IEEE Trans on PAMI, Vol. 21, No.11, November 1999.

Embedding Gestalt Laws in Markov Random Fields

– a theory for shape modeling and perceptual organization

Song-Chun Zhu

Dept. of Computer and Information Science
Ohio State University
szhu@cis.ohio-state.edu
http://www.cis.ohio-state.edu/~szhu/

Abstract

The goal of this paper is to study a mathematical framework of 2D object shape modeling and learning for middle level vision problems, such as image segmentation and perceptual organization. For this purpose, we pursue generic shape models which characterize the most common features of 2D object shapes. In this paper, shape models are learned from observed natural shapes based on a minimax entropy learning theory (Zhu and Mumford 1997, Zhu, Wu and Mumford 1997)[31, 32]. The learned shape models are Gibbs distributions defined on Markov random fields (MRFs). The neighborhood structures of these MRFs correspond to Gestalt laws –co-linearity, co-circularity, proximity, parallelism, and symmetry. Thus both contour-based and region-based features are accounted for. Stochastic Markov chain Monte Carlo (MCMC) algorithms are proposed for learning and model verification. Furthermore, this paper provides a quantitative measure for the so-called *non-accidental statistics*, and thus justifies some empirical observations of Gestalt psychology by information theory. Our experiments also demonstrate that global shape properties can arise from interactions of local features.

Key words: Gestalt laws, perceptual grouping, shape modeling, Markov random field, maximum entropy, shape synthesis, active contour.

1 Introduction and motivations

In psychology, it has long been evident that early human vision strongly favors certain shapes and configurations over others without high level identification. Many theories have been proposed to account for this phenomenon, among which Gestalt psychology is the most influential one. In Gestalt psychology, an enormous number of generic laws have been identified for grouping *parts* to *whole* (Koffka 1935)[17], for example,

> proximity, continuity, co-linearity, co-circularity, parallelism, symmetry, closure, familiarity.

These laws are supposed to be coordinated by the law of Pragnanz[17]:

> "of several geometrically possible organizations that one will actually occur which possesses the best, simplest and most stable shape (p.138)."

But what is meant by a good, simple and stable shape? Gestalt psychologists explained it in terms of "field forces" by analogy to field theories of gravity and electricity, but it is unclear what the "field forces" really are. Even worse, Gestalt psychology only provides a descriptive theory, it does not specify a computational process for achieving the percept from *parts* to *whole*. Although Gestalt laws have been successfully utilized in many perceptual organization algorithms[22, 24], a rigorous mathematical theory has yet to be found.

Besides Gestalt psychology, there are two other theories for perceptual organization. One is the *likelihood principle* (Helmholtz 1867)[9], which assigns a high probability for grouping two elements, such as line segments, if the placement of the two elements has a low likelihood of resulting from *accidental arrangement* (Lowe 1985,1990)[21, 22]. The third theory is the simplicity or *minimum description length principle* (Hochberg 1957) [10], which states that perceptual organization should achieve the shortest coding length by exploring shape symmetry etc.

We argue that the key to understanding the psychophysical phenomena is to pursue the "true" underlying probability distribution for the ensemble of object shapes (or configurations) in a given application domain. With this probability distribution, the three theories can be well unified for the following reasons.

1. According to Shannon's coding theory, a "true" probability yields the minimum expected coding length for the ensemble of shapes (or configurations)[1]. Since the shape distribution accounts for the complexity and frequency of configurations occurring in nature, it gives the genuine likelihood probability for groupings. As we can see in later section, *non-accidental* statistics can also be measured using this distribution.

2. The importance of Gestalt laws can be quantified by studying statistical regularities in the shape ensemble. These Gestalt laws should be reflected in the structures of the probability distribution, provided that they are indeed effective.

In two previous papers, the author, in collaboration with Wu and Mumford, has studied probability models for textures and natural images by a minimax entropy principle[32, 31]. This paper applies the minimax entropy principle to learning probability distributions of 2D shapes. In particular, we are interested in simple closed curves which are non-self-intersect region boundaries in 2D images projected from 3D objects. A basic assumption is that there exists a true underlying distribution for 2D object contours in a given application, and shape modeling is posed as a statistical inference problem.

We shall address the following questions in this paper:

1. Studies in neurosciences have showed that statistics of visual environment (i.e. the ecologic effects) play key roles on the functions of both individual cells and neural systems[1].[2] Are there any ecologic evidences — embedded in the statistics of realistic shapes — for Gestalt laws and early human perception of 2D shapes?

2. How can various Gestalt laws be integrated into a single probability measure?

3. What are the structures and forms of generic probability distributions for 2D shapes?

[1] The coding length should also include the code book, thus simpler models are favored in model selection.

[2] For example, there are considerable functional differences of neurons in retina between rabbit and fish, and these differences are supposed to be explained by statistical properties of their living environments[1]. Here, we are interested in knowing why early human vision is sensitive to Gestalt features.

4. How do we sample shape distributions? What are the typical shapes sampled from such distributions? To what extent can real shapes–with global properties – be characterized by merely a few locally defined Gestalt laws?

We start with exploring statistics of features extracted from observed real shapes, and then a shape distribution is learned such that it reproduces the observed statistics while having the maximum entropy. The learned shape models are Gibbs distributions on Markov random fields. The variables of these MRFs are the coordinates of points along the contours, and the structures of the MRFs correspond to Gestalt laws such as proximity, co-linearity, co-circularity, parallelism and symmetry. The learned shape models are verified by Markov chain Monte Carlo sampling. The paper also provides a quantitative measure for *non-accidental statistics* by comparing observed statistics and statistics of randomly sampled shapes. We demonstrate that global shape properties could arise through propagation of local interactions in Markov random fields.

The paper is organized as follows. We start with reviewing existing theories for shape modeling in section (2), and discuss basic issues in shape modeling in section (3). Section (4) discusses feature extraction, and section (5) demonstrates experiments on statistics of animate shapes. Section (6) presents a maximum entropy theory for probability learning. Section (7) discusses issues of designing stochastic sampling processes in a shape space, and describes experiments of sampling a uniform distribution of shape. Section (8) discusses feature selection by maximizing non-accidental statistics. Section (9) is devoted to experiments on shape learning and sampling. Finally we conclude the paper by a discussion in section (10).

2 Previous theories in 2D shape modeling and perceptual grouping

2D shapes have been studied intensively in many disciplines, ranging from pure mathematics, statistics, psychology, to computer vision. These studies are divided into four areas: 1). a statistical theory of shape, 2). deformable templates and models, 3). non-accidental properties and perceptual grouping, 4). active contour models. In this section, we shall review these theories.

2.1 Statistical theory of shape

The term "theory of shape" was coined by David Kendall in 1977 (see [15]). In Kendall's theory, a shape is defined as a set of k points in m dimensions. Thus a shape is naturally represented as an $m \times k$ matrix. To "filter" out effects from translation, scaling, and rotation, Kendall first defines one point as the origin, which reduces the matrix to a size of $m \times (k-1)$, and then he normalizes the shape by setting the sum of the squared elements of the matrix to one. Thus all his shapes live on the surface of a unit sphere in $m \times (k-1)$ dimensions, which he called the *preshape sphere*. The real shape space, denoted by $\sum_m^k$, is the quotient of the preshape sphere by some transformation groups, e.g. $SO(m)$. Then each $SO(m)$ equivalence class is viewed as a single point–a real shape[15]. For example, assuming independent and identical uniform distribution for the k points, Kendall makes inference about the statistical significance of co-linearity in a given point set, and of the size of a hole in a Delaunay tessellation of Galaxy[15], to answer query about how likely the co-linearities and holes are not accidental arrangements by i.i.d. uniform distributions.

The theory of shape has been also studied by Bookstein in morphometrics – a discipline studying deformations and variabilities of biological organisms[2]. Here the point set is the collection of unique landmarks which correspond biologically from object to object. Recently Mardia and Dryden have studied the theory of shape in the context of image analysis with interesting results[23].

2.2 Deformable templates – high level vision

Another elegant theory for shape modeling was pioneered by Grenander in a discipline which he named *pattern theory*[7]. In pattern theory, shape primitives are selected, such as edge segments, and these primitives are arranged in pre-specified configurations, such as a circular graph. Then transformation groups (rotation, scaling etc.) act on these configurations to account for both global and local deformations.

For example, the contour of a human hand is represented as a ring of linelets[8], and a global transform specifies the location, size, and orientation of the hand, and local transforms define relative orientations of fingers, and so on. Then a probability distribution is defined on these groups. This shape modeling scheme has been used in representing leaves and brain mapping [8, 7].

In computer vision, deformable templates were proposed by Yuille in studying shapes of human eyes, mouth, and eyebrows etc.[28]. In Yuille's templates, shapes consist of piecewise conics, and deformations are modeled based on coefficients. Other deformable models of shapes are defined in terms of basis functions, such as B-splines, sine waves[4], and implicit polynomials[16] and super quadrics[26]. A more general template model for flexible object is studied in the FORMS system (Zhu and Yuille 1996)[30], where object shapes are defined in graphs computed from medial axes, and deformations are specified by PCA modes learned from animal shapes. Other shape models are represented in Bayes networks[5]. In fact, the statistical theory of shape studied in (Bookstein 1986, Mardia and Dryden 1989) can also be considered as shape templates with points being primitives.

2.3 Non-accidental properties and perceptual organization – middle level vision

The problem of perceptual organization was formally studied by Lowe in 1985[21]. Lowe pointed out that the goal of grouping is to identify features that are likely to have arisen from some scene properties rather than accidental arrangements. Lowe proposed some measures that account for how non-accidental the arrangement is for each individual grouping feature, such as, co-linearity and parallelism. This theory has been used in grouping and recognizing rigid objects with interesting results[22, 24]. However it doesn't provide a rigorous probability measure for shapes, and it is unclear how to **de-correlate** multiple shape features in computing the significance of a non-accidental arrangement. Similar to Lowe's theory, non-accidental properties are also studied for convexity of a sequence of line segments by Jacobs in 1992[11].

In the literature of perceptual grouping, there are no rigorous attempts for learning and verifying probabilities models from real shapes.

2.4 Active contour models – low level vision

In early vision tasks, generic shape models are found to be useful. One important generic shape model is the active contour model (SNAKE)[14], and the internal energy in a SNAKE model implicitly defines a prior probability distribution of a curve $\Gamma(s)$:

$$p(\Gamma) = \frac{1}{Z} \exp\{-\int \alpha|\dot{\Gamma}(s)|^2 + \beta|\ddot{\Gamma}(s)|^2 ds\} \tag{1}$$

where Z is a normalization constant, $\dot{\Gamma}(s)$ and $\ddot{\Gamma}(s)$ are the first and the second derivatives of the curve and s is usually the arc-length.

An explicit probability model for open curves, called the Elastica model, was first derived from Brownian motion by Mumford in 1994[25] and it was also studied by Williams and Jacobs in 1997[27]:

$$p(\Gamma) = \frac{1}{Z}\exp\{-\int \beta + \alpha\kappa^2(s)ds\} \tag{2}$$

where $\kappa(s)$ is the curvature of the curve.

In discrete cases, both equation (1) and equation (2) are defined on interactions between nodes in a local neighborhood, illustrated in figure 1.a. Despite the wide applications of active contour models in computer vision, they have two main problems.

Firstly, they do not characterize region based information. For example, in the shape shown in figure 1.b, node A is close to node B in distance, whereas they are far away from each other along the boundary.

Secondly, potential functions are often manually designed, whereas it is desirable to learn them from data.

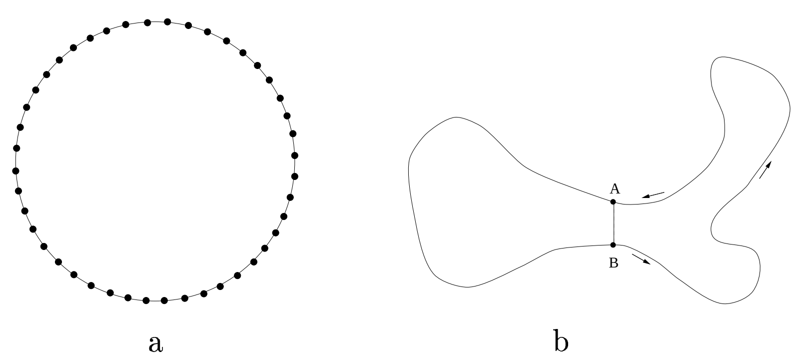


Figure 1 a). an 1D Markov random field where the nodes represent random variables for positions of contour points. b). node A is spatially adjacent to point B, but it is far away from B in the circular neighborhood of a).

In summary, subsections (2.4), (2.3), and (2.2) briefly review existing theories in the low, middle, and high level vision respectively. It is still unclear how these theories can be combined for image understanding in a consistent manner, and the incompatibility of these models (or theories) in the three levels poses a major obstacle for building sophisticated vision systems.

In this paper, we shall study a mathematical framework of 2D shape modeling for middle level vision problems, such as image segmentation and perceptual organization. Motivated by the above discussion, we pose two criteria in our models. I) These shape models should be generic, and thus characterize the most common features of 2D shapes. II) They should be compatible with low level representations, such as raw images and edge maps, and the high level descriptions such as deformable templates.

3 Shape spaces and probability measures

In this section, we discuss the basic issues in shape modeling: the spaces of 2D shapes and the probability measures defined on the shape spaces.

Let $\Gamma(s)$ be a simple and closed contour on 2D plane[3], where $s \in [0,1]$ is the arc-length and $\Gamma(0) = \Gamma(1)$. Fixing the resolution, $\Gamma(s)$ is discretized into a polygon of N vertices,

$$\Gamma = ((x_0, y_0), (x_1, y_1), ..., (x_{N-1}, y_{N-1})), \quad \Gamma \in \Omega_\Gamma \subset R^{2N},$$

where Ω_Γ is the space of shapes satisfying the following hard constraints.

Constraint I: $\Gamma \in \Omega_\Gamma$ is closed and non-self-intersecting.
Constraint II: $\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \in [ds - \epsilon, ds + \epsilon] \; i = 0, 1, ..., N-1.$

In constraint II, the distance of sequential nodes is allowed to vary slightly because of finite precision of lattice in computer implementation. We shall discuss the effect of ϵ when we discuss a Gibbs sampler for sampling shapes in section (7). The structures of Ω_Γ perhaps are too complicated to analyze explicitly.

On Ω_Γ a Lebesgue measure is well defined:

$$P(\Gamma) = p(\Gamma) dx_0 dy_0 dx_1 dy_1 \cdots dx_{N-1} dy_{N-1}, \tag{3}$$

with $p(\Gamma)$ being the density that we shall define in late sections.

Following Kendall's treatment[15], we shall "filter out" some effects of transforms and define a "genuine" space of shape Ω^*_Γ. Ω^*_Γ is the quotient space of Ω_Γ under 2D translation and planar rotation as well as cyclic permutation of nodes. Each

[3]In the rest of the paper, the word shape is referred to a simple, closed and non-self-intersecting contour.

"genuine" shape in Ω^*_Γ is the projection of an *equivalence class* in Ω_Γ under similarity transforms and cyclic permutation. The reasons for such treatment are two folds.

Firstly, object silhouettes are taken into images at arbitrary distances, locations and orientations. Therefore, a shape model should be invariant to translation, rotation and scale.

Secondly, for generic shape models in middle level vision, we assume that any features should have the same chance to appear in any location s in the contour. Thus the shape model is invariant to cyclic permutation of nodes.

In the rest of the paper, $p(\Gamma)$ is assumed to be **homogeneous** with respect to s, and $p(\Gamma)$ is defined on relative positions and orientations. Furthermore, Γ is normalized to have unit length. Thus the density $p(\Gamma)$ is invariant to translation, rotation, scale, and cyclic permutation. As we shall discuss in section (10), non-homogeneous properties can be built at high level representation.

Now we need to define a probability measure $\nu(\Gamma)$ on Ω^*_Γ, which takes the sum of the Lebesgue measure over an equivalence class in Ω_Γ. If we only consider translation and cyclic permutation transforms, the equivalence classes all have the same size, therefore, $\nu(\Gamma)$ is simply the Lebesgue measure multiplied by a constant[4]. When rotation is added, the sizes of the equivalence classes may vary, because of rotational symmetry in shapes. However, this should not be a concern in computer vision, because of the following reason.

Suppose Γ is a rotationally symmetric polygon, for example, figure 2 displays a N-gon with degree $d = 4$ rotational symmetry. Suppose we discretized $[0, 2\pi]$ into 1000 angles, the equivalence class for this shape includes only 997 distinct shapes, while a non-symmetric shape has 1000 distinct shapes in its equivalence class. With precise discretization, the size variation of equivalence classes is negligible except for shapes,

e.g. a perfect circle. However, because we are studying natural animate shapes in this paper, the probability measure for shapes which have high degrees of rotational symmetry is **exponentially** small, that is, these highly symmetric shapes have measure 0. For example, none of the observed and synthesized shapes in the later section demonstrate any rotational symmetry at all!

In a given application domain, the ensemble of shapes is assumed to be governed

[4]With a slight modification, we can define shapes in a finite but large enough 2D rectangular domain instead of an infinite plane, then the equivalence classes have finite size.

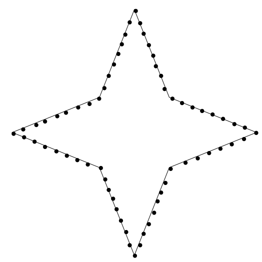


Figure 2 A N-side polygon with rotational symmetry of degree $d = 4$.

by a true underlying distribution $f(\Gamma)$ on Ω_Γ (or Ω_Γ^* with no practical difference). A set of 2D object contours, $\{\Gamma_i^{\text{obs}},\ i = 1, 2, ..., M\}$, are observed as independent samples from $f(\Gamma)$ (see examples in figure (4)). The objective of shape modeling is to learn a probability model $p(\Gamma)$ as an estimation of $f(\Gamma)$. In computer vision, $p(\Gamma)$ is often called the *prior model*, and can be used for image segmentation, contour tracing, shape completion, and so on in the Bayesian framework.

4 Feature selection

To learn a shape model $p(\Gamma)$, we start with exploring both contour-based and region-based features on Γ. We are interested in some simple Gestalt laws: co-linearity, co-circularity, proximity, parallelism, and symmetry, each of which is measured by a continuous function.

Figure 3.a shows a dog shape Γ, which is decomposed into a number of sequential linelets $\ell(s)$. $\ell(s)$ includes attributes $(x(s), y(s), \theta(s))$ for the center coordinates and orientation. Each linelet is represented by a node on a circle in figure 3.b. The circular connection defines a *random field* as in figure 1.a, where $x(s), y(s), \theta(s)$ are random variables at each site s and the neighborhood structures of the random field is shown in figure 3.b (to be discussed later in this section).

Now we discuss how shape features are extracted from random fields.

At each location s,[5] we measure a set of functions $\phi^{(\alpha)}(s)$ with $\alpha = 1, 2, 3, ...$ being the index of shape features. By analogy to feature extraction using Gabor filters in 2D images, $\phi^{(\alpha)}(s)$ is the "response" for a "shape filter" at location s.

We first define two contour-based functions, the curvature $\kappa(s)$ and derivative

[5]For simplicity of notation, we derive the shape models in continuous representation, and then we convert it to the discrete lattice at other places.

of curvature:

$$\phi^{(1)}(s) = \kappa(s) = \frac{d\theta(s)}{ds}, \quad \text{and} \quad \phi^{(2)}(s) = \nabla\kappa(s) = \frac{d^2\theta(s)}{ds^2}, \quad \forall s \in [0,1].$$

Obviously, $\phi^{(1)}(s) = 0$ means that two adjacent linelets are *co-linear*, and $\phi^{(2)}(s) = 0$ means that three sequential linelets are *co-circular*. Other contour-based shape filters can be defined in the same way.

In the following, we proceed to define region based properties.

It is well known in computer vision that real object shapes –resulting from processes of accretion – have natural descriptions in terms of medial axes (Leyton 1992, Zhu and Yuille 1996) [20, 30]. For example, a dog has a skeleton and many elongated parts. As shown in figure 3.a, linelets at the two sides of a limb or a torso are parallel or symmetric with respect to a medial axis. It is also evident in psychophysical experiments that early human vision is sensitive to medial axis (Kovacs and Julesz 1994)[18, 3]. This observation has its deep root in physiology where experiments have demonstrated that some neurons in the primary visual cortex (V1) of monkeys compute symmetric axes as soon as they detect edge elements (Lee, Mumford and Zhu 1996)[19].

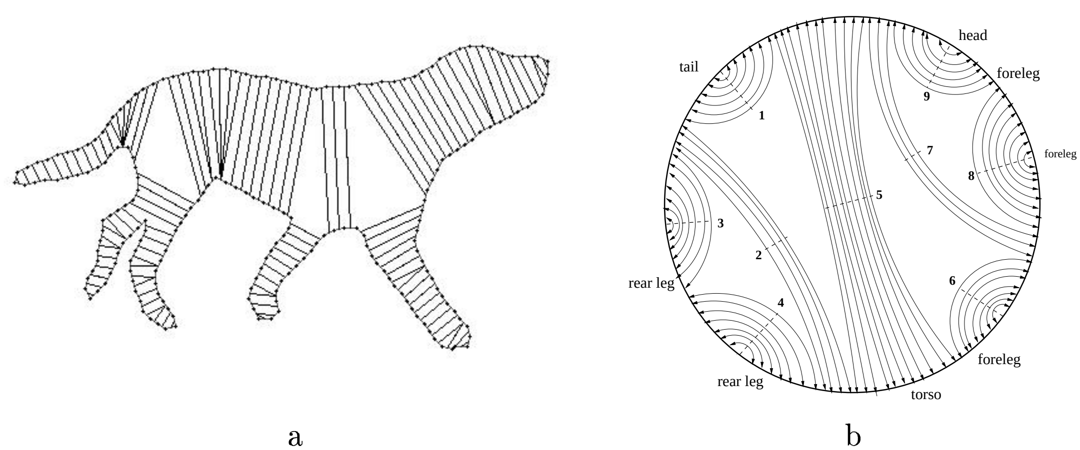


Figure 3 a. A dog shape with region based correspondence detected, b. an abstract planar *adjacency graph* representing the neighborhood structures of the linelets in the dog shape.

To capture region based features of a curve $\Gamma(s)$, the author has defined a symmetry mapping function $\psi(s)$ (Zhu 1998)[33],

$$\psi : [0,1] \to [0,1], \qquad \psi(s) = t \Longleftrightarrow \psi(t) = s.$$

Figure 3 displays the mapping function ψ for a dog shape. As shown in figure 3.b, ψ divides the circular domain $[0,1]$ into 18 disjoint intervals

$$\mathcal{I}_i = (s_{i0}, s_{i1}) \subset [0,1], \quad \mathcal{I}_i \cap \mathcal{I}_j = \emptyset, i \neq j.$$

ψ establishes a piecewisely continuous mapping $(s, \psi(s))$ between these intervals, under a hard constraint that these mapping line segments don't cross each other.

Intuitively each pair of intervals represent an elongated part of a shape. For example, the mapping function ψ for the dog shape, shown in figure 3, has 9 pairs of intervals for the 9 parts of the dog. Note that in figure 3.a some linelets are matched more than once due to the effect of discretization. As illustrated in figure 3.b, the mapping function ψ defines a new neighborhood for nodes on shape Γ, and opens "communication channels" for linelets across regions.

Now we briefly explain how ψ is computed.

The mapping function $\psi(s)$, as well as intervals, is computed by minimizing an energy functional, which is defined to enforce the following two aspects.

1. Two matched linelets $\ell(s)$ and $\ell(\psi(s))$ should be as close, parallel, and symmetric to each other as possible.

2. The number of intervals (or discontinuities of $\psi(s)$) in $[0,1]$ should be as small as possible.

The optimal ψ is computed by a stochastic algorithm (Gibbs sampler) based on local neighborhood in a Markov random field, and the medial axis is then computed based on ψ. Details of the definition of the energy functional and the algorithm is referred to a companion paper[33].

Let $r(s)$ be the distance between two matched linelets $\ell(s)$ and $\ell(\psi(s))$. $r(s)$ is divided by the length of the curve, so it is well normalized with respect to scale. We call $r(s)$ the "rib" length. Then we define three region based shape features,

$$\phi^{(3)}(s) = r(s), \quad \phi^{(4)}(s) = \nabla r(s) = \frac{d\,r(s)}{ds}, \quad \phi^{(5)}(s) = \nabla^2 r(s) = \frac{d^2 r(s)}{ds^2}.$$

$\phi^{(4)}(s) = 0$ means two linelets are parallel to each other, $\phi^{(5)}(s) = 0$ means two pairs of linelets are symmetric to each other. $\phi^{(3)}(s)$ measures the proximity between two linelets across a region as we discussed in figure (1).b.

It is easy to see that all functions $\psi^{(\alpha)}(s), \alpha = 1, 2, 3, 4, 5$ are continuous measures for the shape properties and the Gestalt laws, and they are invariant to translation, rotation, and scaling transforms in a 2D plane.

Another interesting shape feature is,

$$\phi^{(6)}(s) = \begin{cases} 1 & \text{if } \psi(s) \text{ is discontinuous at s} \\ 0 & \text{otherwise} \end{cases}$$

This measures the number of breaks and branches (or parts) of an object.

5 Statistics of animate shapes

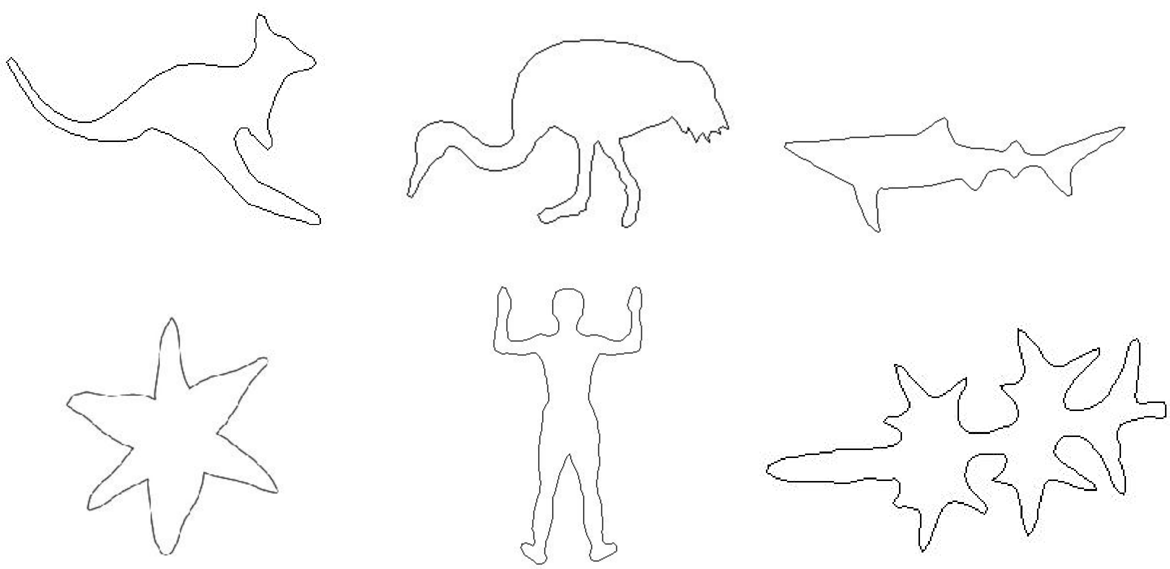


Figure 4 Examples of the observed natural shapes.

In this section we shall discuss how to compute statistics of natural 2D shapes for a set of features $\Phi = \{\phi^{(\alpha)},\ \alpha = 1, 2, ...\},$.

We collect a set of $M = 22$ shapes of animate objects, $\{\Gamma_i^{\text{obs}}, i = 1, 2, ..., 22\}$, six of which are displayed in figure 4. These shapes are assumed to be independent samples from a true distribution $f(\Gamma)$. We are interested in animate shapes because they have richer flexibilities and variations than human made objects.

These animate shapes are acquired at various resolutions from 2D images with perimeter being L_i pixels long for $i = 1, 2, ..., 22$. A shape observed at a high resolution contains rich information. Γ_i^{obs} is represented by $N_i = \frac{L_i}{c}$ nodes or linelets, i.e. a polygon of N_i sides. Generally speaking, the side length c of a polygon should be as small as possible to obtain a good approximation to the continuous curve, on the other hand, c should not be too small, otherwise features, such as $\kappa(s)$, cannot be computed reliably. A good compromise is to make c as big as possible while there is no noticeable artifacts of polygon approximation to human perception. We choose $c = 6$ pixels in this paper. The arc length for each linelet in Γ_i^{obs} is $ds = \frac{1}{N_i}$, therefore all shapes are normalized to the same scale.

In this paper, the statistics are extracted as empirical histograms of features. By homogeneity assumption, the histogram for $\phi^{(\alpha)}(s)$ on Γ_i^{obs} is,

$$H^{(\alpha)}(\Gamma_i^{\text{obs}}, z) = \int \delta(z - \phi^{(\alpha)}(s))ds, \quad \alpha = 1, 2, ..., 6, \quad \forall s.$$

In the above definition, z is a continuous variable for the feature, e.g. $H^{(0)}(\Gamma_i^{\text{obs}}, 0)$ is the number of points on Γ_i^{obs} that have zero curvature. $\delta()$ is the Dirac delta function with unit mass at zero and $\delta(x) = 0$ for $x \neq 0$. For a discretized curve, we have

$$H^{(\alpha)}(\Gamma_i^{\text{obs}}, z) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta(z - \phi^{(\alpha)}(s_j)), \quad \alpha = 1, 2, ..., 6.$$

We further compute average histograms for the M discretized curves

$$\mu_{\text{obs}}^{(\alpha)}(z) = \frac{1}{N_1 + N_2 + ... + N_M} \sum_{i=1}^{M} N_i H^{(\alpha)}(\Gamma_i^{\text{obs}}, z), \quad \alpha = 1, 2, ..., 6. \tag{4}$$

If $\sum_{i=1}^{M} N_i$ is big enough, then $\mu_{\text{obs}}^{(\alpha)}(z)$ is a close estimation of the marginal distribution of the true model $f(\Gamma)$, i.e.

$$\mu_{\text{obs}}^{(\alpha)}(z) \approx \mu^{(\alpha)}(z) = \int f(\Gamma)\delta(z - \phi^{(\alpha)}(s))d\Gamma, \quad \forall z, \ \forall s, \quad \forall \alpha.$$

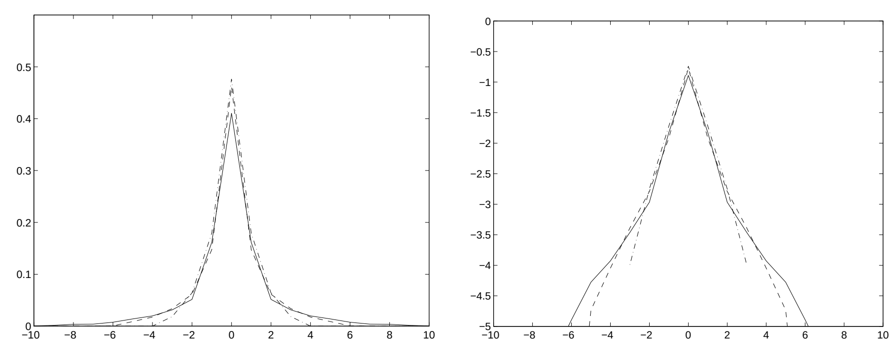
In the rest of the paper, we assume that $\mu_{\text{obs}}^{(\alpha)}(z) = \mu^{(\alpha)}(z), \alpha = 1, 2, ..., 6$.

To study how the observed histograms change with the scales (resolutions) of natural shapes, we subsample each observed shape once and twice, and generate two new sets of observed shapes at scale 1 and scale 2. A shape at scale $i + 1$ has only one half of the nodes of the shape at scale i.

Figure 5.a plots the observed histograms $\mu_{\text{obs}}^{(1)}$ for $\phi^{(1)}$ averaged over 22 animate shapes at scale 0 (solid curve), scale 1 (dashed curve) and scale 2 (dash-dotted curve) respectively, and figure 5.b plots the logarithms of these curves in figure 5.a.

In our experiments, we flipped the 22 observed shapes horizontally to double the observed dataset to 44 shapes for more robust estimation of histograms, because we should have the same chance to observe an animate object from both sides. As a result, the histograms for $\phi^{(\alpha)}, \alpha = 1, 2$ are perfectly symmetric with a peak at zero. This result may not be true if we had not had flipped the shapes. This treatment means that the probability model should also be invariant to a flip transform.

Figure (5) demonstrates two interesting properties. Firstly, the middle part of $\mu^{(1)}_{\text{obs}}$ is close to an exponential curve, but $\mu^{(1)}_{\text{obs}}$ has heavier tails. Secondly, unlike the scale invariant properties found in filter responses of natural images[31], the histogram of curvature is only approximately invariant to scales in these observed shapes. This is mainly because that the sub-sampling procedure smoothes the curve faster at high curvature segments than at low curvature segments. As a result, the curvature histograms at scale $i+1$ has lighter tails and sharper peak near zero than histograms at scale i.



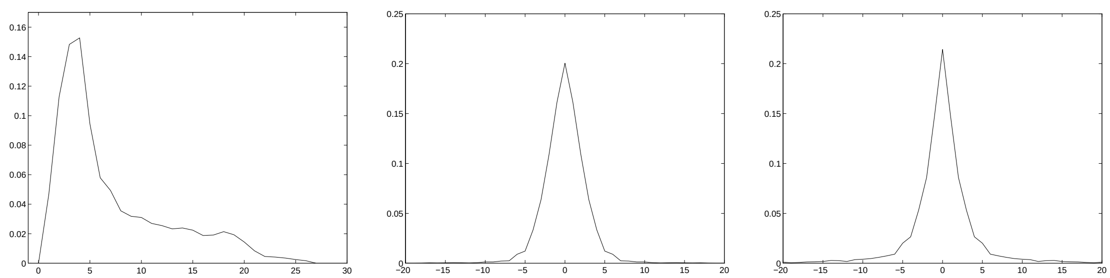
a b

Figure 5 a). the histograms of $\kappa(s)$ averaged over 22 animate objects at scale 0 (solid curve), scale 1 (dashed curve) and scale 2 (dash-dotted curve), the horizontal axis is $\kappa(s)$ with unit $dz = \frac{\pi}{16\times 200}$ b). the logarithm of curves in a).

The average histogram $\mu^{(2)}_{\text{obs}}$ of $\phi^{(2)}$ is very close to the curvature histogram. Currently it is unclear why it is so. We choose $\mu^{(1)}_{\text{obs}} = \mu^{(2)}_{\text{obs}}$ to be the observed curvature histogram averaged over scale 0 and scale 1 for robustness.

We also compute histograms $\mu^{(\alpha)}_{\text{obs}}, \alpha = 3, 4, 5$ for the region based features, $\phi^{(3)}(s) = r(s), \phi^{(4)}(s) = \nabla r(s), \phi^{(5)}(s) = \nabla^2 r(s)$. These computed histograms are shown in figure 6.

Histograms for $\kappa(s), \nabla\kappa(s), \nabla r(s), \nabla^2 r(s)$ are not Gaussian distributions, and they all have a sharp peak at zero, indicating the statistical significance of co-linearity, co-circularity, parallelism and symmetry in natural shapes. The histogram of $r(s)$ has peak near zero, which implies important correlations and proximity of points across regions.



a b c

Figure 6 The observed histograms averaged over 22 animate shapes for region based features a). histogram of $r(s)$, b). histogram of $\nabla r(s)$, c). histogram of $\nabla^2 r(s)$.

6 Shape modeling by maximum entropy

This section shall derive shape models which account for the observed statistics.

6.1 *From empirical histograms to Gibbs distributions*

Given features $\Phi = \{\phi^{(\alpha)}, \alpha = 1, 2, ...\}$ and the marginal distributions of $f(\Gamma)$ $\{\mu^{(\alpha)}_{\text{obs}}(z), \alpha = 1, 2, ...\}$, a model $p(\Gamma)$ should reproduce these marginal distributions. Thus as long as the features in Φ are concerned, $p(\Gamma)$ cannot be distinguished from the true model $f(\Gamma)$. Intuitively if we extract enough number of important features, $p(\Gamma)$ will be a close estimation of $f(\Gamma)$.

Let Ω_p be the set of all probability distributions $p(\Gamma)$ which can reproduce the observed statistics,

$$\Omega_p = \{p(\Gamma) \mid \int p(\Gamma)\delta(z - \phi^{(\alpha)}(s))d\Gamma = \mu^{(\alpha)}_{\text{obs}}(z) \quad \forall s, \forall z, \forall \alpha\}.$$

From Ω_p, a maximum entropy (ME) (Jaynes 1957)[12] distribution is chosen, for it has the least bias in the unconstrained dimensions. Thus a shape model is

$$p^*(\Gamma) = \arg\max - \int p(\Gamma) \log p(\Gamma) d\Gamma$$

subject to constraints,

$$\int p(\Gamma)d\Gamma = 1, \tag{5}$$

$$\int p(\Gamma)\delta(z - \phi^{(\alpha)}(s))d\Gamma = \mu^{(\alpha)}_{\mathrm{obs}}(z) \quad \forall s, \forall z, \forall \alpha. \tag{6}$$

Solving the constrained optimization problem by Lagrange multipliers and calculus of variations, we obtain the following ME distribution,

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{-\sum_{\alpha=1}^{k} \int \lambda(\phi^{(\alpha)}(s))ds\}, \tag{7}$$

or, equivalently,

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{-\sum_{\alpha=1}^{k} \int \lambda(z)H^{(\alpha)}(\Gamma, z)dz\}. \tag{8}$$

In the above equations, Z is the normalization constant, and it is also called the partition function $Z = Z(\Phi, \Lambda)$ in physics. $\Lambda = (\lambda^{(1)}(), \lambda^{(2)}(), ..., \lambda^{(k)}())$ are Lagrange multipliers, or potential functions. Since the constraints are imposed for continuous variables z and s, $\lambda^{(\alpha)}()$ is a continuous function. Because of the homogeneity assumption, $\lambda^{(\alpha)}(), \alpha = 1, 2, ..., k$ are independent of s, and therefore they are one dimensional.

It is easy to see that the active contour models in equation (1) and equation (2) are special examples of equation (7) with Φ and Λ specified.

6.2 Estimation and computation

In model $p(\Gamma; \Phi, \Lambda)$, the Lagrange multipliers (or the potential functions) have yet to be solved from the constraint equations (5) and (6). In practice, due to the complexity of Φ and the shape space Ω_Γ, we can only compute Λ numerically. Our method for computing Λ has been successfully applied to texture modeling and prior learning in [32, 31], and we describe it briefly in this section.

In practice, a histogram $H^{(\alpha)}(\Gamma, z)$ is discretized into m bins, represented by a vector $H^{(\alpha)}(\Gamma) = (H^{(\alpha)}_1, H^{(\alpha)}_2, ..., H^{(\alpha)}_m)$. $\lambda^{(\alpha)}()$ is also approximated by a piecewisely constant function, represented by a vector $\lambda^{(\alpha)} = (\lambda^{(\alpha)}_1, \lambda^{(\alpha)}_2, ..., \lambda^{(\alpha)}_m)$. If m is large enough, $\lambda^{(\alpha)}$ is a close approximation to $\lambda^{(\alpha)}()$. In general m may vary with features.

We replace the integration in equation (8) by an inner product, therefore,

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{-\sum_{\alpha=1}^{k} < \lambda^{(\alpha)}, H^{(\alpha)}(\Gamma) >\}. \tag{9}$$

The Lagrange multipliers $\lambda^{(\alpha)}, \alpha = 1, 2, ..., k$ are solved from the constraint equations, or equivalently by maximum likelihood estimation. They can be computed by the following iterative equations,

$$\frac{d\lambda^{(\alpha)}}{dt} = E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}(\Gamma)] - \mu_{\text{obs}}^{(\alpha)}, \quad \alpha = 1, 2, ..., k, \tag{10}$$

where t is time step, and $E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}(\Gamma)]$ is the expected histogram with respect to the current model $p(\Gamma; \Phi, \Lambda)$. When the above dynamical equation converges, i.e. $\frac{d\lambda^{(\alpha)}}{dt} = 0$, then $p(\Gamma; \Phi, \Lambda)$ duplicates the observed statistics. It is well known that Λ has a unique solution as $\log p(\Gamma; \Phi, \Lambda)$ is straight concave with respect to Λ [32], provided that the constraint equations are consistent.

In equation (10), for a given Λ, $E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}(\Gamma)]$ is hard to compute analytically. In general, one needs to simulate a Monte Carlo Markov Chain(MCMC) walking randomly in the shape Ω_Γ. When this MCMC becomes stationary, it samples the distribution $p(\Gamma; \Phi, \Lambda)$, and we shall discuss the design of MCMC in the next section. Given Φ and Λ, we sample a set of shapes $\Gamma_j^{\text{syn}}, j = 1, 2, ..., M'$ from $p(\Gamma; \Phi, \Lambda)$ and we estimate $E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}]$ by the sample mean $\mu_{\text{syn}}^{(\alpha)}$ computed in the same way as for $\mu_{\text{obs}}^{(\alpha)}$ in equation (4).

So Λ is updated by

$$\frac{d\lambda^{(\alpha)}}{dt} = \mu_{\text{syn}}^{(\alpha)} - \mu_{\text{obs}}^{(\alpha)}, \quad \alpha = 1, 2, ..., k. \tag{11}$$

The whole learning process simulates an inhomogeneous Markov, and this computational scheme has been successfully applied to texture modeling by Zhu, Wu and Mumford [32].

6.3 *The learned shape model*

Suppose we adopt the six features as discussed in section (4), an ME model is,

$$p(\Gamma) = \frac{1}{Z} \exp\{-\int \lambda^{(1)}(\kappa(s)) + \lambda^{(2)}(\nabla\kappa(s)) + \lambda^{(3)}(r(s)) + \lambda^{(4)}(\nabla r(s)) + \lambda^{(5)}(\nabla^2 r(s)) ds + \gamma\|B\|\}. \tag{12}$$

In equation (12), $\|B\|$ is the number of discontinuities of $\psi(s)$. The number of branches of a shape is $\|B\|/4$. The model in equation (12) has some desirable properties.

1. Multiple features and Gestalt laws – both region based and contour based – are fused into a single probability measure on Markov random fields.

2. Shape features are weighted by the learned potential functions, and correlations between features are taken into account in the learning process.

3. With more features selected, $p(\Gamma)$ can be extended to account for the complexity and frequency or likelihood of shapes that occurs in nature. Thus as $p(\Gamma)$ approaches $f(\Gamma)$, it provides the optimal coding length $-\log p(\Gamma)$ for natural shapes.

7 Design MCMC for shape sampling

In this section, we discuss a Markov Chain Monte Carlo method for sampling $p(\Gamma; \Phi, \Lambda)$, and we also demonstrate experiments on sampling uniform distributions in the shape space Ω_Γ.

Designing a sampling process is important for the following reasons. Firstly, it is a necessary step for estimating the potential functions Λ. Secondly, it provides a natural way for verifying the learned models. If a shape model is close to the underlying truth, then the typical set of samples from this model should be similar to the observed shapes judged by human perception. Thirdly, it is also an engine for shape inference from real images when the learned shape model is used as a prior distribution in vision tasks.

In section (3), a shape $\Gamma = (\,(x_0, y_0), (x_1, y_1), ..., (x_{N-1}, y_{N-1})\,)$ is defined on a continuous space Ω_Γ. In computer implementation, the cruel reality is that the nodes $(x_i, y_i), = 0, 1, 2, ...$ can only be described in finite precision, and we denote by τ the unit length of the lattice [6]. The sampling process starts with a simple closed curve, such as a circle, or rectangle with N nodes. At each step, it randomly picks up a node (x_i, y_i), and proposes to move it to $(x_i^{'}, y_i^{'})$ – a small perturbation moving Γ to $\Gamma^{'}$. If $(x_i^{'}, y_i^{'})$ violates the two hard constraints, then the proposal is rejected, otherwise it is accepted with a probability computed from the model $p(\Gamma; \Phi, \Lambda)$. This is known as a Metropolis-Hastings algorithm, and it simulates a stochastic process –random walk in the shape space Ω_Γ. If it runs long enough, this stochastic process converges to its equilibrium. When this occurs, its status Γ is subject to distribution $p(\Gamma; \Phi, \Lambda)$ regardless of its starting status, and Γ is a synthesized shape from $p(\Gamma; \Phi, \Lambda)$.

[6] One can represent the nodes with sub-pixel accuracy by setting $\tau = 1/5$ or $1/10$ of a pixel width.

There are two immediate questions that we have to address before we discuss the algorithm in details. Firstly, in the shape constraint II of section (3), we have introduced the ϵ because of discretization, what is the effect of ϵ in stochastic sampling? Secondly, how do we diagnose whether or not the random walk becomes stationary? There is no precise answer to the first question due to the complexity of the shape space Ω_Γ. If $\frac{\epsilon}{ds} \to 0$, then we expect the ϵ effect should vanish, however if ϵ is too small, the sampling process converges very slowly. We argue that the choice of the ratio $\frac{\epsilon}{ds}$ should also depend on the precision of human perception. We did two comparison experiments. In one experiment, we used $\epsilon = 2\tau$ and $ds = 10\tau$, and in the other, we used $\epsilon = 2\tau$, and $ds = 20\tau$. There were no noticeable differences for the learning process, except that the convergence becomes slower for the latter. There is no good answer for the second question either, the way we use to diagnose the convergence is to monitor the convergence of the marginal distributions $\mu^{(\alpha)}_{\text{syn}}, \alpha = 1, 2, ...$ estimated from the most recent samples, and we monitor the sampled shapes by our perception. The objective of the sampling process is to match $\mu^{(\alpha)}_{\text{syn}}$ to $\mu^{(\alpha)}_{\text{obs}}$, and our experiments in later sections show that the MCMC process can achieve this goal.

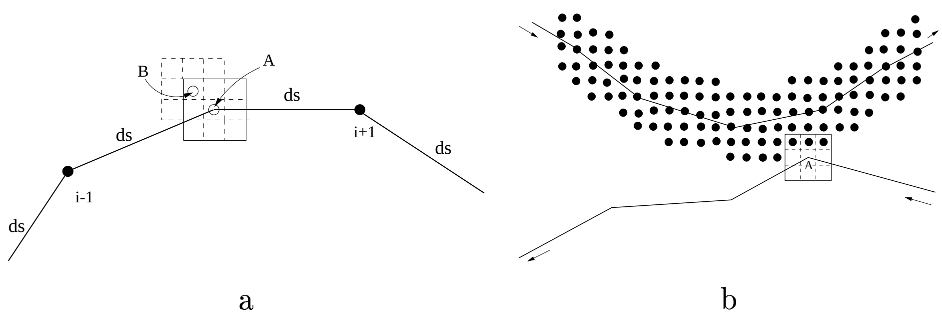


Figure 7 a. The proposal for moving a node from position A to B. b). The firewall preventing the curve from self-intersecting itself.

Now we discuss the details of the sampling algorithm.

Suppose at a certain step, a node $A = (x_i, y_i)$ is chosen at random, and A is allowed to move in a local neighborhood $\{x_i - \tau, x_i, x_i + \tau\} \times \{y_i - \tau, y_i, y_i + \tau\}$-the solid square in figure 7.a. Suppose there are N_A $(1 \leq N_A \leq 9)$ valid moves out of the 9 positions, because putting node A in any of the other $(9 - N_A)$ positions may violate the two hard shape constraints. Now we **propose** to move A to one of the N_A positions B at equal chance $K(A \to B) = \frac{1}{N_A}$. Similarly we compute N_B as the number of valid moves in the neighborhood of B –the dashed square. Obviously it is valid to move from B back to A, and the chance for proposing a move from B to A is $K(B \to A) = \frac{1}{N_B}$. This guarantees that the sampling process is reversible – an important precondition for designing the stochastic process.

The move from A to B is one step of random walk, and we denote by Γ_A and Γ_B the two curves respectively. The proposal is accepted with probability,

$$\alpha(A \to B) = \min\left(\frac{K(B \to A)p(\Gamma_B; \Phi, \Lambda)}{K(A \to B)p(\Gamma_A; \Phi, \Lambda)}, 1\right).$$

In the above equation, $p(\Gamma_A; \Phi, \Lambda)$ and $p(\Gamma_B; \Phi, \Lambda)$ are, due to Markov property, computed as $p((x_i, y_i)| \cup_{j \in \partial i} \{(x_j, y_j)\})$, where ∂i is the local neighborhood set of node i. For example, if $\phi^{(1)} = \kappa(s)$ is the only feature chosen in the model, then the computation of the probability of point (x_i, y_i) will only involve (x_{i-1}, y_{i-1}) and (x_{i+1}, y_{i+1}). If region based properties are integrated in the model, then we compute $r(s)$ from the current mapping function $\psi(s)$ for $p(\Gamma_A; \Phi, \Lambda)$. The mapping function ψ for Γ_B should be updated locally[7], and $r(s)$ at node i is recomputed for estimating $p(\Gamma_B; \Phi, \Lambda)$.

In summary, the transpose probability is $P(A \to B) = K(A \to B)\alpha(A \to B)$, and the random walk satisfies the detailed balance equation

$$p(\Gamma_A; \Phi, \Lambda)P(A \to B) = p(\Gamma_B; \Phi, \Lambda)P(B \to A).$$

This guarantees that if the stochastic process walks long enough, its statuses, i.e. Γ, are samples from $p(\Gamma; \Phi, \Lambda)$.

The algorithm for sampling $p(\Gamma; \Phi, \Lambda)$ is given as follows.

Algorithm I: stochastic algorithm for shape sampling

Step 1:	initialize $\Gamma = ((x_0, y_0), ..., (x_{N-1}, y_{N-1}))$.
Step 2:	initialize the mapping function $\psi(s)$ for Γ.
	if region based features are chosen in Φ.
Step 3:	sweep $\leftarrow 0$.
Step 4:	for *count* $= 1$ to N do.
Step 5:	Pick up $i \in [0, N-1]$ at random, node i is at position A.

[7] As the mapping function $\psi(s)$ is computed in a Markov random field, thus updating $\psi(s)$ only involves local computations[33].

Step 6:	Compute N_A, pick up a position B at random, compute N_B.
Step 7:	Update ψ for the new curve, if necessary.
Step 8:	Compute $\alpha(A \to B)$.
Step 9:	Draw random number $r \in [0, 1)$ at uniform dist.
Step 10:	If $r < \alpha$, then move node i from A to B.
Step 11:	sweep$\leftarrow$ sweep+1
Step 12:	stop, if sweep $>$ threshold, or go to step 4.

In our experiment, to prevent a curve from intersecting itself, we construct a "firewall" of width 5τ along the curve (see the black dots in figure 7.b). The cells along the line segment from node i to node $i+1$, displayed by the dots in figure 7.b, are labeled as i, and a node j is prevented from moving in the firewall cells labeled i if $(|j-i| \, mod \, N) \geq 2$. For example, node A in figure 7 cannot move in the three dotted cells, and $N_A = 6$.

Our first experiment is to sample a uniform distribution on Ω_Γ. 100 random shapes are recorded after 1.5 million sweeps with $N = 200$, $ds = 10\tau, \epsilon = 2\tau$. Figure 8 displays four of the sampled shapes. To study how the statistics of these random shape vary with resolutions, we subsampled the 100 shapes once and twice, and obtained shapes of 100 nodes, and 50 nodes respectively. Figure 9.a shows the average histograms of 100 synthesized shapes at three scales: $H_{200}^{(m)}, m = 0, 1, 2$ by dash-dotted, dashed, and solid curve respectively. m is the times of subsampling. Obviously the histograms change over scales. With a higher resolution, local structures of the random shapes become richer, and display fractal properties. This scale-sensitivity is in sharp contrast to the pseudo-scale-invariant property of natural shapes in figure 5

Our second experiment is to diagnose the convergence of the Markov chain, and to study the effect of the resolution N in the synthesized shapes. We sample a second group of 100 shapes with $N = 400, ds = 10\tau, \epsilon = 2\tau$, and a third group of 100 shapes with $N = 100, ds = 10\tau, \epsilon = 2\tau$. See our technical report for some of these shapes[34].

Figure 9.b shows two groups of histograms $H_N^{(m)}$, where the subscript N is the number of nodes used in the sampling processes, and the superscript (m) denotes the scales. The first group includes two broad histograms (all have 200 nodes): $H_{200}^{(0)}$ (dashed) and $H_{400}^{(1)}$ (solid). The second group includes three sharper curves (all

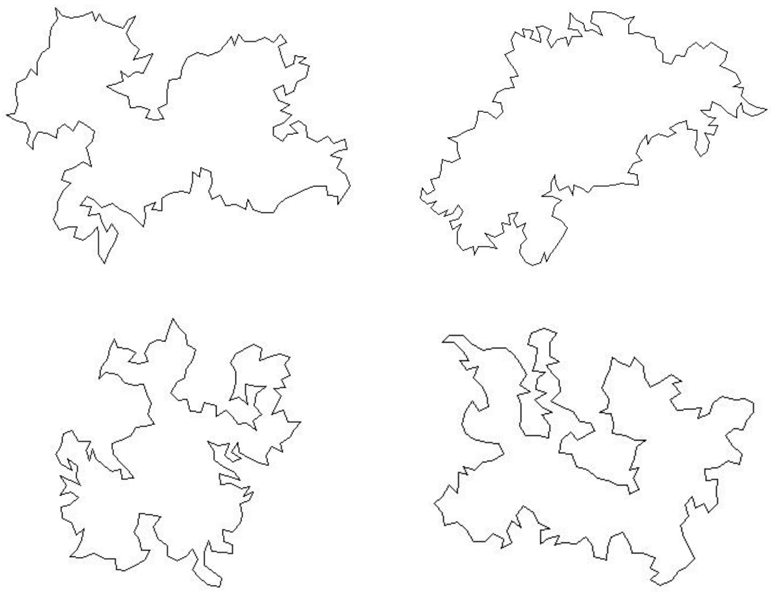
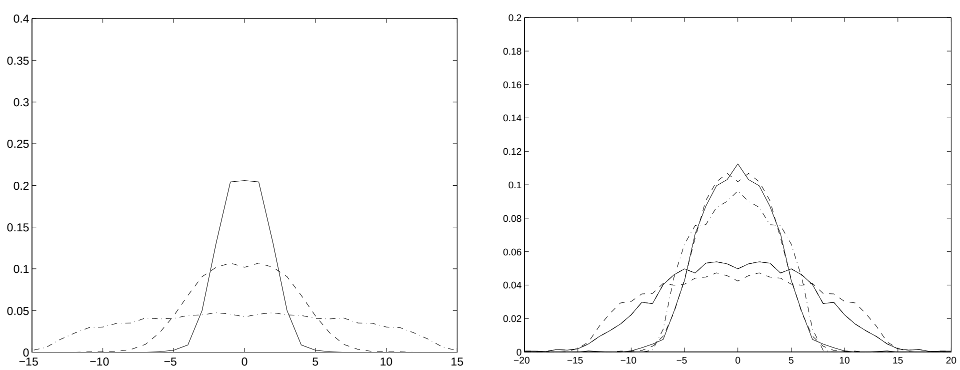


Figure 8 Four of the sampled shapes recorded at different time steps of a stochastic process designed for uniform distribution. $N = 200$, $ds = 10\tau, \epsilon = 2\tau$.



a b

Figure 9 a). Histograms of $\kappa(s)$ averaged over 100 samples at 3 scales $H^{(0)}_{200}$ (dash-dotted), $H^{(1)}_{200}$ (dashed) and $H^{(2)}_{200}$ (solid) respectively. b). The two broad histograms of $\kappa(s)$ are $H^{(0)}_{200}$ (dashed) and $H^{(1)}_{400}$ (solid) respectively, and the three sharper histograms are $H^{(0)}_{100}$ (dash-dotted), $H^{(1)}_{200}$ (dashed) and $H^{(2)}_{400}$ (solid) respectively.

have 100 nodes): $H_{100}^{(0)}$ (dash-dotted), $H_{200}^{(1)}$ (dashed) and $H_{400}^{(2)}$ (solid). It is clear that the histograms observed in the same resolution are very close, which indicates the consistence of the sampling processes with $N = 100$, 200, and 400. Perceptually, the random shapes with $N = 200$ nodes have no noticeable differences with the shapes subsampled from the random shapes with $N = 400$ nodes (see our technical report for some of these shapes[34]).

8 Feature pursuit by maximizing non-accidental statistics

In this section, we discuss how to pursue important features for shape modeling.

We start with $\Phi_0 = \emptyset$, and thus $p(\Gamma)$ a uniform distribution. At each step k, given $\Phi_k = \{\phi^{(\alpha)}, \alpha = 1, 2, ..., k.\}$ and a model $p(\Gamma; \Phi_k, \Lambda_k)$ is learned according to algorithm I. As a result, we obtain a set of synthesized shapes $\{\Gamma_i^{\text{syn}}, i = 1, 2, ..., M'\}$ and $\mu_{\text{syn}}^{(\alpha)} = \mu_{\text{obs}}^{(\alpha)}$, for $\alpha = 1, 2..., k$. Now for any new feature $\phi^{(\beta)} \notin \Phi_k$, we compute the histograms $\mu_{\text{syn}}^{(\beta)}$ and $\mu_{\text{obs}}^{(\beta)}$ from the synthesized and the observed shapes respectively. We notice that $\mu_{\text{syn}}^{(\beta)}$ is the *accidental statistic* for feature $\phi^{(\beta)}$, and it accounts for the correlation between $\phi^{(\beta)}$ and the chosen features in Φ_k.

Definition. *Given Φ_k and $p(\Gamma; \Phi_k, \Lambda_k)$, the* **non-accidental statistic** *for feature $\phi^{(\beta)}$ is the distance between $\mu_{\text{obs}}^{(\beta)}$ and $\mu_{\text{syn}}^{(\beta)}$.*

At step $k+1$, we should choose the feature which has the largest non-accidental statistic. The distance between $\mu_{\text{obs}}^{(\beta)}$ and $\mu_{\text{syn}}^{(\beta)}$ could be measured by a L_1 norm, or a quadratic form – the Mahanalobis distance.

In modeling texture[32], Zhu, Wu and Mumford proposed a minimum entropy principle for selecting the optimal set Φ of features from a dictionary of Gabor filters. It is proven that each feature selection step from $p(\Gamma; \Phi_k, \Lambda_k)$ to $p(\Gamma; \Phi_{k+1}, \Lambda_{k+1})$ is a steepest descent way for minimizing the Kullback-Leibler distance $D(f||p)$ between the true distribution $f(\Gamma)$ and the model $p(\Gamma)$. $D(f||p)$ is a conventional measure for the goodness of the learned model $p(\Gamma)$. For shape modeling, the Gestalt laws play the same role as Gabor filters for texture modeling, however, there are far less Gestalt laws than Gabor filters. So we choose not to discuss the feature selection issue explicitly, and the proof for the following proposition is referred to our texture paper[32].

Proposition. *At each step, choosing a feature which has the maximum non-*

accidental statistic is a steepest descent step of minimizing the Kullback-Leibler distance $D(f\,||\,p)$.

In summary, the overall algorithm for shape learning is list below.

Algorithm II: algorithm for shape learning

Step 1:	given $\{\Gamma_i^{\text{obs}}$, for $i = 1, 2, ...M\}$, compute $\mu_{\text{obs}}^{(\alpha)}, \alpha = 1, 2, ...$
Step 2:	$k = 0$, initialize $\Phi_0 = \emptyset, \Lambda_0 = 0$, and $p(\Gamma; \Phi_0, \Lambda_0)$ a uniform dist.
Step 3:	sample $\Gamma_i^{\text{syn}}, i = 1, 2..., M'$ from $p(\Gamma; \Phi, \Lambda)$ using algorithm I.
Step 4:	compute $\mu_{\text{syn}}^{(\beta)}$ for candidate features $\phi^{(\beta)}$.
Step 5:	adding a new feature, $\Phi \leftarrow \Phi \cup \{\phi^{(k)}\}$, $\lambda^{(k)} \leftarrow 0$, $k \leftarrow k + 1$.
Step 6:	$\lambda^{(\alpha)} \leftarrow \lambda^{(\alpha)} + \Delta t(\mu_{\text{syn}}^{(\alpha)} - \mu_{\text{obs}}^{(\alpha)})$, $\alpha = 1, 2..., k$.
Step 7:	sample $p(\Gamma; \Phi, \Lambda)$ for $10,000$ sweeps using algorithm I.
Step 8:	compute $\mu_{\text{syn}}^{(\alpha)}, \alpha = 1, 2..., k$ from a set of recently sampled shapes.
Step 9:	goto step 6, unless $\|\mu_{\text{syn}}^{(\alpha)} - \mu_{\text{obs}}^{(\alpha)}\| < \rho$ for $\alpha = 1, 2, ..., k$.
Step 10:	goto step 5, unless $\|\mu_{\text{obs}}^{(\beta)} - \mu_{syn}^{(\beta)}\| < \Upsilon$ for all remaining features $\phi^{(\beta)}$.

The whole learning process is computationally very intensive, especially when the region based features are computed.

9 Experiments

In this section, we demonstrate experiments on shape learning, following algorithms I and II.

Experiment I

We start with $\Phi = \emptyset$ and $p(\Gamma; \Phi, \Lambda)$ a uniform distribution. In figure 10.a, we compare $\mu_{\text{syn}}^{(1)}$(dashed) of the uniform distribution against $\mu_{\text{obs}}^{(1)}$(solid). $\mu_{\text{obs}}^{(1)}$ has a much sharper peak than $\mu_{\text{syn}}^{(1)}$ near zero, $\mu_{\text{obs}}^{(1)}$ measures quantitatively how much co-linearity natural shapes have in comparison with $\mu_{\text{syn}}^{(1)}$ the co-linearity resulted from **accidental arrangement** in the uniform shapes while the hard constraints are taken into account. We found that $\|\mu_{\text{syn}}^{(1)} - \mu_{\text{obs}}^{(1)}\| > \|\mu_{\text{syn}}^{(2)} - \mu_{\text{obs}}^{(2)}\|$, which means $\kappa(s)$ is a more significant property than $\nabla\kappa(s)$–the co-circularity in natural shapes. In this initial step, we don't compare other region based features, as it is quite unreliable to compute the symmetric mapping function $\psi(s)$ for the jagged synthesized shapes.

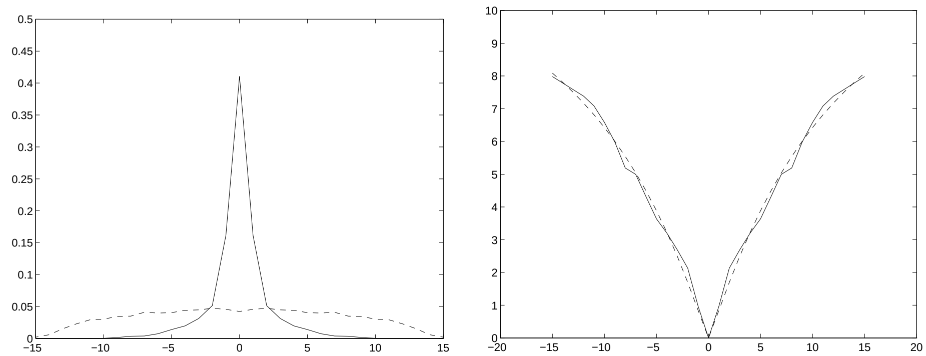


Figure 10 a). $\mu^{(1)}_{\text{obs}}$ (solid) versus $\mu^{(1)}_{\text{syn}}$ of a uniform distribution for $\phi^{(1)}(s)$. These two curves have appeared in figure 5.a and figure 9.a respectively. b). The learned $\lambda^{(1)}()$ for $\phi^{(1)}(s) = \kappa(s)$ in a non-parametric form (solid curve), and it is fit to $\eta(z) = a(1-1/(1+(z/b)^{\gamma}))$ with $a = 15, b = 13, \gamma = 1.1$, $dz = \frac{\pi}{16\times 200}$ is the unit length of the horizontal axis $\kappa(s)$.

Then a Gibbs distribution is learned which could reproduce the marginal distribution for feature $\phi^{(1)} = \kappa(s)$,

$$p(\Gamma; \Phi_1, \Lambda_1) = \frac{1}{Z}\exp\{-\int \lambda^{(1)}(\kappa(s))ds\}.$$

We adopt the continuous notation for consistence, and the learned potential function $\lambda^{(1)}(z)$ is plotted in figure 10.b. $\lambda^{(1)}(z)$ is close to $|z|$ near zero, but it has flat tails to preserve large curvatures. Figure 11 displays six of the sampled shapes from the learned model. From the set of sampled shapes, we compute $\mu^{(\alpha)}_{\text{syn}}, \alpha = 1, 2$. Figure 12.a displays $\mu^{(1)}_{\text{syn}}$ (dashed)–a marginal distribution of $p(\Gamma; \Phi_1, \Lambda_1)$ against $\mu^{(1)}_{\text{obs}}$ (solid), and figure 12.b plots $\log \mu^{(1)}_{\text{syn}}$ (dashed) against $\log \mu^{(1)}_{\text{obs}}$ (solid). Obviously the shape model $p(\Gamma; \Phi_1, \Lambda_1)$ reproduces the observed curvature histogram precisely. $\mu^{(2)}_{\text{syn}}$ is plot against $\mu^{(2)}_{\text{obs}}$ in figure 12.c, so the sampled shapes doesn't have enough co-circularity as in the observed shapes.

For comparison, we also sampled a Gaussian distribution,

$$p(\Gamma) = \frac{1}{Z}\exp\{-\int \kappa^2(s)/\sigma^2 ds\}$$

with σ chosen to reproduce the same mean and the same variance of the curvature as in the observed shapes. We show three sampled shapes in figure 13. These shapes have more jagged boundaries, and have less structures than the shapes in figure 11.

Experiment II

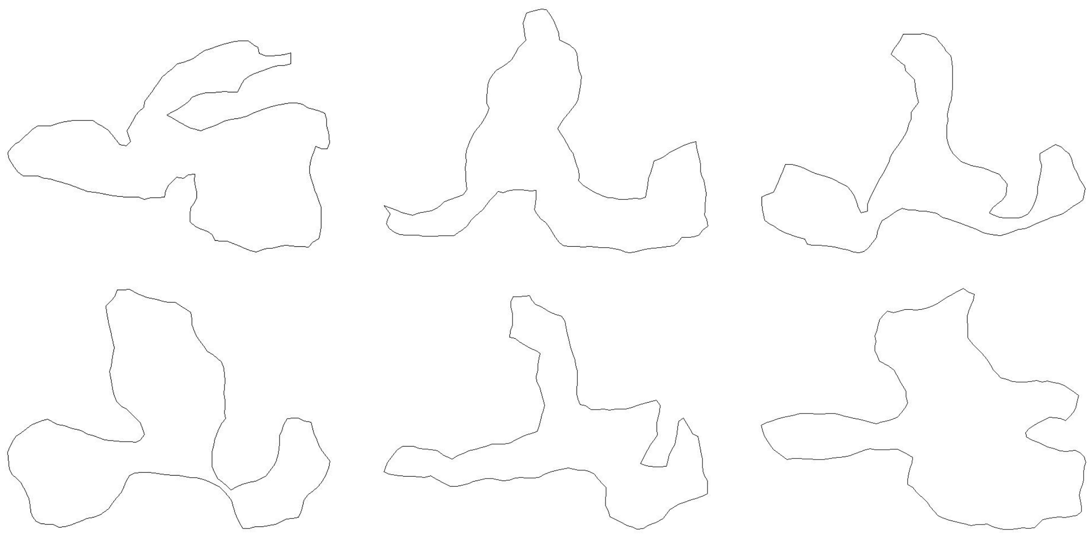
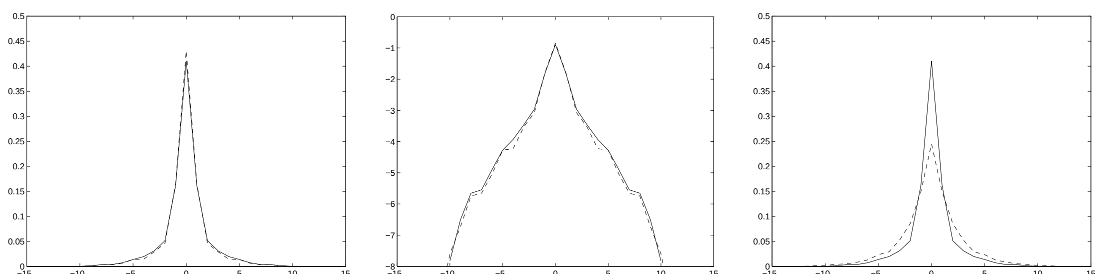


Figure 11 Six of the synthesized shapes with curvature histogram matched to animate shapes, $\mu^{(1)}_{\mathsf{syn}} = \mu^{(1)}_{\mathsf{obs}}$. The histograms of these synthesized shapes are shown by the dashed curves in figure 12.



a b c

Figure 12 a). $\mu^{(1)}_{\mathsf{syn}}$ (dashed) vs. $\mu^{(1)}_{\mathsf{obs}}$ (solid). b) $\log \mu^{(1)}_{\mathsf{syn}}$ (dashed) vs. $\log \mu^{(1)}_{\mathsf{obs}}$ (solid). c). $\mu^{(2)}_{\mathsf{syn}}$ (dashed) vs. $\mu^{(2)}_{\mathsf{obs}}$ (solid).

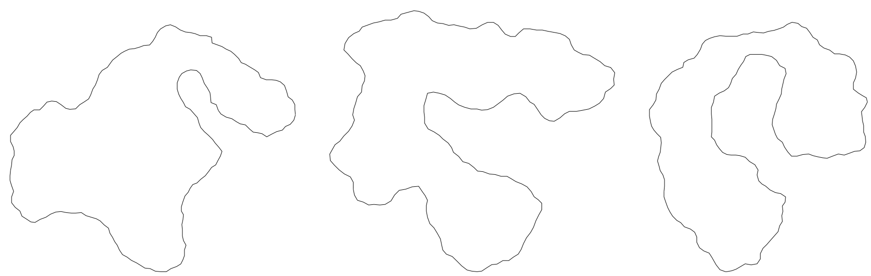


Figure 13 Sampled shapes from a Gaussian model after 2 million sweeps. The variance of $\kappa(s)$ for the animal shape is 0.1665 , the sampled shapes have variance 0.1657, which is a very close match.

In the second experiment, we choose $\Phi_2 = \{\kappa(s), \nabla\kappa(s)\}$, and the model is learned to match both $\mu^{(1)}_{\text{obs}}$ and $\mu^{(2)}_{\text{obs}}$,

$$p(\Gamma; \Phi_2, \Lambda_2) = \frac{1}{Z}\exp\{-\int \lambda^{(1)}(\kappa(s)) + \lambda^{(2)}(\nabla\kappa(s))ds\}.$$

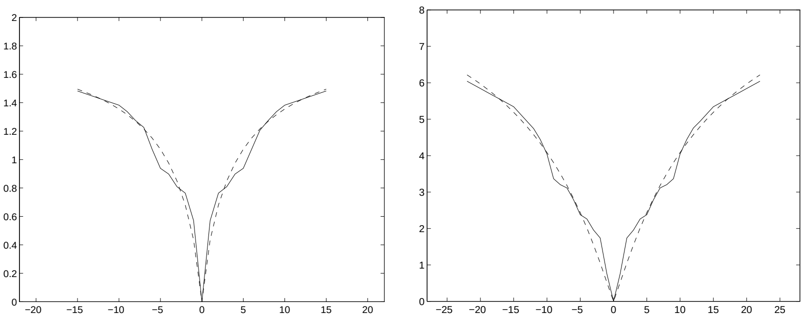
The learned $\lambda^{(1)}(z)$ and $\lambda^{(2)}(z)$ are shown as the solid curves in figure 14.a. and b respectively. Six of the sampled shapes are shown in figure 15. The average histograms $\mu^{(\alpha)}_{\text{syn}}, \alpha = 1, 2$ for 45 synthesized shapes are shown in figure 16.a and b by the dashed curves, and they match $\mu^{(\alpha)}_{\text{obs}}, \alpha = 1, 2$ closely.

The shapes in figure 15 share the same amount of co-linearity and co-circularity as the observed animate shapes. These shapes are smooth, and we also notice that long circular arcs are formed through the propagations of local interactions in Markov random fields.

We want to emphasize two issues in this experiment.

- The synthesized shapes have pseudo scale invariant property like the observed shapes. For example, we collect 45 sampled shapes, and subsample them once and twice, and generate two new sets of shapes at scale 1 and scale 2. Figure 16.c shows the average curvature histograms for the three scales. This indicates that *the choice of N is not critical in shape modeling as long as it is large enough to approximate the curve up to the precision of human perception.*

- The learned potential function $\lambda^{(1)}$ in figure 14.a is different from that in figure 10.b. This change reflects the correlation between the two features. This demonstrates that the learning process could account for inter-dependency between chosen features.



a b

Figure 14 The learned potential functions. a). $\lambda^{(1)}(z)$ (solid), and the fitting curve $\eta(z) = a(1 - 1/(1 + (z/b)^{\gamma}))$ with $a = 1.95, b = 4, \gamma = 0.9$, b). $\lambda^{(2)}(z)$ (solid), and the fitting curve $\eta(z) = a(1 - 1/(1 + (z/b)^{\gamma}))$ with $a = 10, b = 14, \gamma = 1.1$.

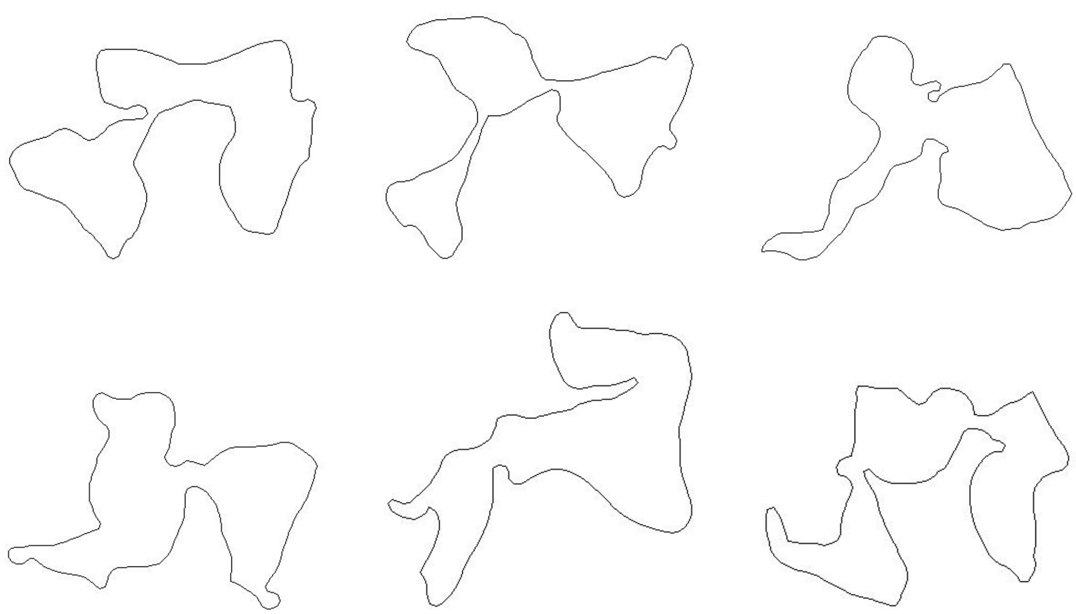
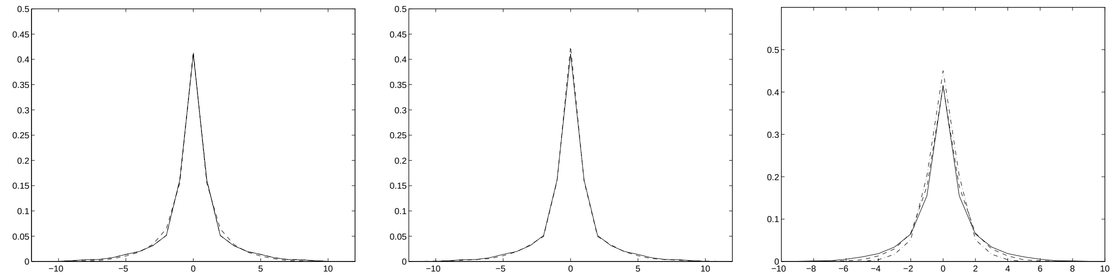


Figure 15 Six of the synthesized shapes with $\mu_{\text{syn}}^{(\alpha)} = \mu_{\text{obs}}^{(\alpha)}, \alpha = 1, 2$. Note that these shapes are smoother than the shapes in figure 11.



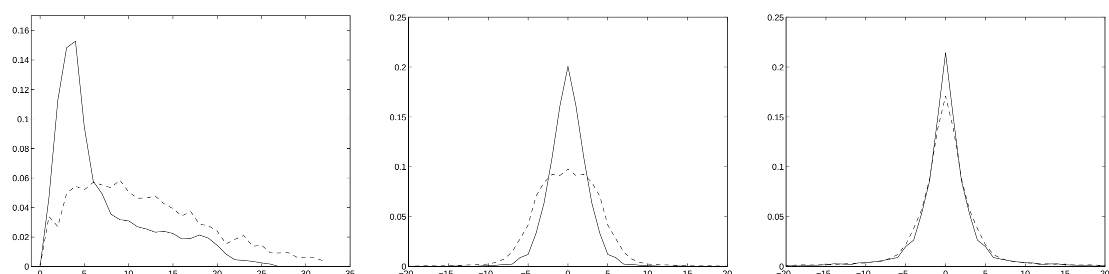
a b c

Figure 16 a). the histogram of $\kappa(s)$ averaged over 45 synthesized shapes (dashed curve) and $\mu^{(1)}_{\text{obs}}$ (solid curve); b). the histogram of $\nabla\kappa(s)$ averaged over 45 synthesized shapes (dashed curve) and $\mu^{(2)}_{\text{obs}}$ (solid curve); c). The histograms of $\kappa(s)$ averaged over 45 synthesized shapes at scale 0 (solid curve), scale 1 (dashed curve), and scale 2 (dash-dotted curve).

Experiment 3

Our third experiment incorporates the region based features into the shape model.

Although the shapes in figure 15 reproduce the exact amount of co-linearity and co-circularity, they are very blob-like, and elongated parts, like limbs of animals, are missing. The lack of region based features are reflected in the histograms plotted in figure 17. We compute $\mu^{(\alpha)}_{syn}, \alpha = 3, 4, 5$ for features $\phi^{(3)}(s) = r(s), \phi^{(4)}(s) = \nabla r(s)$, and $\phi^{(5)}(s) = \Delta r(s)$ from a set of 45 synthesized shapes of $p(\Gamma; \Phi_2, \Lambda_2)$. $\mu^{(\alpha)}_{syn}, \alpha = 3, 4, 5$ are plotted as dashed curves in figure 17.a,b,c respectively. In contrast, $\mu^{(\alpha)}_{\text{obs}}, \alpha = 3, 4, 5$ are the solid curves. Again the differences measure the non-accidental statistics in natural shapes.



a b c

Figure 17 Average histograms of shapes sampled from $p(\Gamma; \Phi_2, \Lambda_2)$. $\mu^{(\alpha)}_{\text{syn}}$ (dashed) vs. $\mu^{(\alpha)}_{\text{obs}}$ (solid), a), $\alpha = 3$, b), $\alpha = 4$, c), $\alpha = 5$.

As the difference between $\mu^{(5)}_{\rm syn}$ and $\mu^{(5)}_{\rm obs}$ is small, we choose two new features $r(s), \nabla r(s)$, and learned a new model,

$$p_4(\Gamma; \Phi_4, \Lambda_4) = \frac{1}{Z}\exp\{-\int \lambda^{(1)}(\kappa(s)) + \lambda^{(2)}(\nabla\kappa(s)) + \lambda^{(3)}(r(s)) + \lambda^{(4)}(\nabla r(s)) ds\}.$$

The learned $\lambda^{(\alpha)}, \alpha = 1, 2, 3, 4$ are shown in figure 18. Six of the sampled shapes are displayed in figure 19 together with their mapping functions $\psi(s)$.

This experiment demonstrates that the region based features are crucial for shape modeling. Although none of these synthesized shapes are identifiable as specific animate objects in our environments, the sampled shapes resemble parts of real world objects, exactly as we expected for middle level vision. We shall debate on this issue in the next section.

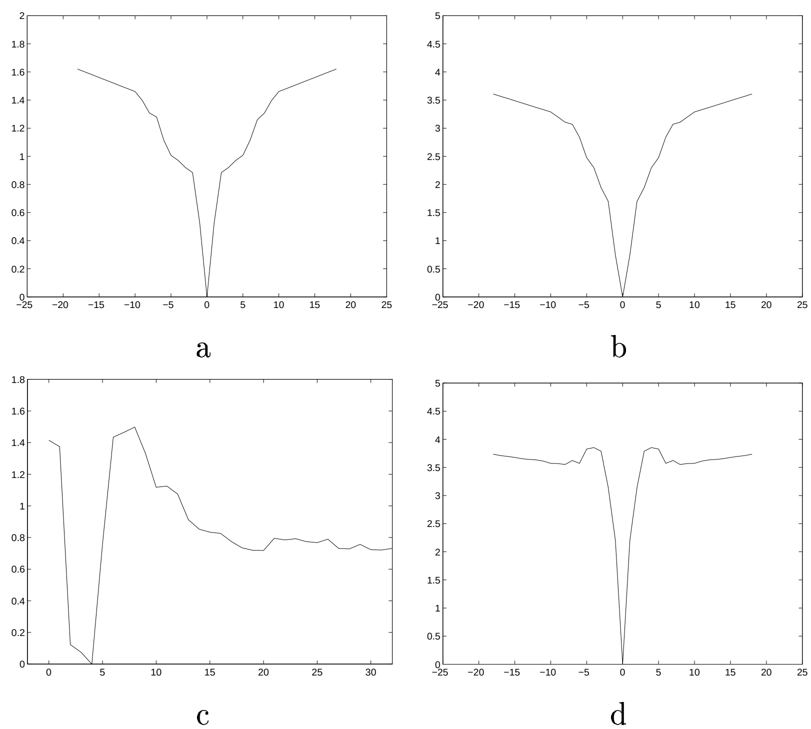


Figure 18 The learned potential functions in $p_3(\Gamma; \Phi_3, \Lambda_3)$. a). $\lambda^{(1)}(z)$, b). $\lambda^{(2)}(z)$, c). $\lambda^{(3)}(z)$, d). $\lambda^{(4)}(z)$.

10 Discussion

In this paper, a theory for shape modeling and learning is proposed based on the maximum entropy principle, and various Gestalt laws are embedded into the Markov random field models. The paper also provides quantitative measures for the non-accidental arrangement by distances between the observed statistics and the statistics of random shapes. The differences between $\mu_{\mathrm{syn}}^{(\alpha)}$ and $\mu_{\mathrm{obs}}^{(\alpha)}$ are evidences for the ecologic reasons underlying the Gestalt laws –co-linearity, co-circularity, proximity, parallelism, and symmetry – identified in experiments of human visual perception. Thus the article provides a firm mathematical justification for some Gestalt laws.

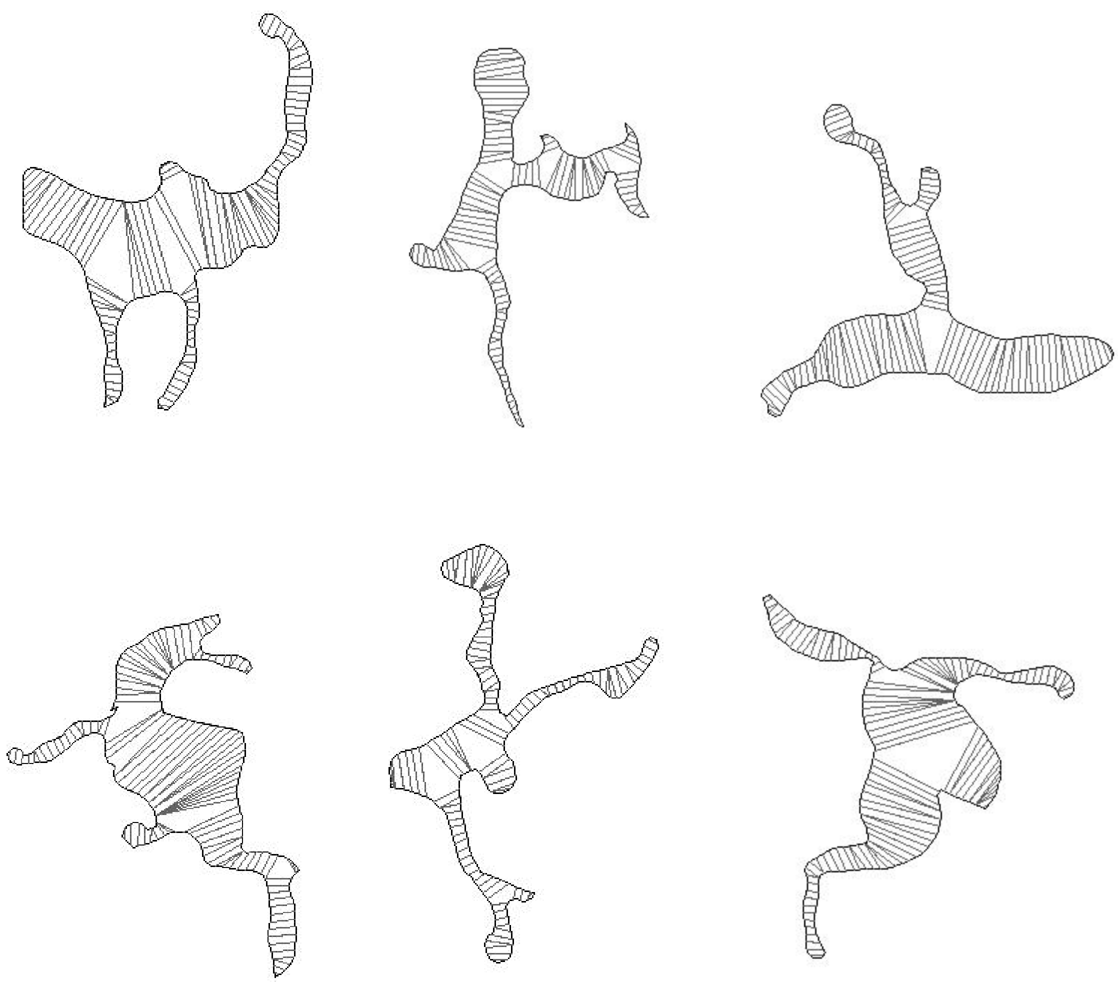


Figure 19 Six of the synthesized shapes from $p_4(\Gamma; \Phi_4, \Lambda_4)$.

The learned models are limited in a few aspects.

Firstly, they are biased by the training shapes that we selected. We have studied

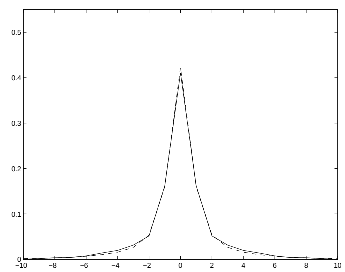


Figure 20 The solid line is the average histogram of curvature $\kappa(s)$ for 22 animals (flipped to 44). The dashed line is the average histogram of $\kappa(s)$ for 1100 fish– a different dataset.

a different database of shapes – 1100 fish contours, and the statistics in this dataset are found to be very similar to those of our training set (see figure 20). It is not very surprising because both datasets are animate objects, and histogram is averaged over large samples (the total number of points in the dataset is over 5,000). However, the statistics are likely to change for the ensemble of polygon shapes such as rectangles and triangles, whose curvature histograms should have sharper peak at zero.

Secondly, the neighborhood structures of our models are limited by the Gestalt laws. There are very few choices for generic shape features. The Gestalt laws are known to be the most important generic features for human perception, if they are not the best. For the same reason, we don't discuss the histogram variations between animate shapes, and we have assumed that $\mu^{(\alpha)}_{\text{obs}}(z) = \mu^{(\alpha)}(z)$ in this paper. Indeed, studying the histogram variation and the estimation error of the observed histograms is important to decide when to stop choosing a feature, this was discussed in our texture modeling paper[32].

Thirdly, the models do not account for high level shape properties. For example, some portions of the sampled shapes in figure 19 may resemble parts of animals, such as legs, tails, and heads, but these parts are not assembled in a proper way.

We now discuss how the models can be extended to overcome the third limitation.

One can easily compute the medial axes (or skeletons) for the shapes in figure 19 by connecting the centers of the mapping line segments, and a detailed algorithm is referred to a companion paper[33]. The skeleton forms a graph representation of shape. For example, figure 21 shows a shape model for high level recognition in the FORMS system proposed by Zhu and Yuille[30]. The skeleton of a dog is generated by a grammar, and the seven deformed parts are joined by two hinge joints. In fact, it is not hard to argue that the deformable models in FORMS are non-homogeneous maximum entropy model too! This non-homogeneous model imposes high level structures on the skeleton graphs to model specific objects, e.g. by imposing constraints on the number of branches at each joint, and the relative orientations of these branches. In this way, one can sample random shapes for a dog.

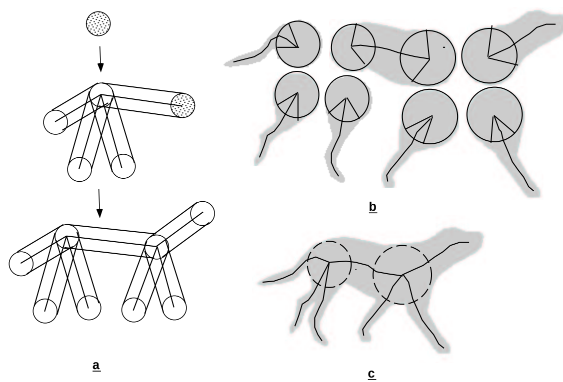


Figure 21 The general deformable model for describing a dog in the FORMS system (Zhu and Yuille 1996). a). the grammar for generating the graph structure of an animal, b). the deformation of parts, c). composing the shape through the hinge joints.

We argue that shape models in the middle level should be **compatible** with both low level and high level representations. The Markov random field shape models in this paper can be easily incorporated into the statistical framework for image segmentation[29]. They also provide a mechanism for computing object shapes as well as their medial axes from raw images, and for passing such a description to the high level.

It is our belief that the random field models may lead to a new way to explain the "field force" in the Gestalt theory. We wish that this paper could stimulate better research results along this direction.

Acknowledgment: The author is supported by an ARO grant DAAH-04-95-1-0494, an NSF grant NSF-9877127, and a Microsoft Research gift. The author is grateful to David Mumford, Yingnian Wu for discussions. The author'd like to thank for two anonymous reviewers for comments that leads to improving the presentation of the paper.

References

[1] R. Balboa and N. M. Grzywacz, "Power spectra of natural underwater images and possible implications for vision", Preprint of the Smith-Kettlewell Eye Research Institute, 1998.

[2] F. L. Bookstein, "Size and shape spaces for landmark data in two dimensions", *Statistical Science*, vol.1, No.2, 181-242, 1986.

[3] C. A. Burbeck and S. M. Pizer, "Object representation by cores: identifying and representing primitive spatial regions", *Vision Research*, 35(13), 1917-1930, 1995.

[4] A. Chakraborty, L. H. Straib, and J. S. Duncan, "Deformable boundary finding influenced by region homogeneity", *Proc. of CVPR*, Seattle, 1994.

[5] S. J. Dickinson, A. P. Pentland, and A. Rosenfeld, "From volumes to views: an approach to 3D object recognition", *CVGIP: Image Understanding*, Vol.55, No.2, pp. 130-154, March 1992.

[6] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, Wiley Publisher, 1998.

[7] U. Grenander, *General Pattern Theory*, Oxford University Press, 1993.

[8] U. Grenander, Y. Chow, and K. M. Keenan, *Hands: A Pattern Theoretical Study of Biological Shapes*, Springer-Verlag, New York, 1991.

[9] H. Helmholtz, *Treatise on Physiological Optics*, Dover, New York 1962 (first published in 1867).

[10] J. E. Hochberg, "Effects of the Gestalt revolution: the Cornell symposium on perception", *Psychological Review*, 64(2), 73-84, 1957.

[11] D. W. Jacobs, *Recognizing 3D objects using 2D images*, Unpublished Ph.D. dissertation, Dept. of EECS, MIT, 1992.

[12] E. T. Jaynes, "Information theory and statistical mechanics", *Physical Review* 106, 620-630, 1957.

[13] G. Kanizsa, *Organization in Vision*, Praeger, New York, 1979.

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models", *Proc. of Int'l Conf. on Computer Vision*, London, 1987.

[15] D. G. Kendall, "A survey of the statistical theory of shape", *Statistical Science*, vol.4, No.2, 87-120, 1989.

[16] D. Keren, D. Cooper, and J. Subrahmonia, "Describing complicated objects by implicit polynomials", *IEEE Trans. on PAMI*, Vol.16, No.1, January 1994.

[17] K. Koffka, *Principles of Gestalt Psychology*, Harcourt, Brace and Company, New York, 1935.

[18] I. Kovacs and B. Julesz, "Perceptual sensitivity maps within globally defined visual shapes", *Nature* 371, 644-646, 1996.

[19] T. S. Lee, D. B. Mumford and S.C. Zhu, and V. Lamme. "The role of V1 in shape representation", *Computational Neuroscience*, Edited by Bower, Plenum Press, New York, 1997.

[20] M. Leyton, *Symmetry, Causality, Mind*. MIT Press. Cambridge, Mass. 1992.

[21] L. D. Lowe, *Perceptual organization and visual recognition*, Kluwer Academic Publishers, 1985.

[22] D. G. Lowe, "Visual recognition as probabilistic inference from spatial relations", in A. Blake and T. Troscianko (eds.) *AI and Eye*, John Wiley & Sons Ltd., 1990.

[23] K. V. Mardia and I. L. Dryden, "Statistical analysis of shape data", *Biometrika*, 76, 271-281, 1989.

[24] R. Mohan and R. Nevatia, "Perceptual organization for scene segmentation and description", *IEEE Trans on PAMI*, vol.14, No.6, June 1992.

[25] D. B. Mumford, "Elastica and Computer Vision", in C.L. Bajaj (ed.) *Algebraic Geometry and Its Applications*, Springer-Verlag, New York, 1994.

[26] A. P. Pentland, "Perceptual organization and the representation of natural form", *Artificial Intelligence*, 28 pp293-331, 1986.

[27] L. R. Williams and D. W. Jacobs, "Stochastic completion fields: a neural model of illusory contour shape and salience", *Neural Computation*, 9, 837-858, 1997.

[28] A. L. Yuille, "Deformable templates for face recognition", *Journal of Cognitive Neurosciences*, Vol.3, No.1, 1991.

[29] S. C. Zhu and A. L. Yuille, "Region competition: Unifying Snake/balloon, Region Growing and Bayes/MDL/Energy for multi-band Image Segmentation", *IEEE Trans. PAMI*, vol.18, no.9, Sept. 1996.

[30] S. C. Zhu and A. L. Yuille, "FORMS: a Flexible Object Recognition and Modeling System", *Int'l Journal of Computer Vision*, Vol.20, No.3, Dec, 1996.

[31] S.C. Zhu and D.B. Mumford, "Prior learning and Gibbs reaction-diffusion", *IEEE Trans. PAMI*, vol.19, no.11, Nov. 1997.

[32] S.C. Zhu, Y.N. Wu and D.B. Mumford, "Minimax entropy principle and its application to texture modeling", *Neural Computation* Vol. 9, no 8, Nov. 1997.

[33] S. C. Zhu, "Stochastic computation of medial axis in Markov random field", *IEEE Trans. PAMI.* Vol21, No. 11, November, 1999.

[34] S. C. Zhu, "Embedding Gestalt laws in Markov random fields", *Technical report*, Department of Computer Science, Ohio State University, 1998.