–355, 2010.

and configurations, the context and hierarchy of shapes have been the theme of recent work [1, 7, 17, 16]. Despite of the acknowledged success of these methods, it is still an open problem to adaptively select the proper shape distances corresponding to different shape categories.

(ii) Early work on learning shape models include learning contour groups with perception organization [19, 26] and learning global modes of variation with the Active Shape Models [3]. In the research of object detection, contour-based features were widely adopted due to their large invariance against lighting conditions and variations in object color and texture [22,8]. However, in the context of noisy edge maps and background clutter, shape contours are often considered as been less discriminative. Some recent hierarchical or part-based object models [10, 23] prefer region (or appearance) features.

In this paper, we argue that automatic object recognition was indeed achievable by employing only shape contour information. Incorporating with multiple types of distance metrics, the proposed shape detectors are robust to capture various shape deformations, and thus enable us to perform stable shape classification and detect shapes against background clutter. Our approach includes two key steps: (i) automatic feature generation and (ii) generative shape model pursuit.

In the first step, for each category, we first extract a massive number of local "prototype" curve segments from roughly aligned shape instances and quantize each of them with three types of different descriptors, i.e. procrustes distance [18], articulation distance and geodesic distance [12]. This is inspired by the classical work of shape analysis [6], which shows that the arbitrary deformation of a shape curve (or contour) can be decomposed into three types: the rigid (affine) transformation, articulation transformation, and distortion (twist). The three distances we employed are proven to accordingly capture the three typical transformations very well. In the rest of this paper, we call these "prototype" curve segments as proto-curves for simplification.

In the perspective of mathematics, each proto-curve quantized by a descriptor can be viewed as a point in the metric space, where this point can be further spanned into a manifold by introducing a statistical fluctuation ϵ. As illustrated in Fig.1 (a), we visually define the manifold centering at a proto-curve as an "ϵ-ball", in the sense that the ball-like manifold is essentially an equivalence class of the proto-curve in the metric space. Moreover, each ϵ-ball encodes the relative location (i.e. global spatial configuration) of the proto-curve with respect to the center of the shape that we extracted the proto-curve from, inspired by the Implicit Shape Model [14]. Therefore, given an input shape, each ϵ-ball can be further defined as a "visual feature" or classifier that decides whether the testing shape has the similar local deformation corresponding to the ϵ-ball.

In the second step of shape model pursuit, we propose a fast greedy feature selection algorithm based on the information projection principle [4]. For each shape category, the training set consists of a small number of positive samples and a certain amount of reference samples chosen over all categories. The algorithm is guided by a generative shape model based on the Pietra's representation [20], in that each feature (ϵ-ball) captures the shape variance explicitly and generatively. In our learning algorithm, different types of features, (i.e. proto-curves quantized by different metrics), are made comparable to each other by an information gain criterion; the shape model pursuit is formulated

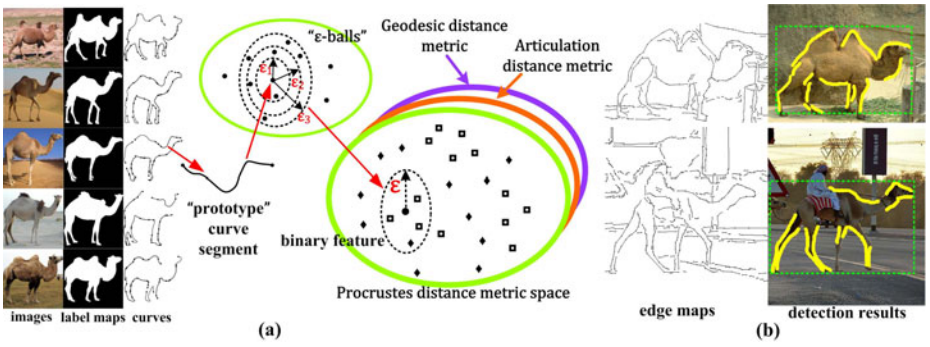

Fig. 1. (a) Illustrates that we extract "prototype" curve segments from the roughly aligned shapes and quantize each of them with three metrics. Each quantized proto-curve is spanned to a few manifolds. We name each manifold "ϵ-ball" and define it as binary feature. Here the circle, square and diamond denote the proto-curve, positive sample and reference (negative) sample respectively. (b) shows two significant clutter edge maps and their corresponding detection results. The detected curves and shape bounding box are plotted in yellow and green respectively.

as the procedure of maximizing the log-likelihood ratio of positive samples against the reference samples, with stepwise feature selection. By pruning the correlated features within the feature selection, we assume that the likelihood ratio can be factorized into individual likelihood ratios of the features. As a result, the shape model is in the form of the weighted sum of a small number of ϵ-balls. In the testing stage, given a learned shape model, we adopt the sliding window approach to fast localize and match shapes from clutter edge maps, as shown in Fig.1 (b).

The key contribution of this paper is summarized as follows. (1) We propose a general approach to produce shape features by growing manifolds from curve segments in different distance metrics. (2) We present a simple algorithm to learn generative shape models consisting of multiple types of features by an information gain criterion. (3) Our approach is tested on two challenging datasets, such as the ETHZ shape dataset [8] and a 40 categories image dataset chosen from LHI database [24], and shows the-state-of-the-art performance.

The remainder of this paper is arranged as follows. We first introduce our curve-based features in sec.2, including three distance metrics and feature generation. The algorithm for pursuing shape models and a shape matching algorithm are proposed in sec.3. The experimental evaluations are presented in sec.4. The paper concludes with a summary in sec.5.

2 Feature Generation via ϵ-Balls

In this section, we will introduce the procedure of curve-based feature generation with three types of shape distances.

In our method, a shape $\mathbf{S}$ is represented by a batch of curve segments $\{\mathbf{c}\}$. As illustrated in Fig.2, a curve segment c from the shape $\mathbf{S}$ is described as a two tuple

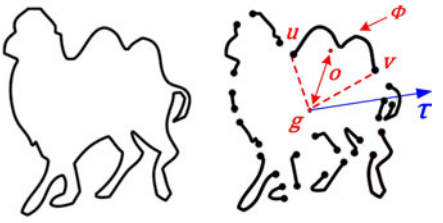

Fig. 2. A shape **S** is represented by a batch of curve segments. We encode the relative position of the curve segment with respect to the shape. u, v denote two end points of the curve, the mass center of the shape is g and the orientation of a shape is τ.

$\{\Phi, \Gamma = (\theta_u, \theta_v, o)\}$, where Φ is the set of interpolated landmarks along c. $\Gamma = (\theta_u, \theta_v, o)$ indicates the relative position of the curve segment c with respect to the shape **S**. Supposing u, v denote two end points of c, the mass center of the shape is g and the orientation of the shape is τ, θ_u denotes the relative angle between gu and τ, and θ_v is defined similarly at end point v. o is the offset of the curve centroid related to g. Note that the orientation of a shape can be calculated by the PCA method.

At the first step of feature generation, we extract a number of curve segments, namely, proto-curve, from the shape instances in the training set. We denote this proto-curve as boldface letter **c**. Then we quantize each curve segment with three different distances corresponding to various shape deformation. In the three metric spaces, each quantized proto-curve **c** is further spanned into a number of ball-like manifold, called "ϵ-ball", as follows,

$$\Omega^w(\mathbf{c}) = \{c : \mathcal{D}^w(c, \mathbf{c}) < \epsilon\}, \tag{1}$$

where $w \in \{'p', 'a', 'g'\}$ indicates the type of distance metrics, i.e. procrustes distance, articulation distance and geodesic distance. We will introduce three distance metrics later on. Each ϵ-ball can be viewed as an equivalent class bounded by residual ϵ, in which each element c may share the same statistical characteristics with respect to **c**.

Furthermore, an ϵ-ball can be naturally transformed to a binary feature (weak classifier), $h_i = \Omega^w(\mathbf{c}), i = 1, \ldots M$, ($M$ indicates the size of the feature set), and given a testing shape **S**, its response is defined as,

$$r_i(\mathbf{S}) = \begin{cases} 1, \mathcal{D}^w(c', \mathbf{c}) < \epsilon, \exists c' \in \mathbf{S} \; s.t. \; \Gamma_{c'} \approx \Gamma_{\mathbf{c}}, \\ 0, \qquad\qquad\qquad otherwise, \end{cases} \tag{2}$$

where $\Gamma_{c'} \approx \Gamma_{\mathbf{c}}$ indicates that two curve segment c' and **c** have the similar relative position with respect to the shape. Intuitively, one shape **S** is predicted as positive by the feature h is equivalent to existing a curve c' in **S** that has the almost the same relative position as well as the same similarity the distance metric similar deformation compared to the proto-curve **c**. Unlike the discriminative boundaries in many previous work [8], the proposed manifold features "keep silence" (equal to 0) to a shape **S** not falling into them.

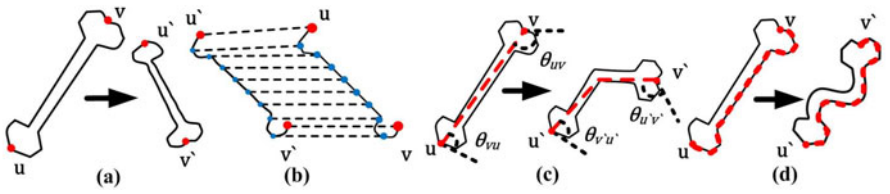

Fig. 3. (a) Shows the rigid transformation between two bones and (b) shows bijection correspondence between two curve segments using procrustes analysis. (c) and (d) illustrate the articulated transformation and distortion (twist) between two shapes respectively.

2.1 Quantizing Curve Segments

For each proto-curve $\mathbf{c}$, we map it into three different metric space, in that the proto-curves are transformed as a quantized point. It worth mentioning that though the various distance metrics are not entirely uncorrelated, they capture different characteristics of the shape.

Procrustes distance metric: $\Omega^p(\mathbf{c})$**.** We adopt the squared Procrustes distance [6] to measure the goodness of match between a pair of curve segments. By writing the coordinates of $x_i = (\xi_i, \eta_i)$ in c and $y_i = (\xi_i^{'}, \eta_i^{'})$ in $\mathbf{c}$ in complex form, namely, X and Y , respectively, we have

$$\mathcal{D}^p(c, \mathbf{c}) = 1 - \frac{|Y^* \cdot X|^2}{Y^* \cdot Y \cdot X^* \cdot X}, \tag{3}$$

where X^* and Y^* are the conjugate forms of X and Y. One exemplar of this metric is illustrated in Fig.3 (a), which shows the rigid transformation between two bones. Fig.3 (b) exhibits bijection correspondence of curves computed by procrustes analysis [9].

Articulation distance metric: $\Omega^a(\mathbf{c})$. In order to capture articulation invariance shown in Fig.3 (c) between a pair of curve segments, we design articulation distance metric by employing three geometrical shape descriptors as,

$$\mathcal{D}^a(c, \mathbf{c}) = \min \mid \Upsilon(c) - \Upsilon(\mathbf{c}) \mid, \tag{4}$$

where $\Upsilon(\cdot)$ denotes a six dimensional vector of the curve segment, (ID,A1,A2,S1,S2,S3), that combines the following three shape descriptors together.

• *Inner-distance between the ends (ID):* The traditional computing process of the inner distance [13] which refers to the shortest path between a pair of points within the whole shape silhouette is suitable for label maps during the learning procedure, but suffers pain when dealing with clutter edge maps in the testing stage, since there is no information about the inner and the outer parts (Fig.1 (b)). Thus we improve it by finding two shortest pathes between u, v following (i) respectively building two weighted undirected-graphs with the landmarks as their nodes on both sides of the curve, and (ii) individually applying the shortest route algorithm (e.g. Bellman-Ford) over these graphical structures.

• *Relative angles (A1, A2):* As shown in Fig.3 (c), we achieve the angles A1 = θ_{uv}, A2 = θ_{vu} between the path uv and the tangents at u and v respectively.

• *Articulated-invariant curve signature (S1, S2, S3):* Let d^{in} be the inner distance matrix of a curve segment calculated between each pair of landmarks. While mapping d^{in} into Euclidean distance space d^{eu} by multidimensional scaling (MDS), the landmark point set Φ will be transformed into a new one Φ', which can be computed by minimizing the following equation,

$$\Phi'^{*} = \arg\min_{\Phi'} \sum_{i}^{|\Phi|} \sum_{j}^{|\Phi|} \frac{(d^{in}(i,j;\Phi) - d^{eu}(i,j;\Phi'))^2}{d^{in}(i,j;\Phi)^2}. \tag{5}$$

And the articulated-invariant curve signature is defined to be a triple (S1, S2, S3), that it is the l_2-norm between $< u', v' >$, $< u', v'_{cen} >$ and $< v'_{cen}, v' >$ respectively, where $u', v' \in \Phi'$ and v'_{cen} is the center point of the mapped curve segment.

Geodesic distance metric: $\Omega^g(\mathbf{c})$. The geodesic distance between each pair of points on a 3D shape keeps stationary even though the shape is distorted. And this would be the same case if two 2D shapes to be matched have approximately identical view. Thus, we use the contour distance as an analogue to model distortive transformation (Fig.3 (d)). And $\mathcal{D}^g(c, \mathbf{c})$ indicates the Euclidean distance between the contour length of c and the length of the proto-curve. Due to different lengths of the curve segments, $\Omega^g(\mathbf{c})$ has been proved a discriminative distance metric in practice (see sec.4).

2.2 Feature Evolution

We conduct a procedure called "feature evolution" to calculate the residual ϵ for each manifold feature $\Omega^w(\mathbf{c})$. In practice, we generate three ϵ-balls for each proto-curve $\mathbf{c}$ in each metric space.

Recall that the all proto-curves are quantized points in the metric space. Intuitively, for each $\mathbf{c}$, we grow the ϵ starting from an initial small number, and the more neighboring proto-curves will fall into the growing ball when the ϵ increases. The specific value of ϵ relies on the number of neighboring proto-curves in the ϵ-ball. In our implementation, the discretized value of ϵ is computed by the ball containing 0.1%, 0.3% and 0.5% amount of total proto-curves.

We ensure feature independence by pruning those redundant ϵ-balls with high relevance, i.e. having the same similarity and relative position. We thus calculate the mutual correlation between arbitrary two features following the theory of Pearsonian Correlation Coefficient in Statistics,

$$corr(h_i \mid h_j) = \frac{\sum_k r_i(\mathbf{c}_k) \cdot r_j(\mathbf{c}_k)}{\sum_k r_j(\mathbf{c}_k)}. \tag{6}$$

Note the correlation is non-symmetric measured. For example, if a feature h_i is totally covered by h_j, then $corr(h_i \mid h_j) < 1$ and $corr(h_j \mid h_i) = 1$.

3 Learning Shape Models via Information Projection

With a large amount of "ϵ-balls" as features, we introduce a novel learning algorithm based on information projection [4], embedded with a loop named "MaxMin-KL", to

pursue the generative shape models that implicitly form the quantized curve segments based deformable templates.

3.1 Learning Procedure

We pursue the generative shape model on a given training set $\{(\mathbf{S}_1, l_1), ..., (\mathbf{S}_N, l_N)\}$, where $l \in \{1, 0\}$ denotes the label of each sample.

Let $f(\mathbf{S})$ be the target distribution of a shape category. To learn a generative model with a few positive examples, we gradually pursue a series of models $p_1(\mathbf{S}), p_2(\mathbf{S}), ...,$ $p_t(\mathbf{S})$ to approach $f(\mathbf{S})$ from a background model $q(\mathbf{S})$ (reference samples) by stepwise selecting the most informative features, which lead to the fastest decreasing of the information gain. Since any shape $\mathbf{S}$ is projected into the feature spaces, we can redefine our problem that $p_t(r_1, r_2, ..., r_t)$ must agree upon dimensions $(r_1, r_2, ..., r_t)$ with the target distribution $f(r_1, r_2, ..., r_t)$, where r_t is the response of the selected feature h_t. Therefore, in each step, we choose a feature h_t to maximize KL divergence $\mathcal{KL}(p_t(r_t) \parallel p_{t-1}(r_t))$. Since the overlapping features have been roughly pruned, we may assume that $p_{t-1}(r_t) \approx q(r_t)$ (i.e. feature independence). With this independence assumption, the likelihood ratio of $f(r_t)$ and $q(r_t)$ can be factorized into individual likelihood ratios for the features. Thus, our shape model has the following form,

$$p_T(\mathbf{S}) = q(\mathbf{S}) \prod_{t=1}^{T} \frac{1}{Z_t} \exp\{\lambda_t r_t(\mathbf{S})\}, \tag{7}$$

where $Z_t = E_q[\exp\{\lambda_t r_t(\mathbf{S})\}]$ is the normalized term and each λ_t is found by $E_f[r_t] = E_{p_t}[r_t]$. And the following log-linear equation would provide a matching score against background for a given shape,

$$H(\mathbf{S}) = \log \frac{p_T(\mathbf{S})}{q(\mathbf{S})} = \sum_{t=1}^{T} (\lambda_t r_t(\mathbf{S}) - \log Z_t), \tag{8}$$

which can be combined with a threshold γ ($\gamma = 0$ in our implementation) for object classification.

We repeat two steps called "MaxMin-KL" for pursuing the shape model, that is selecting feature h_t and calculating the parameters λ_t and Z_t in Eq.(8). During the t-th pursuit iteration, we perform:

1) a max-step to argumentatively maximize the $\mathcal{KL}(p_t(r_t) \parallel q(r_t))$ for choosing a most distinct feature h_t.

Proposition I: *Let $f_i^{obs} = E_f[r_i]$ and $q_i^{ref} = E_q[r_i]$ be the expectations of any feature h_i responding to positives and reference samples respectively. We select a most distinct feature by maximizing KL-divergence in iteration t as*

$$h_t^* = \arg\max (f_t^{obs} - q_t^{ref})^2. \tag{9}$$

Proof: *Let* $r_t = E_{p_t}[r_t]$ *be a variable, we establish a function* $\Psi(r_t) = \mathcal{KL}(p_t(r_t) \parallel q(r_t)) = \lambda_t r_t - \log Z_t$. *Perform Taylor expansion of* $\Psi(r_t)$ *at point* $E_{p_{t-1}}[r_t]$,

$$\begin{aligned}\Psi(r_t) &\approx \underbrace{\Psi(E_{p_{t-1}}[r_t])}_{=0} + \underbrace{\frac{\partial \Psi(E_{p_{t-1}}[r_t])}{\partial r_t}}_{=0}(r_t - E_{p_{t-1}}[r_t]) \\ &+ \underbrace{\frac{\partial^2 \Psi(\frac{r_t + E_{p_{t-1}}[r_t]}{2})}{2\partial r_t^2}(r_t - E_{p_{t-1}}[r_t])^2 + \ldots}_{\approx (r_t - E_{p_{t-1}}[r_t])^2} \\ &\approx (r_t - E_{p_{t-1}}[r_t])^2 = (E_{p_t}[r_t] - E_q[r_t])^2 \end{aligned} \tag{10}$$

from which we can choose Eq.(9) as an approximation.

Intuitively, after selecting h_t, we can simply update f_i^{obs} and q_i^{ref} for each feature h_i as,

$$\begin{aligned} f_i^{obs} &= E_{p_t}[r_i] \cong \frac{1}{N^+}(1 - corr(h_t \mid h_i)) \sum_{j=1}^{N^+} r_i(\mathbf{S}_j), \\ q_i^{ref} &= E_q[r_i] \cong \frac{1}{N^-}(1 - corr(h_t \mid h_i)) \sum_{j=1}^{N^-} r_i(\mathbf{S}_j), \end{aligned} \tag{11}$$

where N^+, N^- stand for the number of positives and reference examples respectively and $corr(h_t \mid h_i)$ is defined in Eq.(6), which guides to perform the sparse feature set. In each iteration, the features that have their correlations with the selected feature h_t exceed a threshold δ will be directly excluded ($\delta = 0.2$ in our implementation).

2) a min-step to compute λ_t, Z_t for the selected h_t in order to meet the constraint $E_f[r_t] = E_{p_t}[r_t]$.

Proposition II: *Given the selected feature* h_t, *the parameters* λ_t *and* Z_t *of the current model is,*

$$\lambda_t = \log \frac{f_t^{obs}(1 - q_t^{ref})}{(1 - f_t^{obs})q_t^{ref}} \quad and \quad Z_t = e^{\lambda_t} q_t^{ref} + 1 - q_t^{ref}. \tag{12}$$

Proof: *As discussed above,* $Z_t = E_q[\exp\{\lambda_t r_t(\mathbf{S})\}] = \sum_{\mathcal{D}^{w_t}} q(r_t) \exp\{\lambda_t r_t(\mathbf{S})\}$, *which can be partitioned by* $\mathcal{D}^{w_t}$ *as*

$$\begin{aligned} Z_t &= \textstyle\sum_{\mathcal{D}^{w_t} < \epsilon_t} q(r_t) \exp\{\lambda_t\} + \sum_{\mathcal{D}^{w_t} \geq \epsilon_t} q(r_t) \\ &= e^{\lambda_t} E_q[r_t] + 1 - E_q[r_t] = e^{\lambda_t} q_t^{ref} + 1 - q_t^{ref}. \end{aligned} \tag{13}$$

Similarly, the analytical solution of λ_t *can be easily proved in the same way.*

The stepwise learning algorithm is summarized in Alg.1.

3.2 Shape Matching from Clutter Background

While the learned shape model in sec.3.1 consists a sparse feature set incorporated with different distance metrics, the corresponding proto-curves of features are directly used

Algorithm 1. Features pursuit

Input: A small set of positive shapes (i.e. we name it the "proto set") for extracting proto-curves and a training set, which contains a small number of positive samples and a certain amount of reference samples. The positive examples and the shapes of the proto set have no intersection, and have been normalized to the same scale.

Initialization: determining ϵ for each "ϵ-ball" to generate shape features by feature evolution (sec.2.2); computing correlation between each pair of features by Eq.(6).

Loop t=1 to T

max-step: select a distinct feature h_t^* by Eq.(9); update f_i^{obs}, q_i^{ref} for any feature h_i by Eq.(11) and prune correlated features with δ.

min-step: calculate λ_t, Z_t for h_t^* by Eq.(12).

⋆ *until information gain is smaller than a threshold (say 0.05) then stop.*

Output: A generative model (i.e. a strong classifier with a threshold $\gamma = 0$) for a shape category,
$H(\mathbf{S}) = \sum_{t=1}^{T}(\lambda_t r_t(\mathbf{S}) - \log Z_t)$.

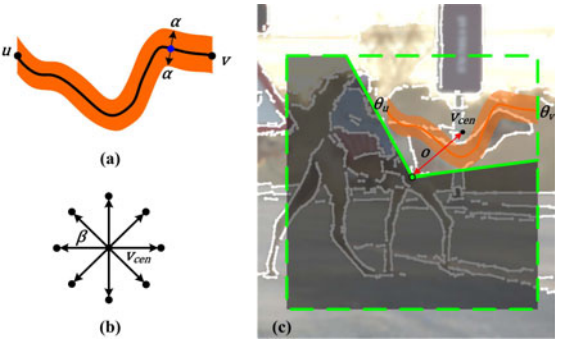

Fig. 4. (a) Shows a proto-curve as in the left of Fig.1 (a). We detect object contours from the edge map inside a ribbon of each proto-curve. (b) illustrates the idea of moving the ribbon around its eight neighborhood. (c) shows that each ribbon is used as the deformable template.

as deformable templates to match shape from clutter background in this section. We adopt a coarse-to-fine sliding window approach [5] and normalize the points inside each detection window to the same scale as the training step (Alg.1). Our goal is to sample curves and calculate response of the feature by Eq.(2). Since the object boundaries in clutter edge map are usually broken and surrounded by noise, it is natural to sample curves by scanning specific regions according to the spatial configuration of the proto-curve.

A template is defined as a four tuple $\{\mathbf{c}, v_{cen}, \alpha, \beta\}$, where v_{cen} indicates the center point of the proto-curve $\mathbf{c}$ and α, β are two radii acting on each landmark of $\mathbf{c}$ and v_{cen} respectively. Each template is working as follows. (i) We detect object boundary inside a ribbon of $\mathbf{c}$, that is obtaining by marching each landmark off its normal direction with a small distance α as illustrated in Fig.4 (a). (ii) A partial scan strategy is proposed to sample curves from clutter edge map. we first place this ribbon inside the detection window referring to the position of v_{cen}, which can be accurately calculated by its end

directions and center offset (Fig.4 (c)), and then move it around with a small radius β as shown in Fig.4 (b). (iii) Finally we compute the minimum distance of the sampled curves and the proto-curve using the related distance metric.

The above method is inspired by the part-based detection work [2], which uses single shape contour as template and gradients of each point as feature. The experimental results in sec.4 demonstrate that our promotion is successful. Combining with multiple metrics, the pursued shape model is robust and flexible to account for various deformation, occlusion and noise.

4 Experiments

We evaluate the proposed shape detectors with the following three experiments.

• **Experiment I. Shape detection from cluttered edge maps.** We select four classes (e.g. Bottles, Giraffes, Mugs and Swans) from the ETHZ image dataset [8] for this experiment. For each category, we partition the images into two half for training and testing respectively. Due to too few amount of this dataset, we use another 30 shapes LHI database [24] for extracting proto-curves. It worth mentioning that there are no any overlapped data between the two datasets. In this experiment, our method takes only about 2 minutes to learn four shape models.

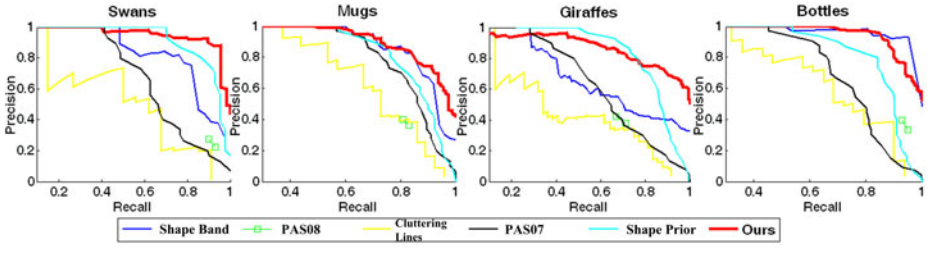

Fig. 5. Comparison of precision vs recall (PR) curves for four classes on ETHZ

Table 1. The precisions are compared to [11, 2, 19, 8] at the same recall rates

	Bottles	Giraffes	Swans	Mugs
Our precision/recall	83.9%/92.7%	83.4%/70.3%	**88.5%/93.9%**	**84.4%/83.4%**
Shape Prior CVPR09 [11]	39.6%/92.7%	**88.7%/70.3%**	60.4%/93.9%	69.9%/83.4%
Shape Band CVPR09 [2]	**95.0%/92.7%**	56.0%/70.3%	44.1%/93.9%	83.3%/83.4%
Cluttering Lines ICCV09 [19]	41.3%/92.7%	37.5%/70.3%	19.8%/93.9%	40.1%/83.4%
Ferrari 2008 [8]	33.3%/92.7%	43.9%/70.3%	23.3%/93.9%	40.9%/83.4%

Detecting results: the results of our algorithm are summarized in Fig.5. We compare with the results by [11, 2, 19, 8] using precision vs recall (PR) curves, where the advantage of our method is clearly identified. We also compared our approach to those

Fig. 6. Some selected results on ETHZ dataset [8] with the detected bounding boxes and contours plotted in green and yellow respectively

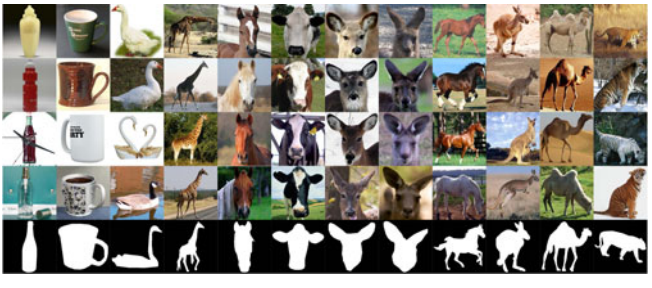

Fig. 7. Several data of the 40 image categories dataset chosen from LHI database [24] are illustrated. The last row shows some corresponding label maps.

methods at the same recall rates in Tab.1. Fig.6 shows some representative results on this dataset with the detected boxes (in green color) and localized curves by our system.

• **Experiment II. Shape-based categorization from images.** We further evaluate our method on a 40 categories image dataset selected from LHI database [24]. It contains about 3600 images and each with its corresponding edge map and label map. Our task is to detect and classify shapes from the edge maps. Due to the the heavy occlusion, shading, and surrounding clutter, this task is more complex. For example, the animal faces in Fig.7 are hardly distinguished. In the training stage, for each category, we separate the images into three equal parts: one for extracting proto-curves, one for learning models and the other for testing.

Results: we obtain an overall classification rate approaching 87.3%, which is better than the 81.4% reported in [15], and Fig.8 shows several selected results on this dataset.

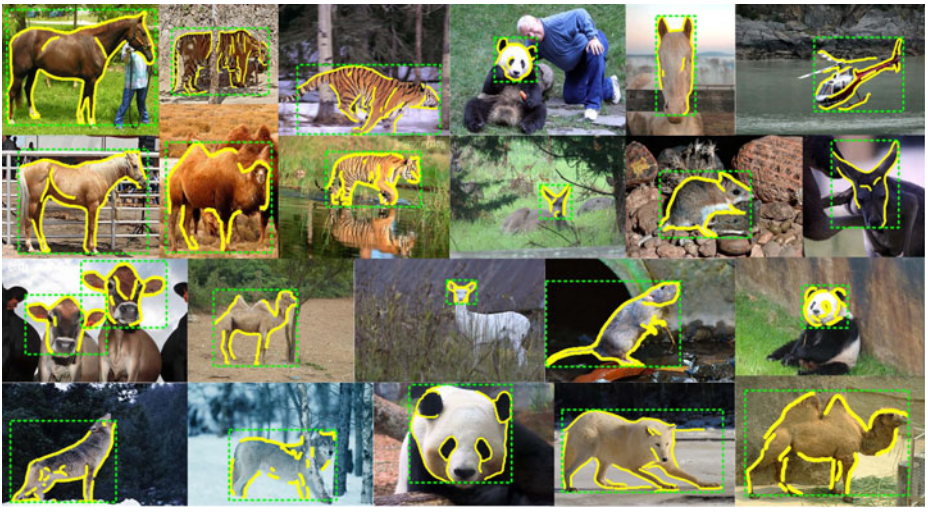

Fig. 8. We show some selected results on LHI dataset [24] and demonstrate that the shape detectors are robust to various deformations, background clutter and occlusion

Additional details of Experiment I and II are summarized here: the label maps used during the training stage are roughly annotated and aligned manually. Each of them is normalized to 256×256 with the aspect ratio preserved, and vectorized to 200 landmark points. For each shape in the proto set, we extract about one hundred prototype curves that represented by 30∼120 spaced landmarks. There are thus more than 10^4 features in total. The maximum iteration number T in Alg.1 is set to be 500. In the testing stage, we obtain edge maps of the images by canny detector and also normalize each detection window to 256 × 256. The radius α of each ribbon is set as 15 and the radius β is 10 (see sec.3.2).

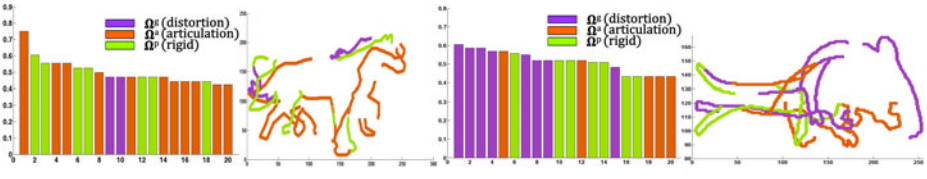

Fig. 9. Top 20 most informative features of horse (left) and mouse (right) are both plotted and visualized . Different colors indicate three distance metrics. From these results, we conclude that horses are more likely to perform articulation and mice are usually distorted, which matches our intuition very well. Moreover, we find that articulation mostly occurs on four limbs while distortion happens more often on the back and tail of animals. The shape models consisting of ϵ-balls can be viewed as the implicit deformable templates that includes different local shape variance.

• **Experiment III. Evaluating for feature selection.** It is an interesting experiment to reveal which types of features, corresponding to different deformation metrics, are

effective for different shape categories. We use two categories, horse and mouse, from the data in the Experiment II. As shown in Fig.9, top 20 informative features (i.e. first selected by our algorithm) are plotted and visualized respectively. Features with different distance metrics are denoted by different colors (green for procrustes metric Ω^p, orange for articulation metric Ω^a and purple for geodesic metric Ω^g). The results show that horses are more likely to perform articulation and mice are usually distorted, which matches our intuitive observation very well. Moreover, from these results we find that articulation mostly occurs on the limbs while distortion happens more often on the back and tail of animals. The shape models consisting of ϵ-balls can be viewed as the implicit deformable templates that includes different local shape variance.

5 Conclusion

In this paper, we learn shape models using local curve segments with multiple types of distance metrics. These shape models consisting of quantized curve segments can be viewed as the implicit deformable templates that incorporate different local shape variance. We show that our method significantly improves the shape classification and detection results on two public datasets. In the future work, we will implement a sophisticated design for further modeling distinct shape transformations and supporting wide range of shape descriptors more generally.

Acknowledgements. The work at LHI was supported by NSF China grants 90920009 and 60970156, and China 863 programs 2008AA01Z126 and 2009AA01Z331. The authors are thankful to Dr.Song-Chun Zhu and Dr. Yingnian Wu at UCLA for their insightful discussions and helpful comments.

References

1. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. TPAMI 24(4), 509–522 (2002)
2. Bai, X., et al.: Shape Band: A Deformable Object Detection Approach. In: CVPR, pp. 1335–1342 (2009)
3. Cootes, T.F., et al.: Active Shape Models - their training and application. CVIU 61, 38–59 (1995)
4. Csiszar, I., et al.: Information Theory and Statistics: A Tutorial. Commun. Inf. Theory 1(4), 417–528 (2004)
5. Dalal, N., et al.: Histograms of Oriented Gradients for Human Detection. In: CVPR, vol. 1(1), pp. 886–893 (2005)
6. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley and Son, Chichester (1998)
7. Felzenszwalb, P.F., Schwartz, J.D.: Hierarchical Matching of Deformable Shapes. In: CVPR (2007)
8. Ferrari, V., Jurie, F., Schmid, C.: From Images to Shape models for Object Detection. Intern'l Jour. of Computer Vision (2009)
9. Goodall, C.: Procrustes Methods in the Statistical Analysis of Shape. Jour. Royal Statistical Society 53, 285–339 (1991)

10. Gu, C., Lim, J.J., Arbelaez, P., Malik, J.: Recognition using Regions. In: CVPR (2009)
11. Jiang, T., Jurie, F., Schmid, C.: Learning Shape Prior Models for Object Matching. In: CVPR, pp. 848–855 (2009)
12. Klassen, E., Srivastava, A., Mio, W., Joshi, S.: Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces. IEEE Trans. PAMI 26(3), 372–383 (2004)
13. Ling, H., Jacobs, D.W.: Shape Classification Using the Inner-distance. TPAMI 29(2), 286–299 (2007)
14. Leibe, B., et al.: Combined Object Categorization and Segmentation With An Implicit Shape Model. In: ECCV Workshop, pp. 17–32 (2004)
15. Lin, L., et al.: An Empirical Study of Object Category Recognition: Sequential Testing with Generalized Samples. In: ICCV, vol. 1, pp. 419–426 (2007)
16. Lin, L., Wu, T., Xu, Z., Porway, J.: A Stochastic Graph Grammar for Compositional Object Representation and Recognition. Pattern Recognition 42(7), 1297–1307 (2009)
17. Lin, L., Liu, X., Zhu, S.C.: Layered Graph Matching with Composite Cluster Sampling. IEEE Trans. PAMI (2010)
18. McNeill, G., et al.: Hierarchical Procrustes Matching for Shape Retrieval. In: CVPR, vol. 1, pp. 885–894 (2006)
19. Ommer, B., et al.: Multi-Scale Object Detection by Clustering Lines. In: ICCV (2009)
20. Pietra, V.D., et al.: Inducing Features of Random Fields. TPAMI 19, 380–393 (1997)
21. Siddiqi, K., et al.: Shock Graphs and Shape Matching. IJCV 35(1), 13–32 (1999)
22. Shotton, J., Blake, A., Cipolla, R.: Multi-Scale Categorical Object Recognition Using Contour Fragments. IEEE Tran. PAMI (2008)
23. Todorovic, S., Ahuja, N.: Unsupervised Category Modeling, Recognition, and Segmentation in Images. IEEE Tran. PAMI (2008)
24. Yao, B., Yang, X., Lin, L., Lee, M.W., Zhu, S.C.: I2T: Image Parsing to Text Description. Proceedings of IEEE (2010) (to appear)
25. Yu, X., Yi, L., Fermuller, C., Doermann, D.: Object Detection Using Shape Codebook. In: BMVC (2007)
26. Zhu, Q., et al.: Contour Context Selection for Object Detection: A set-to-set contour matching approach. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 774–787. Springer, Heidelberg (2008)