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 shape modeling for middle level vision problems, such as image segmentation and perceptual organization. This goal suggests that I) the shape models should be generic, and thus characterizes the most common features of 2D shapes, II) the shape models should be compatible with low level representation, such as raw images and edge maps and the high level descriptions such as deformable templates. In two previous papers (Zhu and Mumford 1997, Zhu, Wu and Mumford 1998)[38, 39], the author and his colleagues have proposed a minimax entropy principle for learning probability models for natural images and textures. In this paper, we apply this learning scheme to shape modeling. In contrast to the Gabor filters in image modeling, the Gestalt laws plays a central role in characterizing 2D shapes! The shape models are of the forms of Gibbs distributions defined on the Markov random fields, whose neighborhood structures correspond to the Gestalt laws –co-linearity, co-circularity, proximity, parallelism, and symmetry. Thus both contour-based and region-based features are accounted for. The potential functions of the Gibbs distributions are learned from observed shape set so that the shape models duplicate the observed statistics of natural shapes. Stochastic algorithms are proposed for learning the shape distributions and for sampling random shapes from the models by Markov chain Monte Carlo. Furthermore, this paper provides a quantitative measure for the *non-accidental statistics*, and put the Gestalt theory on a firm mathematical framework. Our experiments demonstrate that global shape properties can arise from the local interactions of local features.

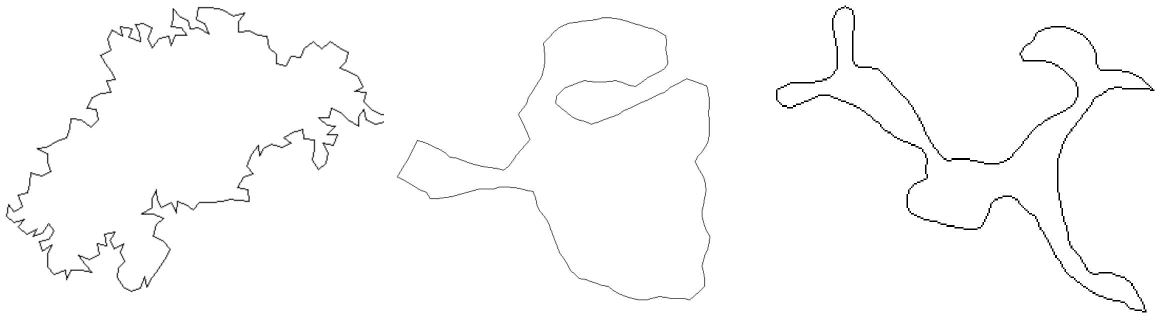
Key words: Gestalt laws, perceptual grouping, shape modeling, Markov random field, maximum entropy, shape synthesis, active contour.

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. For example, figure 1 shows three simple closed curves, which are random shapes sampled from three probability distributions (see later sections). Although neither of them resemble any objects that are familiar to us, the curve on the right has more regular features (co-linearity, co-circularity and symmetry) than the one on the left, and thus it should be favored by pre-attentive human visual perception.

Many theories have been proposed to account for this phenomenon, among which the Gestalt psychology is the most influential one. The Gestalt psychologists proposed an enormous number of laws, which govern the grouping from *parts* to *whole* (Koffka 1935)[21], for examples,

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



a b c

Figure 1 Random shapes sampled from probability distributions of shape at three stages of learning.

In Gestalt psychology, the coordination of these rules is guided by the law of Pragnanz, stated in (Koffka 1935)[21]:

> "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? The Gestalt psychologists explained it in terms of the "field forces" by analogy to the field theory of gravity and electricity, but it is unclear what the "field force" really is. Even worse, the Gestalt psychology only provides a descriptive theory, it does not specify the computational process for achieving the percept from *parts* to *whole*.

Besides the Gestalt laws, there are two other theories for perceptual organization. One is the *likelihood principle* (Helmholtz 1867) [13], which assigns a high probability for grouping two elements, such as line segments, if the placement of the two elements have a low likelihood of resulting from *accidental arrangement* (Lowe 1985,1990)[26, 27]. The other theory is the simplicity or *minimum description length principle* (Hochberg 1957) [14], which states that perceptual organization should achieve the shortest coding length by exploring shape symmetry etc.

We argue that the three theories could be unified into a single framework, provided that the following two conditions are observed.

1. The likelihood probability accounts for the complexity and frequency of configurations occurring in nature, and thus yields the expected minimum coding length. As we know, the description length depends on the language of coding, and by information theory, the optimal coding length of a signal should be the minus logarithm of the "true" probability. [1]

2. The structures of the likelihood probability reflect the Gestalt laws. Indeed, we argue that the Gestalt laws can only be effective *provided that they correspond to statistical regularities in real shapes.*

Therefore, the key issue is to find the generic probability distribution for all possible shapes and configurations in a given application. [2]

Unlike many shape templates used in high level vision, the Gestalt laws are not associated with any specific objects, instead they characterize some generic properties of object shapes. These laws play a dominant role in designing algorithms for solving the middle level vision problems, such as, image segmentation, contour tracing, edgel grouping, and occlusion recovery. Although many successful algorithms have been reported for perceptual organization in the literature[27, 29], a rigorous

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

[2] One probability distribution is better than the others if it leads to smaller expected coding length, or equivalently it has a smaller Kullback-Leibler divergence from the *true distribution*. The optimal shape distribution should be the true distribution itself.

mathematical framework for shape modeling and for integrating the Gestalt laws has yet to be found.

In two previous papers, the author, in collaboration with Wu and Mumford, has successfully studied various probability models for textures and natural images by the minimax entropy principle[39][38]. In this paper, we continue to learn probability distributions for simple closed curves on 2D planes. By simple closed curves, we mean closed, 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. Thus shape modeling is posed as a statistical inference problem, that is, learning the underlying distribution from a set of observed 2D shapes.

We shall address the following questions in this paper:

1. Studies in neurosciences showed that the statistics of the visual environment (i.e. the ecologic effects) play key roles on the behavior of both individual cells and the neural systems[31, 1].[3] Are there any ecologic evidences — embedded in the statistics of realistic shapes — for the Gestalt laws and early human perception of 2D shape?

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

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

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 Gestalt laws?

In this paper, 2D shape models are defined on the Markov random fields (MRF). The variables of the MRF are the coordinates of points along the contour, and the structures of the MRF correspond to the Gestalt laws such as proximity, co-linearity, co-circularity, parallelism and symmetry. We start with exploring the statistics of features extracted from observed real shapes, and then a shape distribution is

[3]For example, there are considerable functional differences of the neurons in the retina between rabbit and fish, and these difference 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.

learned such that it reproduces the observed statistics while having the maximum entropy. The learned shape models are of the Gibbs forms where properties, which correspond to the Gestalt laws, are weighted in the potential function. Furthermore the shape model is verified by stochastic sampling using Markov chain Monte Carlo. The paper also provides a quantitative measure for the non-accidental statistics by comparing the observed statistics and the statistics of the random 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 the basic issues in shape modeling in section (3). Section (4) discusses feature extraction for shape modeling, and section (5) demonstrates experiments on statistics of real shapes. Section (6) presents the maximum entropy theory for probability learning. Section (7) discusses the issues of designing stochastic sampling process in the shape space, and describes experiments of sampling the uniform distribution of shape. Section (8) discusses the issue of feature selection by maximizing the 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 work 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. The literatures on shape modeling and recognition are divided into four areas: 1). 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 [19]). 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 the 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 group , e.g. $SO(m)$. Then each $SO(m)$ equivalence class is viewed as a single point–a real shape[19]. Then he studies the probability structures in this shape space. For example, assuming independent 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 the Delaunay tessellation of Galaxy[19]. For example, he studied how likely the co-linearity or hole is not an accidental arrangement by the i.i.d. uniform distribution.

However, in the past two decades, people knew very little about the structures of this elegant shape space, even some seemingly trivial quests often turn out to be too hard for the most gifted mathematicians to answer.

The theory of shape is also studied by Bookstein in morphometrics – a discipline studying the deformations and variabilities of biological organisms[3]. 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[28]. For example, given two sets of points, they compute the best transformation which matches one to the other under assumption of i.i.d. Gaussian models, and the computation is done by maximum likelihood estimate (MLE). The preconditions for these applications are: 1). the set of landmarks must be extracted from images reliably, and 2). there is a well defined correspondence between the points in the two sets.

2.2 Deformable templates – the high level vision

Another elegant theory for shape modeling was pioneered by Grenander in a discipline which he named *pattern theory* [10]. In pattern theory, shape primitives are selected, such as edge segments, and then these primitives are arranged in a pre-specified configuration, e.g. a circular graph. Then transformation groups (rotation, scaling etc.) act on the subgraphs to account for global and local deformations.

For example, the contour of a human hand is represented as a ring of linelets[11], and a global transform specifies the location, size, and orientations of the hand, and local transforms define the relative orientation of fingers, and so on. Then a probability distributions is defined on these groups. This shape modeling scheme has been used in representing leaves, and brain mapping [12].

In computer vision deformable templates was proposed by Yuille in studying shapes of human eyes, mouth, and eyebrow etc.[35]. In Yuille's templates, shapes consist of piecewise conics, and deformations are modeled by the coefficients. Other deformable models of shapes are defined in terms of basis function, such as B-splines, sine waves[5], and implicit polynomials[20] and Super quadrics[32]. A more general template model for flexible object is studied by (Zhu and Yuille 1996) in the FORMS system, in which object shapes are defined in graphs computed from medial axis, and deformation are specified by the PCA modes learned from animal shapes[37]. Other shape models are represented in Bayes networks [23, 6]. In fact, the statistical theory of shape by (Bookstein 1986, Mardia and Dryden 1989) can also be considered as shape templates with simple points being primitives.

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

As in Kendall's shape theory, non-accidental properties are also studied in perceptual organization – grouping short line segments (or edgels) into object shapes following the edge detection process.

The problem of perceptual organization was first studied by (Lowe 1985)[26]. Lowe pointed out that the goal of grouping is to identify features that are likely to have arisen from some scene property rather than reflecting some accidental arrangement in the image. Suppose that in an image line segments are identically and independently distributed according to a uniform density with respect to orientations, positions, and scales, Lowe computed the expected number E of line segments which are within the same degree of proximity, parallelism or co-linearity as in the two given line segments. The smaller E is, the less likely the two line segments are from accidental arrangement, and the larger the probability is for grouping them together.

This theory has been successfully used in the grouping and recognition of rigid objects[27][29]. However it doesn't provide a probability measure for shape, and it is unclear how to **de-correlate** multiple properties in computing the significance of non-accidental arrangement.

Similar to Lowe's theory, non-accidental properties are also studied for convexity of a sequence of line segments by (Jacobs 1992) [15]. A quantitative measure of closure of shape is given by (Elder and Zucker 1994)[8].

In the literature of perceptual grouping, attempt has not yet been made for learning and verifying probabilities models from real shapes.

2.4 Active contour models – the low level vision

In the early stage of vision computation, it is impractical to build shape templates for all objects and match these templates against the images, unless the domain of application is well specified. Thus in early vision tasks, generic shape models are found to be successful. The first important generic shape model in the literature is the active contour model (SNAKE) [18], and the internal energy in the snake model implicitly defines a prior probability distribution of 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 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 1994)[30] and it was also studied by (Williams and Jacobs 1997)[33]:

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

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

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

Firstly, it does not characterize region based information. For example, in the shape shown in figure 2.b, node A is close to node B in distance, whereas they are far away from each other along the boundary.

Secondly, the forms of the potential function are often manually designed, whereas it is desirable that the right potential function should be learned from examples.

In summary, subsections (2.4), (2.3), and (2.2) present the existing theories in the low, middle, and high level vision respectively. It is unclear how these theories can be combined for image understanding in a consistent manner, and the incompatibility

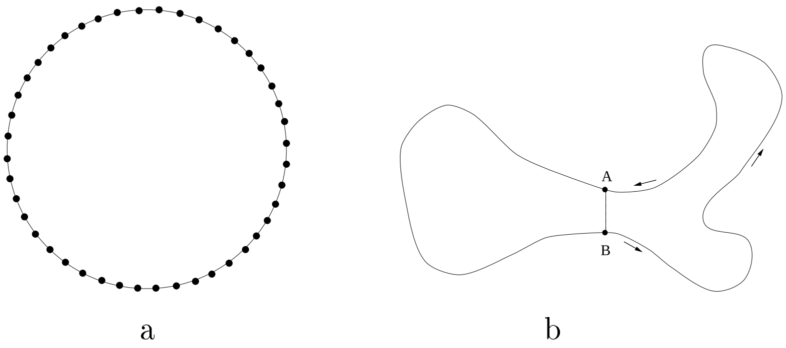


Figure 2 a). an 1D Markov random field where the nodes represent the random variables for the positions of the contour points. b). node A is spatially adjacent to point B, but is far away from B in the circular neighborhood of a).

of these models and the 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) the shape models should be generic, and thus characterizes the most common features of 2D shapes, II) the shape models should be compatible with low level representation, such as raw images and edge maps and the high level descriptions such as deformable templates.

3 Basic issues in shape modeling and learning

Let $\Gamma(s)$ be a simple, closed and non-self-intersecting contour on 2D plane[4], $s \in [0,1]$ is the arc-length and $\Gamma(0) = \Gamma(1)$. Fixing the resolution, we sample N points in $[0,1]$ at equal distance $ds = \frac{1}{N}$ and represent the curve as,

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

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

Constraint I: Γ is closed and non-self-intersecting.

Constraint II: $\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \in [ds - \epsilon, ds + \epsilon]$ $i = 0, 1, ..., (x_N, y_N) = (x_0, y_0)$.

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

In constraint II, the distance of sequential nodes is allowed to vary slightly because the node coordinates have finite precision (lattice) in computer implementation. We shall discuss the effect of ϵ when we discuss the Gibbs sampler for sampling the shape in section (7).

Ω_Γ is a subspace that contains all the valid shapes Γ in the $2N$-space, and the structures of Ω_Γ perhaps are too complicated to analysis explicitly.

On Ω_Γ the Lebesgue measure for the probability of shapes 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 will define in late sections.

To make it interesting, we impose two criteria in defining the generic probability models of shape, as our goal of shape modeling is for low/middle level vision tasks, such as general image segmentation and perceptual grouping without high level object identifications.

Firstly, the density $p(\Gamma)$ should be invariant to translation and rotation transforms, as object silhouettes are taken into images at arbitrary distances, locations and orientations. Also the shape is better to be normalized to unit length for scale invariance.

Secondly, for generic shape models, we assume that any features should have the same chance to appear in any location s in the contour. So $p(\Gamma)$ is assumed to be **homogeneous** with respect to s, and it is invariant under cyclic permutations of the labels of the points on the curve. As we shall discuss in section (10), non-homogeneous properties can be built at the high level of the hierarchy of representation.

In the rest of the paper, shapes Γ are normalized to have unit length, and the translation, rotation and cyclic permutation invariances are easily observed as $p(\Gamma)$ is defined on relative positions and orientations.

Following Kendall's treatment[19], it is best to define a "genuine" space of shape which we denote by Ω^*_Γ. Ω^*_Γ is the quotient of Ω_Γ subject to the similarity transform (2D translation and planar rotation) as well as the cyclic permutation of nodes. Each "genuine" shape in the space Ω^*_Γ is the projection of an *equivalence class* in Ω_Γ under similarity transforms and cyclic permutation. Now we need to define a probability measure $\nu(\Gamma)$ on Ω^*_Γ, which takes the sum of the Lebesgue measure of the equivalence class in Ω_Γ. If we only consider translation and cyclic permutation transforms, the

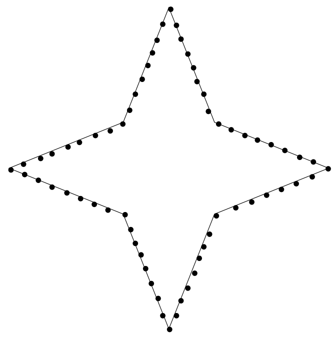


Figure 3 A N-side polygon with rotational symmetry of degree $d = 4$.

equivalence classes all have the same size, therefore, $\nu(\Gamma)$ is simply the Lebesgue measure multiplied by a constant[5]. When rotation is added, the equivalence classes may vary in size, because of rotational symmetry in shape. However, this should not be a concern in computer vision, because of the following two reasons.

Firstly, suppose Γ is rotational symmetric polygon, for example figure 3 displays a N-gon with degree $d = 4$ rotational symmetry. Suppose we discretize $[0, 2\pi]$ into $W = 1000$ angles, the equivalence class for this shape includes only 996 distinct shapes, while a non-symmetric shape has 1000 distinct shapes in its equivalence class. With precise discretization, the size variation of the equivalence classes can be negligible except for shapes like a perfect circle. However, because we are studying natural animate shapes in this paper, the probability measure for shapes which have high degree of rotational symmetry is **exponentially** rare, i.e. have measure 0. For example, none of the observed and synthesized shapes in the later section demonstrates any rotational symmetry at all!

Secondly, one can start with defining a probability measure on Ω_Γ^* directly,

$$P^*(\Gamma) = p(\Gamma)\nu(\Gamma).$$

$P^*(\Gamma)$ enjoys the invariant properties discussed above, as long as $p(\Gamma)$ is invariant to these transforms. $P^*(\Gamma)$ can be sampled by simulating Markov chain Monte Carlo sampling $P(\Gamma)$ in Ω_Γ, and then we project the samples to Ω_Γ^*. In practice, it has no difference with the samples in Ω_Γ, because it is literally impossible to hit the rotational symmetric shapes.

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

From now on, we shall assume that there is a true underlying distribution $f(\Gamma)$ on Ω_Γ (or Ω^*_Γ with no practical difference), from which we observe a set of 2D object contours in real world imagery, $\{\Gamma_i^{\rm obs},\ i = 1, 2, ..., M\}$. Due to the invariant assumptions, one may add more shapes to the observed shape set by rotating each observed shapes. Our goal is to learn a probability model $p(\Gamma)$ as an estimation of $f(\Gamma)$. In this way, we pose the problem of learning shape model as a statistical inference problem. 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.

A trivial solution for learning $p(\Gamma)$ is the non-parametric model,

$$p(\Gamma) = \frac{1}{M}\sum_{i=1}^{M} G(\Gamma - \Gamma_i^{\rm obs}). \tag{4}$$

where $G(x) = \delta(x)$ is a Dirac function with unit mass centered at 0, or a Gaussian window function on the deformations from Γ to $\Gamma_i^{\rm obs}$ when there are well defined correspondences from the points in Γ to the points in $\Gamma_i^{\rm obs}$ and principal components analysis can be conducted as in the statistical theory of shape in section (2.1)[7]. $p(\Gamma)$ in equation (4) is the mixture of deformable templates. This model has two advantages: 1. it is easy to compute if M is small, 2). it approaches the true model $f(\Gamma)$, if $M \to \infty$. However, if M is large, [6] then $p(\Gamma)$ becomes infeasible to compute. [7] On the other hand, when M is small, this non-parametric model cannot effectively predict the unobserved samples from $f(\Gamma)$.

We argue that an effective model $p(\Gamma)$ should capture the most common characteristics of natural shapes, as by the Gestalt laws. A quantitative comparison between a generic model and a non-parametric model is studied in the context of image modeling by Zhu and Mumford[38], where local generic properties such as intensity gradients are found to be more important – measured in light of an information criterion – than the overall images examples as in the non-parametric model.

[6] In the N-space, a modest number for M should be at least in the order of 10^6, otherwise the non-parametric model are too sparse to be useful in general vision problems.

[7] This is the reason why we cannot solve vision problems using template matching, although, theoretically speaking, vision can be solved by purely top-down methods.

4 Feature selection

To learn a shape model $p(\Gamma)$, we start with exploring both contour-based and region-based features on Γ. Our choices of features are focused on the most general Gestalt laws: co-linearity, co-circularity, proximity, parallelism, and symmetry. We should define a continuous measure for each of these laws. Such measures can be viewed as "shape filters" by analogy to the Gabor filters in image modeling.

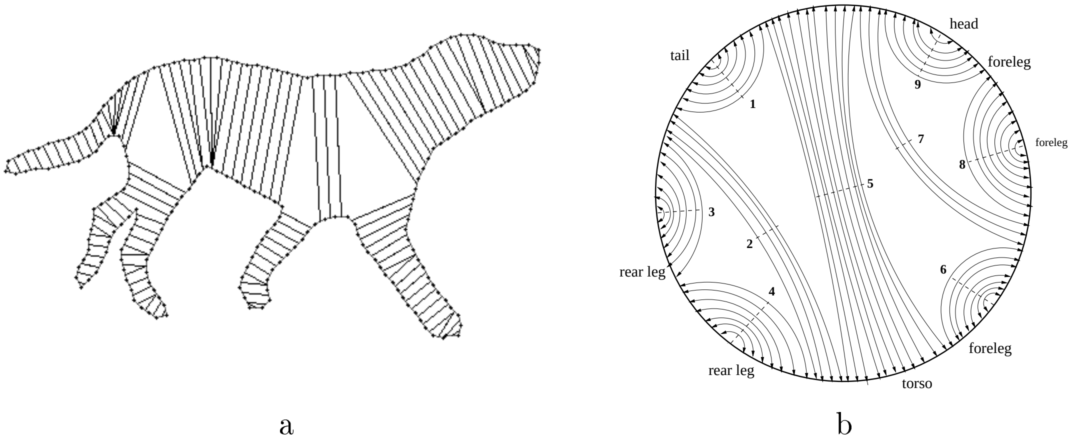


Figure 4 a. A dog shape with region based correspondence detected, b. an abstract planar *adjacency graph* representing the neighborhood structures of the linelets in the dog shape.

Figure (4).a shows a curve Γ – a dog shape, which is decomposed into a number of sequential linelets $\ell(s)$, each of which is represented by a node on the circle in figure (4).b. $\ell(s)$ has attributes $(x(s), y(s), \theta(s))$ for the center coordinates and orientation of the linelet. The circular connection defines a *random field* as in figure (2).a, where $x(s), y(s), \theta(s)$ are the random variables at each site s and the neighborhood structures of the random field is shown in figure (2).b (to be discussed later in this section).

Now we discuss how shape features are extracted from the random field.

At each location s,[8] we measure a set of functions $\phi^{(\alpha)}(s)$ with $\alpha = 1, 2, 3, ...$ being the index of the shape features. By analogy to feature extraction using filters in 2D images, $\phi^{(\alpha)}(s)$ is the "response" for a "shape filter" at location s.

we first measure contour-based functions by defining,

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

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

i.e. the curvature $\kappa(s)$ and the derivative of curvature. It is easy to see that $\phi^{(1)}(s) = 0$ means two adjacent linelets are *co-linear*, and $\phi^{(2)}(s) = 0$ means three sequential linelets are *co-circular*. Other contour-based shape filters can be defined in the same way.

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

It is well known in computer vision that real object shapes –resulting from the process of growth – have natural description in terms of the medial axis (Blum 1973, Leyton 1992, Zhu and Yuille 1996) [2, 25, 37]. For example, a dog has a skeleton and many elongated parts, and as shown in figure (4).a the linelets at the two sides of the limbs or torso are parallel or symmetric with respect to the medial axes. It is evident in psychophysics experiments that early human vision are sensitive to the medial axis(Kovacs and Julesz 1994) [22, 4], and it is also discovered in physiology experiment that some neurons in the primary visual cortex of monkeys compute the symmetric axes as soon as they detect the edge elements (Lee, Mumford and Zhu 1996)[24]

To capture the region based properties of a given curve $\Gamma(s)$, the author has defined a symmetry mapping function $\psi(s)$ as a transform (Zhu 1998) [40].

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

Figure (4) displays the mapping function ψ for the dog shape. As shown in figure (4).b, ψ divides the circular domain $[0,1]$ into a set of 18 disjoint intervals

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

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

Intuitively each pair of intervals represent an elongated part of the shape. For example, the mapping function ψ for the dog shape is shown in figure (4) has 9 pairs of intervals for the 9 parts of the dog. Note that in figure (4).a some linelets are matched more than once, this is due to the effect of discretization.

Now we briefly explain how ψ is computed.

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

1. The 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 the Markov random field, and the medial axis is computed based on ψ. Details of the definition of the energy functional and the algorithm is referred to [40].

Medial axes computed in this statistical framework have advantages over the purely geometrically defined ones[2]. Two of these advantages are important for shape modeling.

- As continuities at large scale can be propagated through local interactions in the Markov random field, the computed mapping function and medial axis are robust to small perturbations on the curve.

- As ψ is computed by random sampling based on local neighborhood, thus when the nodes on the curve move, the mapping function is updated easily based on only local computation – a crucial property for sampling the shape model $p(\Gamma)$ in later sections.

The mapping function ψ defines a planar graph structure – the new neighborhood – for nodes on shape Γ, as illustrated in the abstract graph in figure (4).b. On this graph, region based properties are extracted. Let $r(s)$ be the distance between two matched linelets $\ell(s)$ and $\ell(\psi(s))$. $r(s)$ is divided by the total arc length of the curve, thus is well normalized with respect to scale, and we call it 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, and $\phi^{(3)}(s)$ measures the proximity between two linelets across the region as we discussed in figure (2).b.

It is easy to see that all the 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 the 2D plane.

Another shape measure that one may use is $\phi^{(6)}(s)$,

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

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

5 Statistics of animate shapes

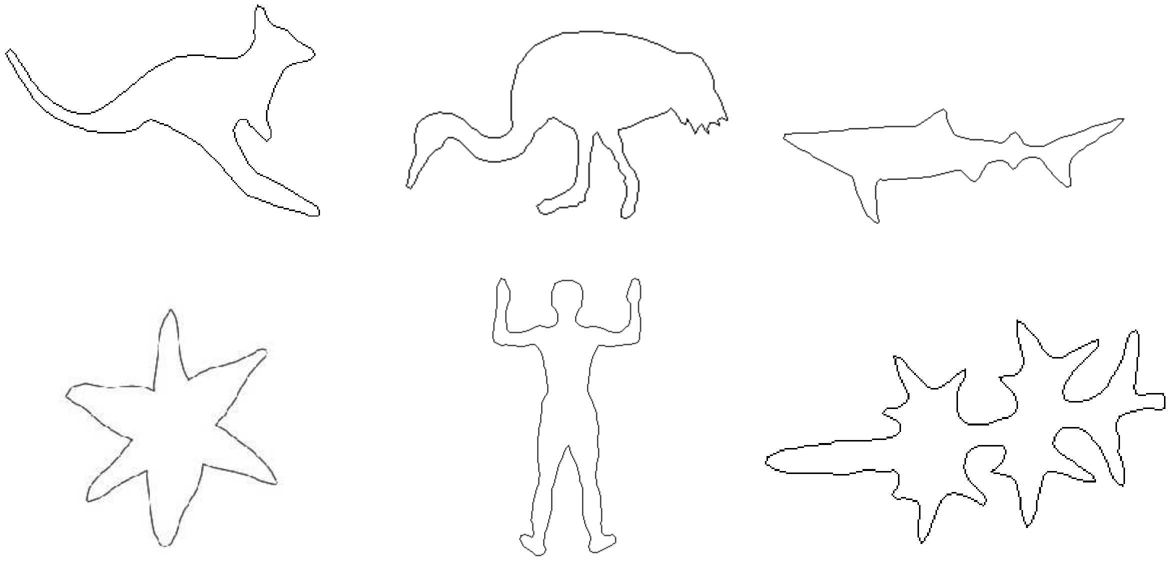


Figure 5 Examples of the observed natural shapes.

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

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 5. These shapes are viewed as independent samples from the true distribution $f(\Gamma)$ in a given domain of application. We are interested in animate shapes because these shapes are the results of the natural growth, and have richer flexibilities and variations than human made objects.

These animate shapes are observed at various resolutions with perimeter $L_i, i = 1, 2, ..., 22$ pixels in the image lattice. A shape observed at a high resolution contains rich information, and thus 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 the 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 the features, such as $\kappa(s)$, cannot be computed reliably. A good compromise is to make c as big as possible while still keeping the curve continuous to human perception. We choose $c = 6$ pixels, as shown in figure 5, there is no noticeable artifacts of polygon approximation. The arc length for each linelet in Γ_i^{obs} is $ds = \frac{1}{N_i}$, therefore the shape is normalized to have the same arc-length.

In this paper, the statistics of a given observed shape Γ_i^{obs} and a given feature in Φ is extracted as the empirical histograms. By the 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 $x \neq 0$. In 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 the average histograms for the M discretized curves

$$\mu_{\text{obs}}^{(\alpha)}(z) = \frac{1}{N_1 + N_2 + ... + N_M} \sum_{i=1}^{M} N_i H^{(\alpha)}(\Gamma_i^{\text{obs}}, z), \quad \alpha = 1, 2, ..., 6. \tag{5}$$

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 $\mu_{\text{obs}}^{(\alpha)}(z) = \mu^{(\alpha)}(z), \alpha = 1, 2, ..., 6$.

Remarks: these one dimensional distributions include mean, variance as well as high order moments.

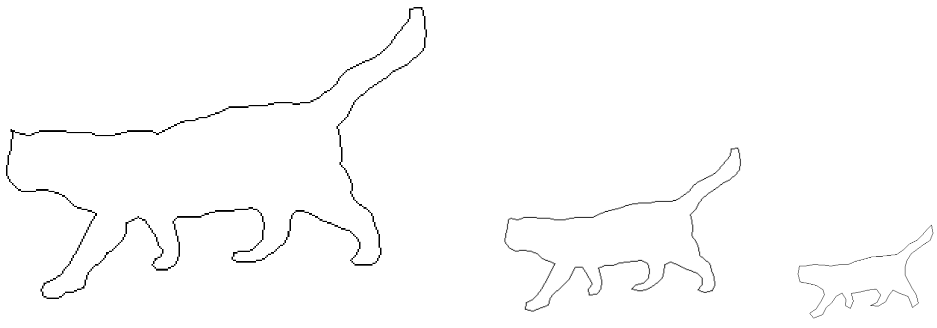


Figure 6 The cat shapes at scale 0,1,2 with $N, N/2, N/4$ nodes respectively.

To show how the observed histograms change with the scales of shape, we sub-sample each observed shape once and twice, and generate two new sets of observed shapes at scale 1 and scale 2. The Figure 6 displays the cat shapes at three scales. The shape at scale $i+1$ has only one half the nodes of the shape at scale i.

Figure 7.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 7.b plots the logarithm of the curves in figure 7.a.

Figure (7) demonstrates two interesting properties.

Firstly, the middle part of $\mu_{\text{obs}}^{(1)}$ is close to an exponential curve (see the dashed curve in figure 8.a), but $\mu_{\text{obs}}^{(1)}$ has heavier tails.

Secondly, unlike the scale invariant properties found in the image intensity of natural images [38], 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 the high curvature portions than at the low curvature portions. As a result, the curvature histograms at scale $i+1$ has lighter tails and sharper peak near zero than scale i. We conjecture that the observed histogram μ_{obs} should be scale-invariant, if the sub-sampling artifact could be eliminated.

The average histogram $\mu_{\text{obs}}^{(2)}$ of $\phi^{(2)}$ is very close to the curvature histogram. Therefore, we choose $\mu_{\text{obs}}^{(1)} = \mu_{\text{obs}}^{(2)}$ to be the observed curvature histogram averaged over scale 0 and scale 1 for robustness. $\mu_{\text{obs}}^{(\alpha)}, \alpha = 1, 2$ is plotted in figure 8.a, and its logarithm is plot in figure 8.b.

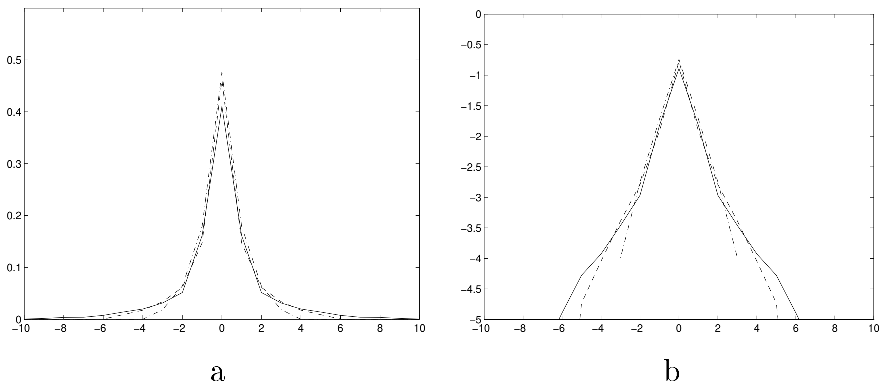


Figure 7 a). the histograms of $\kappa(s)$ averaged over 22 animate objects at scale 0 (solid curve), scale 1 (dashed curve) and scale 2 (dash-dotted curve), the horizontal axis is $\kappa(s)$ with unit $dz = \frac{\pi}{16 \times 200}$ b). the logarithm of curves in a).

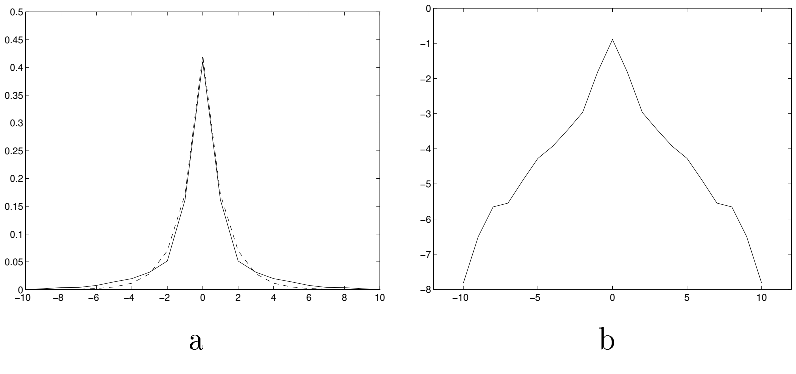
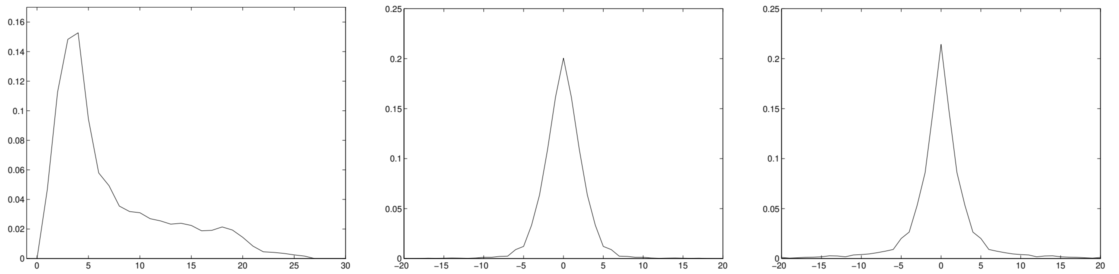


Figure 8 a). the averaged histogram for both $\kappa(s)$ and $\nabla\kappa(s)$ (solid curve), and the fitting curve $e^{-0.9|\kappa(s)|}$ (dashed curve) , b). the logarithm of the solid curve in a). 0.9 is obtained by least square fitting of b) to two straight lines.

We also compute the histograms for the region based shape filters, $\phi^{(3)}(s) = r(s), \phi^{(4)}(s) = \nabla r(s), \phi^{(5)}(s) = \nabla^2 r(s)$, and their histograms $\mu_{\text{obs}}^{(\alpha)}, \alpha = 3, 4, 5$ are shown in figure 9.a, b, c respectively.

The 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 the important correlations and proximity of points across the region – such as the points at the two sides of the animal limbs.



a b c

Figure 9 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 $\{\mu_{\text{obs}}^{(\alpha)}(z)\}$ which are marginal distributions of $f(\Gamma)$, 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,

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

Any model $p(\Gamma) \in \Omega_p$ will be equally good, as long as the selected features and their statistics are concerned. From Ω_p we choose the one which has the maximum entropy (ME), i.e.,

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

subject to the constraint equations,

$$\int p(\Gamma) d\Gamma = 1 \tag{6}$$

$$\int p(\Gamma)\delta(z - \phi^{(\alpha)}(s)) d\Gamma = \mu^{(\alpha)}_{\text{obs}}(z) \quad \forall s, \forall z, \forall \alpha. \tag{7}$$

The ME principle was originally proposed by (Jaynes 1957)[16] in statistical mechanics. When a massive physical system reaches equilibrium, its micro-states are subject to a maximum entropy distribution. Since among all distributions in Ω_p, $p^*(\Gamma)$ has the highest entropy or randomness in the unconstrained dimensions, it provides an unbiased estimation.

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

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

In equation (8), 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 the Lagrange multipliers. Since the constraints are imposed for continuous variables z and s, $\lambda^{(\alpha)}()$ is a continuous function. Due to homogeneity assumption $\lambda^{(\alpha)}()$ is independent of s, and therefore it is one dimensional.

We can rewrite equation (8) as,

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{- \sum_{\alpha=1}^{k} \int \lambda(\phi^{(\alpha)}(s)) ds\}. \tag{9}$$

where $\lambda^{(\alpha)}(), \alpha = 1, 2, ..., k$ is called the potential function, or

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{- \sum_{\alpha=1}^{k} \int \lambda(z) H^{(\alpha)}(\Gamma, z) dz\}. \tag{10}$$

It is easy to see that the active contour models in equation (1) and equation (2) are special examples of equation (9) with Φ and Λ specified.

6.2 Estimation and computation

In model $p(\Gamma; \Phi, \Lambda)$, the Lagrange multipliers or the forms of potential functions have yet to be solved from the constraint equations (6) 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 [39][38], we describe it briefly in this section.

In practice, the histogram $H^{(\alpha)}(\Gamma, z)$ is discretized into m bins, and it is represented by a vector $H^{(\alpha)}(\Gamma) = (H_1^{(\alpha)}, H_2^{(\alpha)}, ..., H_m^{(\alpha)})$. Therefore $\lambda^{(\alpha)}()$ is approximated by a piecewisely constant function, i.e. it is represented by a vector $\lambda^{(\alpha)} = (\lambda_1^{(\alpha)}, \lambda_2^{(\alpha)}, ..., \lambda_m^{(\alpha)})$ in a non-parametric form. In general the sizes of the bins may vary, and m doesn't have to be the same for different features. If m is large enough, $\lambda^{(\alpha)}$ is a close approximation to $\lambda^{(\alpha)}()$.

Thus equation (10) becomes,

$$p(\Gamma; \Phi, \Lambda) = \frac{1}{Z} \exp\{-\sum_{\alpha=1}^{k} < \lambda^{(\alpha)}, H^{(\alpha)}(\Gamma) >\} \tag{11}$$

where $<,>$ is the inner product of two vectors.

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,

$$\frac{d\lambda^{(\alpha)}}{dt} = E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}(\Gamma)] - \mu_{\text{obs}}^{(\alpha)}, \quad \alpha = 1, 2, ..., k, \tag{12}$$

where t is the time steps, 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 process 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 [39], provided that the constraint equations are consistent.

In the above equation, for a given Λ, $E_{p(\Gamma;\Phi,\Lambda)}[H^{(\alpha)}(\Gamma)]$ is hard to compute analytically. In general, one needs to simulate a stochastic process by Markov Chain Monte Carlo (MCMC), and this Markov chain walks randomly in the shape space Ω_Γ. When this stochastic process becomes stationary, it samples the distribution $p(\Gamma; \Phi, \Lambda)$, and we shall discuss the design of the MCMC in late 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 (5).

So Λ is updated by

$$\frac{d\lambda^{(\alpha)}}{dt} = \mu^{(\alpha)}_{\mathrm{syn}} - \mu^{(\alpha)}_{\mathrm{obs}}, \quad \alpha = 1, 2, ..., k. \tag{13}$$

The whole learning process simulates an inhomogeneous Markov chain[34], and this computational scheme has been successfully applied to texture modeling by Zhu, Wu and Mumford [39].

6.3 *The learnt shape model*

Suppose we adopt the six features as discussed in section (4), we derive the following shape model,

$$p(\Gamma) = \frac{1}{Z}\exp\{-\int \lambda^{(1)}(\kappa(s))+\lambda^{(2)}(\nabla\kappa(s))+\lambda^{(3)}(r(s))+\lambda^{(4)}(\nabla r(s))+\lambda^{(5)}(\nabla^2 r(s))ds+\gamma\|B\|\}. \tag{14}$$

In equation (14), B is the set of breaking points– discontinuities in the symmetry mapping function $\psi(s)$, and $\|B\|$ is the number of breaking points. The number of branches of a shape is $\|B\|/4$. The model in equation (14) has some desirable properties.

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

2. Features are weighted by the potential functions. Instead of being manually specified, these potential functions are learned so that $p(\Gamma)$ can reproduce the observed statistics, and correlations between features are taken into account in the learning process, as shown in later section.

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.

In summary, the above shape model provides a unifying framework for the three theories–Gestalt laws, likelihood principle, and minimum description length principle–discussed in section (1).

7 Design the sampling process by MCMC

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

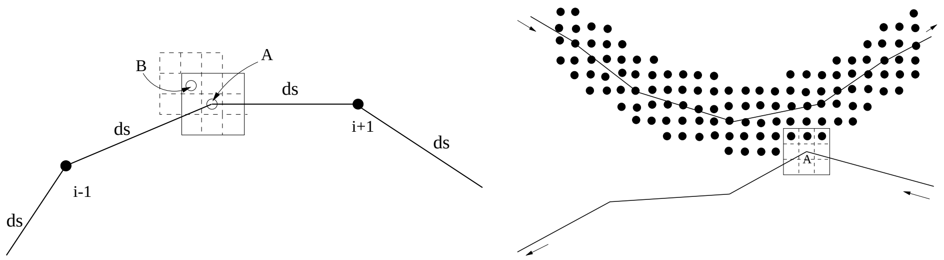
Designing the sampling process is important for the following reasons. Firstly, as shown in equation (13), the sampling process is a necessary step for computing the potential functions Λ. Secondly, the most obvious way to verify a learned model is to sample the model. If the shape model is close to the underlying truth, then the typical set of samples from the model should be similar to the observed shapes judged by human perception. Thirdly, the learned shape model will eventually be used as prior distribution in tasks such as computing object shapes from input images by maximizing *a posteriori* (MAP) estimation, and this sampling process is the engine for shape inference.

In section (3), shape $\Gamma = (\,(x_0, y_0), (x_1, y_1), ..., (x_{N-1}, y_{N-1})\,)$ is defined on Ω_Γ. In computer implementation, the cruel reality is that the nodes $(x_i, y_i), = 0, 1, 2, ...$ can only be described in a finite precision, and we denote by τ the unit length of the lattice [9]. We design the sampling process in the following way. One starts with a simple closed curve, such as a circle, or rectangle with N nodes. At each step, one 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 the Metropolis algorithm, and it simulates a stochastic process –random walk in the shape space. If one samples long enough, this stochastic process converges to its equilibrium, and becomes a stationary process. 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 $p(\Gamma; \Phi, \Lambda)$.

There are two immediate questions that we have to address before we discuss the algorithm in details. Firstly, what is the effect of the ϵ in the shape constraint? Secondly, how do we diagnose if the random walk becomes stationary? It is hopeless to have a precise answer to the above questions due to the complexity of the shape space Ω_Γ. For question one, 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

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

the choice of the ratio $\frac{\epsilon}{ds}$ should also depend on the precision of human perception. In our experimentation, we select $\epsilon = 2\tau$ and $ds = 10\tau$. There is no noticeable difference when we use $\epsilon = 2\tau$, and $ds = 20\tau$ except that the convergence becomes slower. There is no mathematical 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 meanwhile we monitor the sampled shapes by human 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.



a b

Figure 10 a. The proposal for moving a node from position A to B. b). The firewall preventing the curve from self-intersecting itself. See text for interpretations.

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 10.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 violates the two hard constraints. Now we **propose** to move A to one of the N_A positions B at equal chance $K(A \to B) = \frac{1}{N_A}$. Then 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,

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

In the above equation, $p(\Gamma_A; \Phi, \Lambda)$ and $p(\Gamma_B; \Phi, \Lambda)$ are, due to Markov property, computed as $p((x_i, y_i)| \cup_{j \in \partial i} \{(x_j, y_j)\})$ based on nodes in the local neighborhood set ∂i of node i. For example, if $\phi^{(1)} = \kappa(s)$ is the only feature chosen in the model, then the computation of (x_i, y_i) will only involves (x_{i-1}, y_{i-1}) and (x_{i+1}, y_{i+1}). If region based properties like $\phi^{(3)}(s)$ is 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[10], and $r(s)$ at node i is recomputed for estimating $p(\Gamma_B; \Phi, \Lambda)$.

In summary, the transpose probability $P(A \to B) = K(A \to B)\alpha(A \to B)$. Therefore 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 $p(\Gamma; \Phi, \Lambda)$. Detailed discussion of the MCMC method is referred to [34].

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.
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.

[10] As the mapping function $\psi(s)$ is computed in a Markov random field, thus updating $\psi(s)$ only involves local computations [40].

In our experiment, to prevent a curve from intersecting itself, we construct a "firewall" of width 5τ along the curve, as shown by the dots in figure 10.b. The cells along the line segment from node i to node $i+1$, displayed by the dots in figure 10.b, are labeled as i, and a node j cannot move into the firewall cell labeled i if $|j-i| \, mod \, N \geq 2$. For example, node A in figure 10 cannot move to the three dotted cells, though they are within the square– the neighborhood of A.

Our first experiment is to sample the uniform distribution, we sampled a group of 100 shapes after 1.5 million sweeps with $N = 200$, $ds = 10\tau, \epsilon = 2\tau$. Figure 11 displays four of the sampled shapes. To study the effect of the resolution N in the synthesized shapes, we sample another group of 100 shapes with $N = 400$ nodes, four of which are displayed in figure 12. Then we subsample these 400-node shapes by discarding the odd number nodes, and 4 of the resulting 200-node shapes are displayed in figure 13. Perceptually it seems that the shapes in figure 11 and the shapes in figure 13 are sampled from the same stochastic process–without noticeable difference. The quantitative comparison of their curvature histograms is given in figure 14.

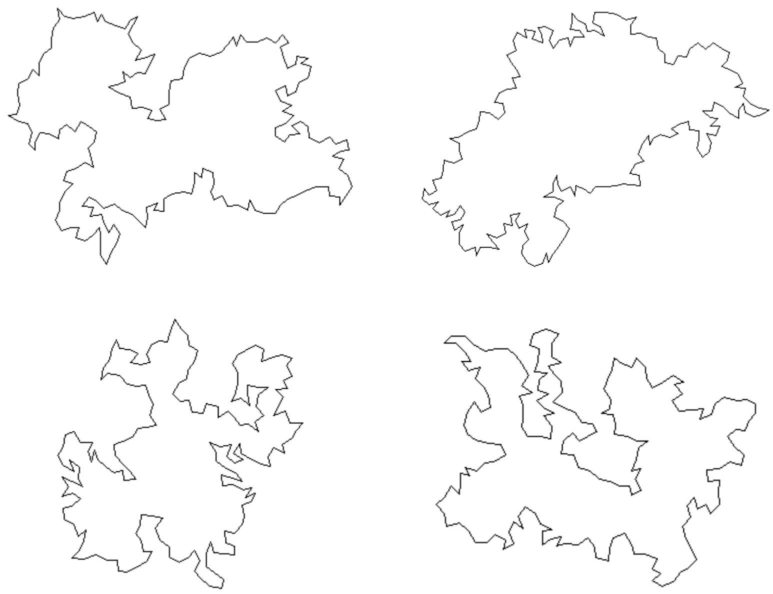


Figure 11 4 of sampled shapes recorded at different time steps of the stochastic process designed for uniform distribution. $N = 200$, $ds = 10\tau, \epsilon = 2\tau$. Note that one of these shapes is displayed in figure 1.a.

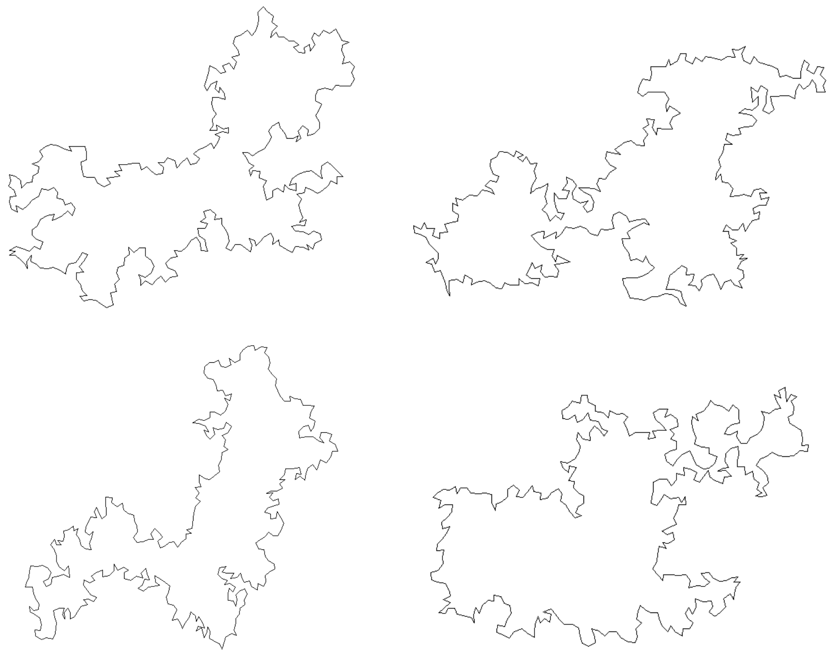


Figure 12 Shapes of uniform distribution with 400 nodes.

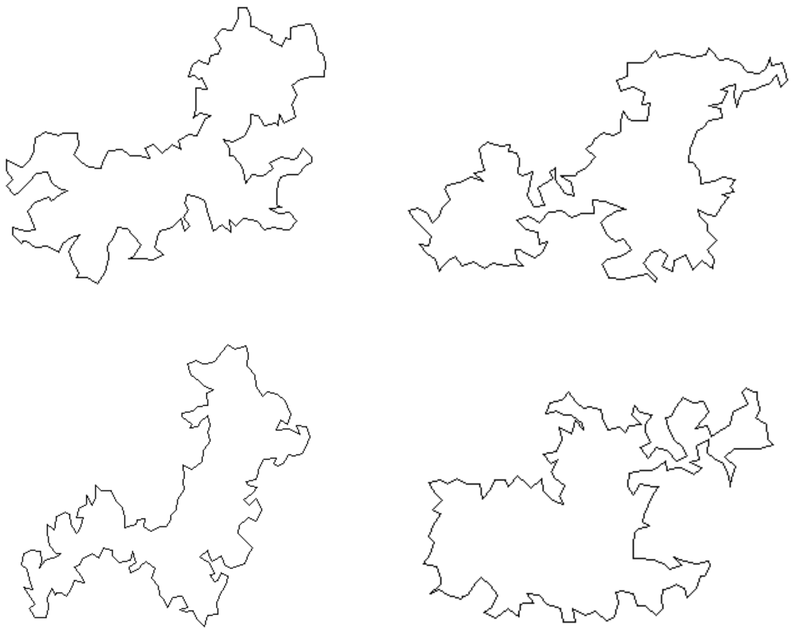


Figure 13 Shapes of 200 nodes resulted from sub-sampling the shapes of 400 nodes, in comparison with the samples of 200 nodes before.

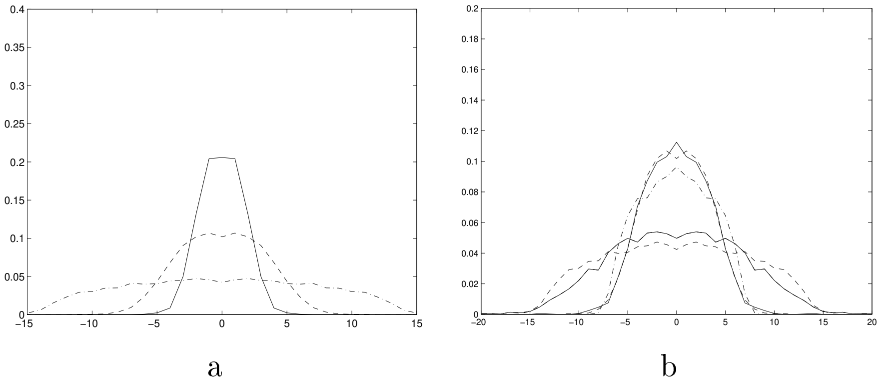


Figure 14 a). Histograms of $\kappa(s)$ averaged over 100 samples at 3 scales $H_{200}^{(0)}$ (dash-dotted), $H_{200}^{(1)}$ (dashed) and $H_{200}^{(2)}$ (solid) respectively. b). Two broad histograms of $\kappa(s)$ are $H_{200}^{(0)}$ (dashed) and $H_{400}^{(1)}$ (solid) respectively, and the three sharper histograms are $H_{100}^{(0)}$ (dash-dotted), $H_{200}^{(1)}$ (dashed) and $H_{400}^{(2)}$ (solid) respectively. See text for explanation.

In figure 14, each curve is the histogram $H_N^{(m)}$ of the curvature $\phi^{(1)}(s) = \kappa(s) = \frac{d\theta}{ds}$ averaged over 100 synthesized shapes, where the subscript N is the number of nodes using in the sampling process, and the superscript (m) denotes the scale –the times of sub-sampling. For example, figure 14.a shows $H_{200}^{(m)}, m = 0, 1, 2$ by the dash-dotted, dashed, and solid curve respectively. That is, as we subsample the shapes once and twice (obtaining curves of 100 nodes, and 50 nodes respectively), the histogram becomes sharper and sharper. The histogram changes over scales, because the total curve length is normalized with $ds = \frac{1}{N}$. Therefore, with higher resolutions, the local structure of the curve of uniform distribution becomes richer and richer, and present fractal properties. This scale-variability is in sharp contrast to the pseudo-scale-invariant property of natural shapes in figure 7

Figure 14.b shows two groups of histograms. 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 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 process with N=100, 200, and 400.

8 Feature pursuit by maximizing non-accidental statistics

In this section, we discuss how to select important features for shape modeling in a stepwise manner.

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 learnt 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 observed shapes respectively. We notice that $\mu_{\text{syn}}^{(\beta)}$ is the statistics of *accidental arrangement* for feature $\phi^{(\beta)}$, and it accounts for the correlation between $\phi^{(\beta)}$ and the feature set 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 L_1 norm, or the quadratic form–Mahanalobis distance.

In modeling texture[39], 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 above 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 of various orientations and scales. So we choose not to discuss the feature selection issue explicitly, and the proof for the following proposition is referred to our texture paper[39].

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

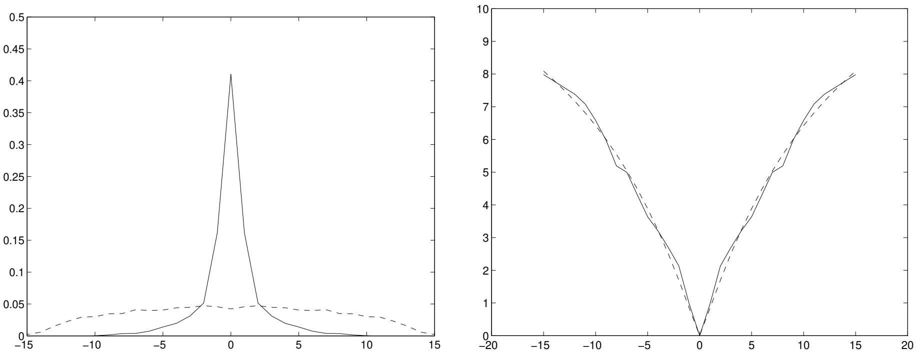


Figure 15 a). The $\mu_{\rm obs}^{(1)}$ (solid) versus $\mu_{\rm syn}^{(1)}$ of the uniform distribution for the chosen feature $\phi^{(1)}(s)$. These two curves have appeared in figure 8.a and figure 14.a respectively. b). The learned $\lambda^{(1)}()$ for $\phi^{(1)}(s) = \kappa(s)$ in a non-parametric 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 $\kappa(s)$.

We start with $\Phi = \emptyset$ and $p(\Gamma; \Phi, \Lambda)$ a uniform distribution. We compare $\mu_{\rm syn}^{(1)}$(dashed) of the uniform distribution against $\mu_{\rm obs}^{(1)}$(solid) in figure 15.a. $\mu_{\rm obs}^{(1)}$

has a much sharper peak than $\mu^{(1)}_{\text{syn}}$ near zero, it measures quantitatively how much co-linearity natural shapes have in comparison with the co-linearity resulted from **accidental arrangement** in the uniform shapes while the hard constraints are taken into account. We found that $\|\mu^{(1)}_{\text{syn}} - \mu^{(1)}_{\text{obs}}\| > \|\mu^{(2)}_{\text{syn}} - \mu^{(2)}_{\text{obs}}\|$, 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.

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 15.b. $\lambda^{(1)}(z)$ is close to $|z|$ near zero, but it has flat tails to preserve large curvatures. Figure 16 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 17.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 17.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 very precisely. $\mu^{(2)}_{\text{syn}}$ is plot against $\mu^{(2)}_{\text{obs}}$ in figure 17.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 18. These shape have more jagged boundary, and have less structures than the shapes in figure 16.

Experiment II

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 19.a.b respectively. Six of the sampled shapes are shown in figure 20. The average his-

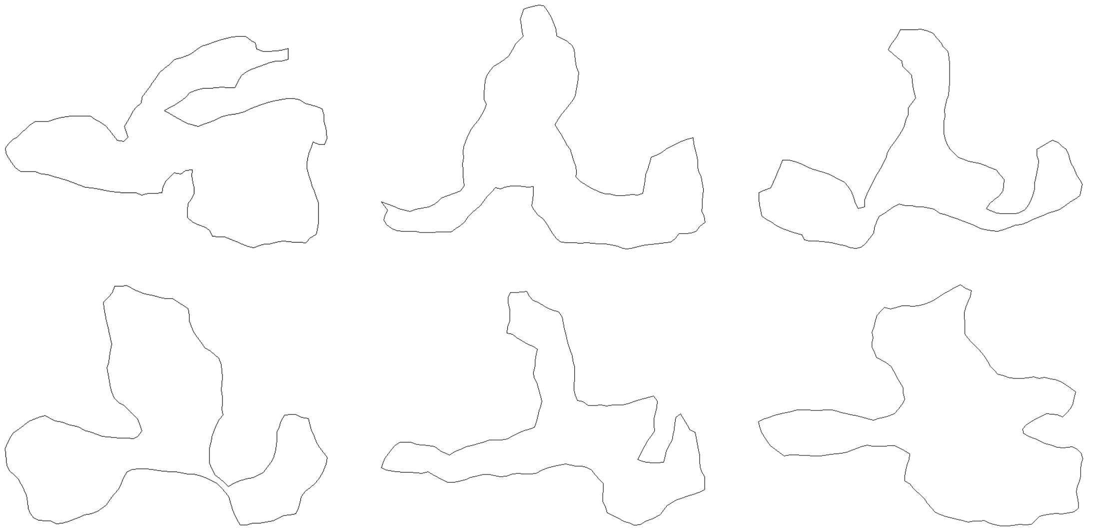


Figure 16 Six of the synthesized shapes with $\mu^{(1)}_{\mathsf{syn}} = \mu^{(1)}_{\mathsf{obs}}$. Another random shape of this set is displayed in figure 1.b. The histograms of these synthesized shapes are shown by the dashed curves in figure 17.

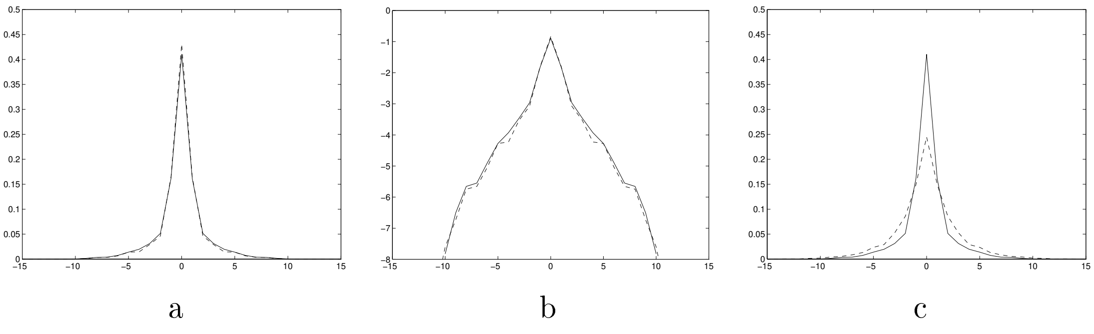


Figure 17 a). $\mu^{(1)}_{\mathsf{syn}}$ (dashed) vs. $\mu^{(1)}_{\mathsf{obs}}$ (solid). b) $\log \mu^{(1)}_{\mathsf{syn}}$ (dashed) vs. $\log \mu^{(1)}_{\mathsf{obs}}$ (solid). c). $\mu^{(2)}_{\mathsf{syn}}$ (dashed) vs. $\mu^{(2)}_{\mathsf{obs}}$ (solid).

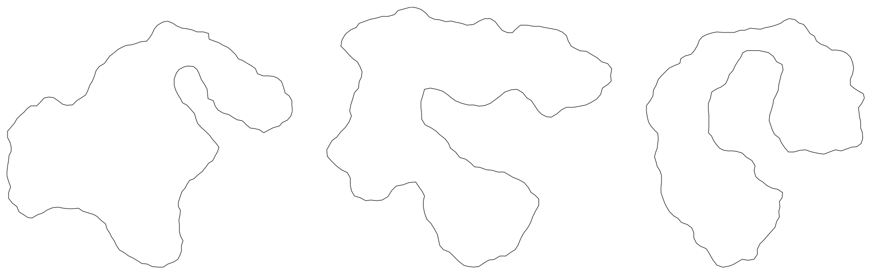


Figure 18 Sampled shapes from the Gaussian model after 2 million sweeps. The variance of $\kappa(s)$ for the animal shape is 0.1665 , the sampled shapes have variance 0.1657, which is a very close match.

tograms $\mu^{(\alpha)}_{\text{syn}}, \alpha = 1, 2$ for the synthesized shapes are shown in figure 21.a and b by the dashed curves, and they match $\mu^{(\alpha)}_{\text{obs}}, \alpha = 1, 2$ closely.

In summary, the shapes in figure 20 share the same amount of co-linearity and co-circularity as in the observed animate shapes. These shapes have very smooth contour as piecewisely continuous spline curves, and we also notice that long circular arcs are formed through the propagation of local interactions in the Markov random field.

We want to emphasize two issues on the results of this experiment.

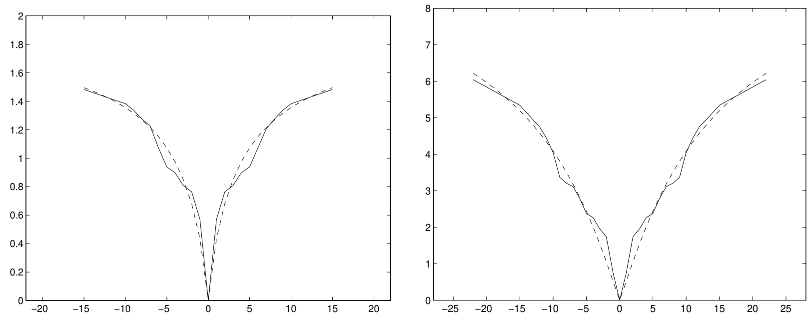
- The synthesized shapes have pseudo scale invariant property as in the observed shape. For example, we collect 45 sampled shapes, and subsample them once and twice, and generate two new sets of shapes at scale 1 and scale 2. Figure 21.c shows the average curvature histograms for the three scales. This indicates that *the choice of N is not critical 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 19.a is very different from that in figure 15.b. This change reflects the correlation between the two features. This demonstrates that the learning process could account for inter-dependency between the chosen features.

Experiment 3

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

Although the shapes in figure 20 reproduce the exact amount of co-linearity and co-circularity as in animate objects, these shapes are very blob-like, and elongated



a b

Figure 19 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 $a = 10, b = 14, \gamma = 1.1$.

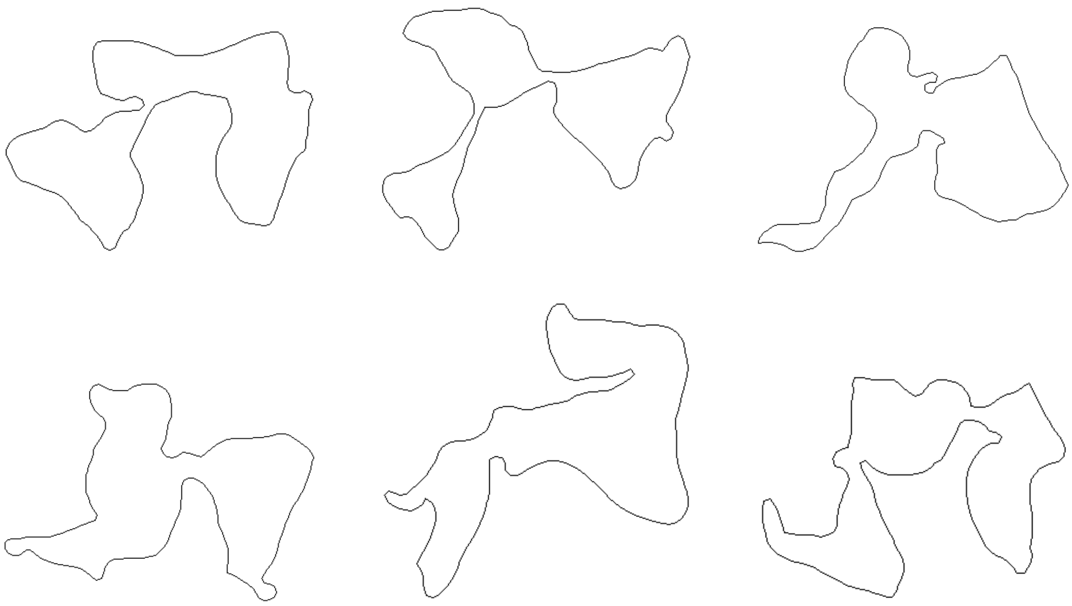
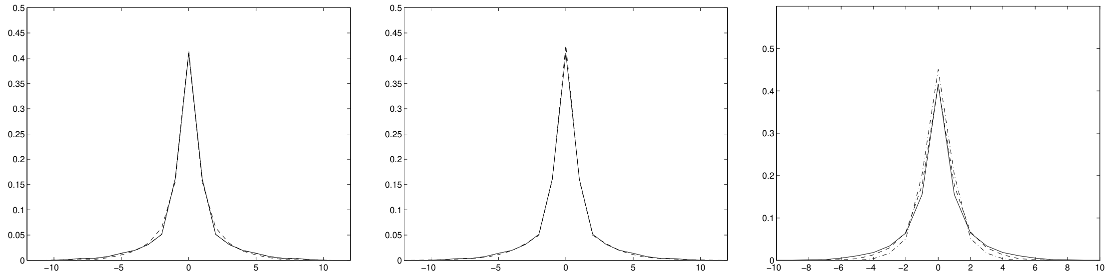


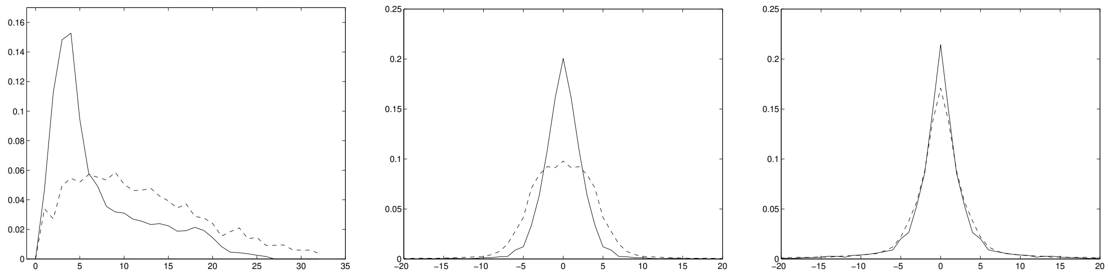
Figure 20 Six of the synthesized shapes with $\mu^{(\alpha)}_{\mathsf{syn}} = \mu^{(\alpha)}_{\mathsf{obs}}, \alpha = 1, 2$. Note that these shapes are smoother than the shapes in figure 16.



a b c

Figure 21 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).

parts, such as the limb of animals, are missing. The lack of region based features are reflected in the histograms plotted in figure 22. 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 $p(\Gamma; \Phi_2, \Lambda_2)$. $\mu^{(\alpha)}_{syn}, \alpha = 3, 4, 5$ is plot as dashed curves in figure 22.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 properties in natural shapes.



a b c

Figure 22 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 $\mu^{(5)}_{\text{syn}}$ and $\mu^{(5)}_{\text{obs}}$ is small, we choose two new features $r(s), \nabla r(s)$, and learned a new model,

$$p_3(\Gamma; \Phi_3, \Lambda_3) = \frac{1}{Z} \exp\{-\int \lambda^{(1)}(\kappa(s)) + \lambda^{(2)}(\nabla\kappa(s)) + \lambda^{(3)}(r(s)) + \lambda^{(4)}(\nabla r(s)) ds\}.$$

The learned $\lambda^{(\alpha)}, \alpha = 1, 2, 3, 4$ are shown in figure 23. Six of the sampled shapes are displayed in figure 24 together with their mapping functions $\psi(s)$.

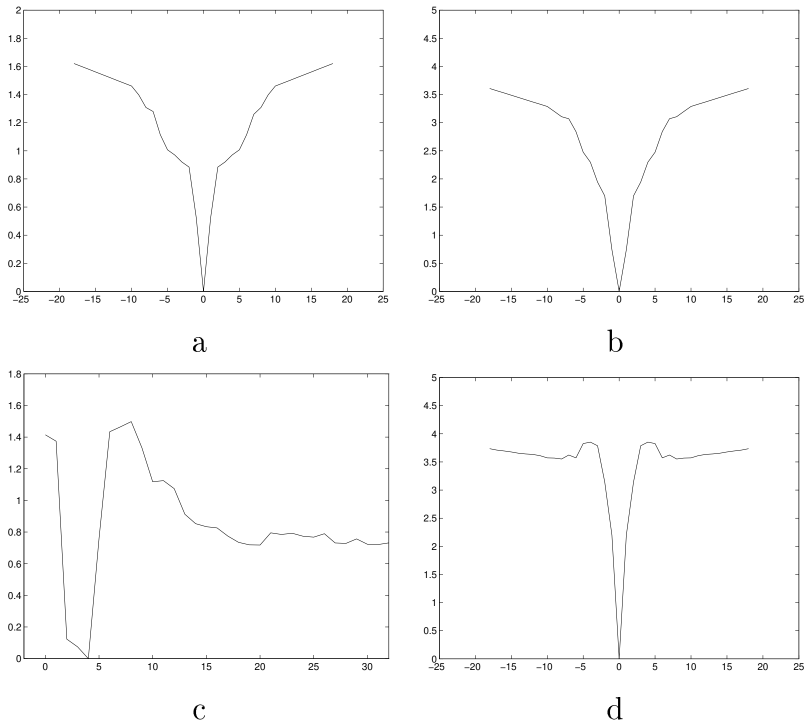


Figure 23 The learned potential functions in $p_3(\Gamma; \Phi_3, \Lambda_3)$. a). $\lambda^{(1)}(z)$, b). $\lambda^{(2)}(z)$, c). $\lambda^{(3)}(z)$, d). $\lambda^{(4)}(z)$.

This experiment demonstrates that the region based features are crucial for shape modeling.

Although none of them are identifiable as specific objects in our environments, the sampled shape resemble parts of real world objects – this is exactly the goal of middle level vision. We shall debate on this issue in the next section.

10 Discussion

In this paper, a novel 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 the quantitative measures for the non-accidental arrangement by the distance between the observed statistics and the statistics of the random shapes. We find that maximizing the non-accidental

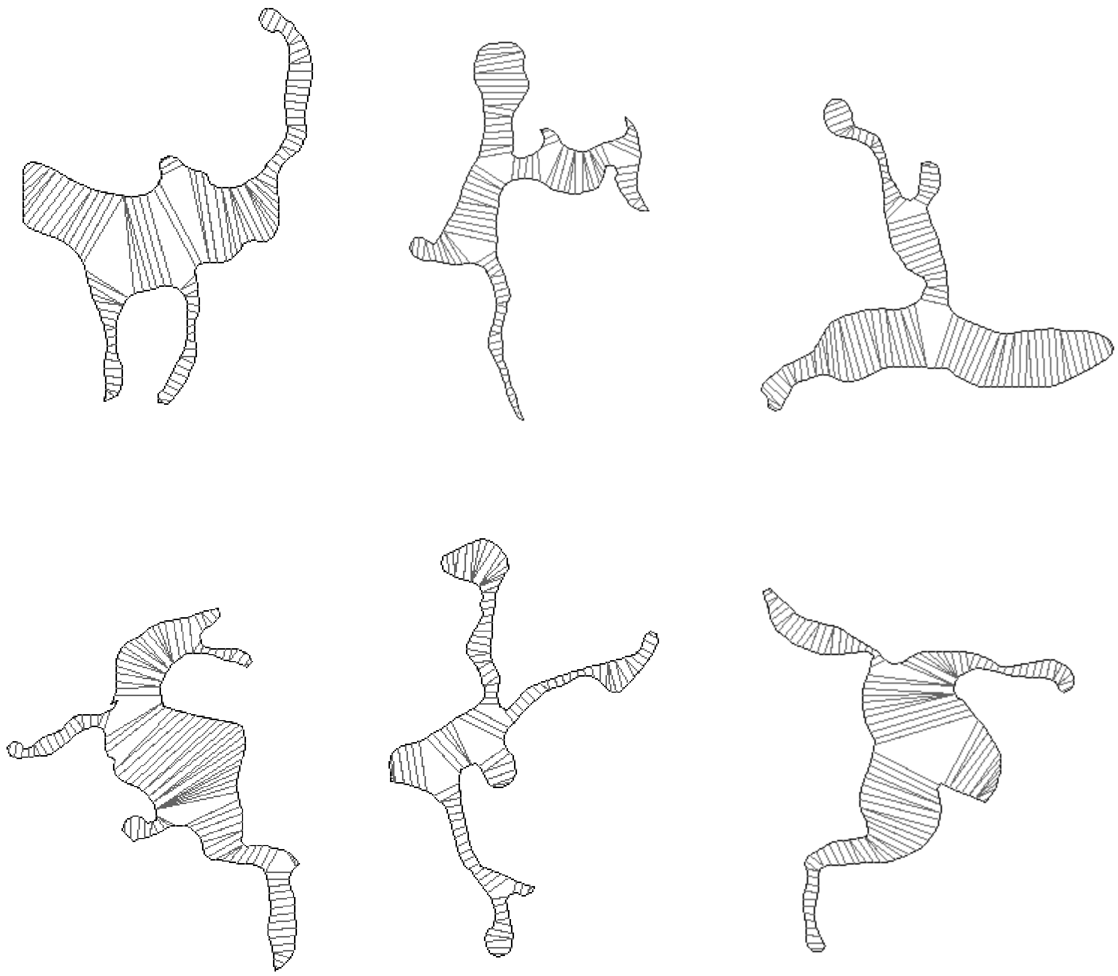


Figure 24 Six of the synthesized shapes from $p_3(\Gamma; \Phi_3, \Lambda_3)$. Another random shape of this set is displayed in figure 1.c.

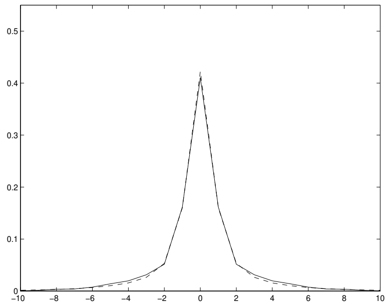


Figure 25 The solid line is the average histogram of curvature $\kappa(s)$ for 22 animals – the dataset used in this paper. The dashed line is the average histogram of $\kappa(s)$ for 1100 fish– a different dataset.

statistics is equal to minimizing the Kullback-Leibler distance between the true distribution and the model. The differences between $\mu^{(\alpha)}_{\text{syn}}$ and $\mu^{(\alpha)}_{\text{obs}}$ are evidences for the ecologic reasons underlying the Gestalt laws –co-linearity, co-circularity, proximity, parallelism, and symmetry – identified in experiment of human visual perception. Thus the article provides a firm mathematical framework for Gestalt psychology.

The learned models are limited in a few aspects.

Firstly, they are biased by the training shapes that we selected. We have studied a different database of shapes – 1100 fish contours, and the statistics in this dataset are found to be very similar to those of the animate objects used in our experiments. Figure 25 plots the curvature histogram averaged over 22 animals (the dataset used in this paper) against the curvature histogram averaged over 1100 fish contours. It is not very surprising to have similar histograms 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, if our shape ensemble consists of human made objects only, such as polygons, obviously it is very likely to derive different models from this framework. Thus our goal here is to learn shape models for a certain data ensemble.

Secondly, the neighborhood structures of our models are limited by the Gestalt laws. Ideally, one should pursue the best set of features for shape perception. However, there is very little 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.

Thirdly, the models do not account for high level shape properties. For example, some portions of the sampled shapes in figure 24 may resemble parts of animals, such as legs, tails, and heads, but these parts are not assembled in the 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 24 by connecting the centers of the mapping line segments, and a detailed algorithm is referred to a companion paper[40]. The skeleton forms a graph representation of shape. For example, figure 26 shows a shape model for high level recognition in the FORMS system proposed by Zhu etc. [37]. 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.

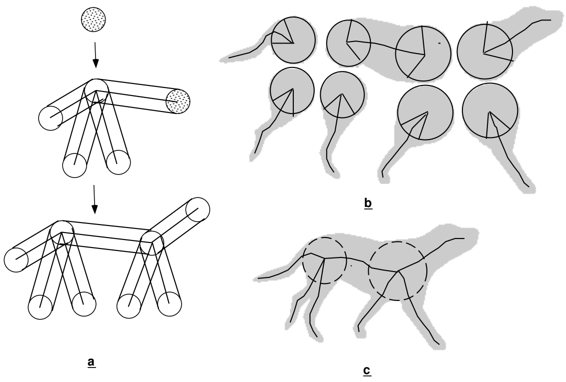


Figure 26 The general deformable model for describing a dog in the FORMS system (Zhu and Yuille 1996). a). the grammar for generating the graph structure of an animal, b). the deformation of parts, c). composing the shape through the hinge joints.

We argue that one should not impose such specific shapes structures in the lower and middle level models. The essential goal of middle level vision and perceptual organization is to bridge the gap between lower level vision algorithms, such as edge detection, image segmentation, and the high level algorithms for object recognition. Therefore the models in the middle level should be **compatible** with both the low and high level representations. We argue that the Markov random field shape models in this paper can be easily incorporated into the statistical framework for image segmentation[36], and 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 old Gestalt theory. We wish that this paper could stimulate better research results along this direction.

Acknowledgment: The author is supported by a grant from the ARO Center for Image Sciences DAAH-04-95-1-0494, and a research gift donated by the Microsoft Research. The author is grateful to David Mumford, Yingnian Wu for discussions over the period when this work was developed. The author'd like to thank for the anonymous reviewers for comments that leads to improving the presentation of the paper.

References

[1] R. Balboa, Ph.D. Thesis. Department of Computer Science, University of Alicante, Spain, 1997.

[2] H. Blum, "Biological shape and visual science". *J. of Theoretical Biology*, 33, pp205-287, 1973.

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

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

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

[6] 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.

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

[8] J. Elder and S. Zucker, "A measure of closure", Technical Report, Center for Intelligent Machines, MaGill Univ. 1994.

[9] J. Feldman, "Regularity-based perceptual grouping", TR-21, Center for Cognitive Science, Rutgers University, New Brunswick, New Jersey.

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

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

[12] U. Grenander and M. Miller, *Pattern Theory: From Representation to Inference*, 1998 (in press).

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

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

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

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

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

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

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

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

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

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

[23] D. J. Kriegman and T. O. Binford, "Generic models for robot navigation", *DARPA Proc. of Image Understanding Workshop*, Morgan-Kaufman, 1988.

[24] 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.

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

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

[27] 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.

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

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

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

[31] B. A. Olshausen and D. J. Field, "Emergence of simple-cell recpetive field properties by learning a sparse code for natural images" *Nature*, 381, 607-609, 1996.

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

[33] 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.

[34] G. Winkler, *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods*, Springer-Verlag, 1995.

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

[36] 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.

[37] 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.

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

[39] S.C. Zhu, Y.N. Wu and D.B. Mumford, "Filters, Random fields, And Maximum Entropy (FRAME): towards a unified theory for texture modeling", *Int'l Journal of Computer Vision* 27(2) 1-20, March/April, 1998.

[40] S. C. Zhu, "Stochastic computation of medial axis in Markov random field", *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998.

Relevant references are available at the author's web page: http://www.cis.ohio-state.edu/~szhu/publication.html