PERGAMON

Pattern Recognition 35 (2002) 1507–1526

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

On deformable models for visual pattern recognition ☆

Kwok-Wai Cheung*, Dit-Yan Yeung, Roland T. Chin

Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Received 18 December 2000; accepted 30 May 2001

Abstract

This paper reviews model-based methods for non-rigid shape recognition. These methods *model*, *match* and *classify* non-rigid shapes, which are generally problematic for conventational algorithms using rigid models. Issues including model representation, optimization criteria formulation, model matching, and classification are examined in detail with the objective to provide interested researchers a roadmap for exploring the field. This paper emphasizes on 2D deformable models. Their potential applications and future research directions, particularly on deformable pattern classification, are discussed. © 2002 Published by Elsevier Science Ltd on behalf of Pattern Recognition Society.

Keywords: Deformable models; Model representation; Criteria formulation; Matching; Classification; Topology adaptation; Regularization; Optimization; Initialization; Constraint incorporation

1. Introduction

1.1. Deformable model-based recognition

Model-based recognition is a process in which an a priori model is searched for in an input image and subsequently its occurrence and location are determined. This approach has been successfully applied to recognize rigid objects, such as machinery parts and printed characters under noisy environment. However, its performance degrades significantly if the shapes to be recognized are non-rigid, such as human faces, cells, gestures and handwritten characters. Although multiple models could be used for each shape in order to represent different possible deformations, such an extension is generally not always feasible due to the requirement of a large model set, implying a high computational cost, yet still with no guarantee that all possible deformations be taken care of.

Deformable models (DM),[1] on the other hand, refer to models which possess shape-varying ability, making them suitable for representing non-rigid patterns . By matching DMs to imagery data, target shapes with possible deformations can be extracted. As a result, multi-class classification is feasible by defining a set of DMs, each containing its own pertinent shape information with an allowable range of deformation specified using a priori information or by training.

1.2. A common formulation

For the sake of subsequent discussion, a common formulation for DM-based recognition is first introduced.

☆ The research reported in this paper was supported in part by the Hong Kong Research Grants Council (RGC) under Competitive Earmarked Research Grants HKUST 746/96E and HKUST 6081/97E.

* Corresponding author. Tel.: +852-233-95965; fax: +852-233-97892.

E-mail addresses: william@comp.hkbu.edu.hk (K.-W. Cheung), dyyeung@cs.ust.hk (D.-Y. Yeung), roland@cs.ust.hk (R.T. Chin).

[1] They are also often called *flexible models.*

PII: S0031-3203(01)00135-2

1.2.1. Modeling

Let $\mathscr{H}_j$ denote a DM based on a particular representation. The model shape is characterized by a parameter vector $\boldsymbol{w}$, where the parameter values span the parameter space of $\mathscr{H}_j$. The model $\mathscr{H}_j$ with a particular parameter vector $\boldsymbol{w}$ is denoted as $\mathscr{H}_j(\boldsymbol{w})$. Model deformation is then specified by varying $\boldsymbol{w}$ in $\mathscr{H}_j(\boldsymbol{w})$.

1.2.2. Matching

Let $\boldsymbol{D}$ denote the observed imagery data. Given $\boldsymbol{D}$ and some initial model parameter vector $\boldsymbol{w}^0$, model matching can be viewed as a search process in the parameter space, resulting in a sequence of DMs $\mathscr{H}_j(\boldsymbol{w}^0), \mathscr{H}_j(\boldsymbol{w}^1), \ldots, \mathscr{H}_j(\boldsymbol{w}^f)$, where $\boldsymbol{w}^f$ denotes the parameter vector of the final matched model.

Furthermore, the search process is commonly implemented as a multi-criterion optimization problem, although there are studies [1,2] which use heuristic matching processes instead. The following defines the common optimization criteria involved.

Model deformation criterion: This criterion measures the degree of model deformation, i.e., the discrepancy between the current model $\mathscr{H}_j(\boldsymbol{w})$ and the reference (i.e., undeformed) model $\mathscr{H}_j(\bar{\boldsymbol{w}})$, where $\bar{\boldsymbol{w}}$ is the parameter setting of the reference model. This function is denoted as $E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}})$, which represents the "energy" of the current model with respect to the reference model.

Data mismatch criterion: This criterion measures the degree of data mismatch, i.e., the discrepancy between the current model $\mathscr{H}_j(\boldsymbol{w})$ and the given data $\boldsymbol{D}$. This function is denoted as $E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D})$, which represents the "energy" of the data with respect to the current model. [2]

Combined criterion function: By combining the model deformation and data mismatch criteria, the "total energy" of the current stage of deformation is defined as $E_{tot}(\boldsymbol{w}, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}}) = \mathscr{Y}(E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}}), E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}), \alpha)$ where α is a regularization parameter used to regulate the significance of model deformation relative to that of data mismatch. The function $\mathscr{Y}$ denotes a specific combination rule, for example, simple summation or weighted averaging. The matching process thus attempts to minimize the total energy E_{tot} to obtain the final solution $\boldsymbol{w} = \boldsymbol{w}^f$ using $\mathscr{H}_j$ and $\bar{\boldsymbol{w}}$ as a priori information about the model.

1.2.3. Classification

Let $\mathscr{H} = \{\mathscr{H}_1, \mathscr{H}_2, \ldots, \mathscr{H}_M\}$ denote a set of M different DMs. Classification can be formulated as direct comparison of the model candidates based on the discriminant measure $\mathscr{Z}(\boldsymbol{w}^f, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}})$ associated with each $\mathscr{H}_j$. The model with the minimum value of $\mathscr{Z}(\boldsymbol{w}^f, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}})$ is taken as the classified output. If the matching and classification processes share the same criterion, i.e., $\mathscr{Z} = \mathscr{Y}$, then pattern matching and classification may simply be implemented as one single integrated step.

1.2.4. Bayesian framework

It has been pointed out by many researchers [3–5] that DM-based recognition can be formulated using a Bayesian framework. Under the framework, the deformation of a model $\mathscr{H}_j(\boldsymbol{w})$, quantified by $E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}})$, can be interpreted as the uncertainty that it is indeed deformed from its reference model $\mathscr{H}_j(\bar{\boldsymbol{w}})$. The data mismatch, quantified by $E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D})$, can be understood as the uncertainty that the input data $\boldsymbol{D}$ indeed comes from the model $\mathscr{H}_j(\boldsymbol{w})$. Using the Gibbs distribution, such uncertainties can be represented as probabilities, where E_{def} corresponds to the prior distribution of the model parameters and E_{mis} corresponds to the likelihood function. Using the Bayes rule, E_{tot} then corresponds to the posterior distribution. Mathematically, they are written as

$$\text{Prior of } \boldsymbol{w}\text{: } p(\boldsymbol{w}|\alpha, \mathscr{H}_j, \bar{\boldsymbol{w}}) = \frac{1}{Z_{def}(\alpha)} \exp(-\alpha E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}}),$$

$$\text{Likelihood of } \boldsymbol{w}\text{: } p(\boldsymbol{D}|\boldsymbol{w}, \mathscr{H}_j) = \frac{1}{Z_{mis}(\boldsymbol{w})} \exp(-E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D})),$$

$$\text{Posterior of } \boldsymbol{w}\text{: } p(\boldsymbol{w}|\boldsymbol{D}, \alpha, \mathscr{H}_j, \bar{\boldsymbol{w}}) = \frac{1}{Z_{tot}(\alpha, \boldsymbol{D})} \exp(-E_{tot}(\boldsymbol{w}; \alpha, \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}}),$$

where $Z_{def}(\alpha)$, $Z_{mis}(\boldsymbol{w})$ and $Z_{tot}(\alpha, \boldsymbol{D})$ are the partition functions for normalization and α is the regularization parameter. Thus, minimizing the combined criterion $E_{tot}(\boldsymbol{w}; \alpha, \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}})$ is equivalent to maximizing the posterior distribution $p(\boldsymbol{w}|\boldsymbol{D}, \alpha, \mathscr{H}_j, \bar{\boldsymbol{w}})$ with respect to $\boldsymbol{w}$.

1.3. Development milestones

The development of DM-based recognition has a history of more than two decades. The earliest works include *rubber marks* proposed by Widrow [6], *spring model* proposed by Fischler et al. [7] and *elastic matching* proposed by Burr [1]. The study of DMs blossomed in late 1980s due to the work of *active contour models* (or *snakes*) proposed by Kass et al. [8], where DM matching is formulated as optimization of a combined criterion with an internal energy (i.e., model deformation) term E_{int} and an external energy (i.e., data mismatch) term E_{ext}. The optimization process can be equivalently represented by the dynamics of a physical process, where

[2] In the literature, the data mismatch criterion is often defined to take also negative values. In that case, the criterion should be interpreted as a relative measure in the sense that its value is only meaningful when compared with another value.

$\partial E_{int}/\partial \mathbf{w}$ acts as internal force and $\partial E_{ext}/\partial \mathbf{w}$ as external force. The model will then deform and settle at an equilibrium point of the dynamical system. [3]

Most of the existing works on DMs are purely based on an optimization framework, which, as mentioned in Section 1.2.4, can have a probabilistic interpretation. Introducing probabilistic interpretation to computer vision tasks, e.g., boundary detection, can be dated back to Cooper's early work [10] in 1979, where a maximum likelihood method was proposed for detecting blob boundaries in noisy images. In 1992, Staib et al. [3] formulated a DM matching problem as *maximum* a posteriori (MAP) estimation using Bayesian techniques. Also, Revow et al. [4] proposed to perform DM matching using the expectation–maximization (EM) algorithm [11], which is originated from statistics for maximum likelihood estimation (MLE) problems with missing data. Besides, they also combined the probabilistic interpretation with MacKay's evidence framework [12] and formulated a DM-based classification problem, previously formulated in a heuristic manner, formally as a model selection problem.

1.4. Paper organization

In the literature, there exist, to the best of our knowledge, at least two related survey papers, one by McInerney et al. [13] and another one by Jain et al. [14]. McInerney et al.'s paper emphasizes 2D and 3D DMs for medical image analysis, with particular interest in matching and tracking of non-rigid biological objects in 2D and 3D medical images. Jain's paper is focused on 2D DMs and it provides detailed descriptions on a particular type of DMs and their related applications. To contrast with these two existing surveys, we attempt to provide another survey of the field from the perspective of the three main steps of model-based pattern recognition—*modeling*, [4] *matching* (*or extraction*) and *classification*. Instead of discussing the variety of DM-based systems one by one, we deliberately separate our discussions according to the three main steps and discuss related issues and proposed remedies accordingly. As the considerations for the three steps are rather orthogonal, various combinations of the corresponding techniques can in fact be explored according to specific applications. [5] So, we believe that this organization can help the readers to compare and contrast different approaches more easily and eventually can gain more insights for solving the underlying problems. Also, since DM-based pattern recognition is by far mainly focused on 2D problems, we restrict our discussions to 2D models. However, as there exist many inspiring and important ideas that have been proposed for 3D DMs only, they will also be included in this paper as far as they can help in explaining some related ideas.

In the literature, there also exist some other related matching algorithms like relaxation labeling [15] and self-organizing map [16]. Since their formulations are quite different from the common framework described in Section 1.2, they will not be further described in this paper. The remaining paper is organized as follows.

Representative DM representations proposed in the literature are first described and categorized in Section 2. Sections 3, 4 and 5 address different issues related to criterion formulation, matching and classification. In addition, various proposed solutions are described and compared. Current applications of DMs to pattern recognition and related research problems are discussed in Section 6. Finally, Section 7 concludes the paper.

2. Model representation

2.1. Representation dilemma between reconstruction and recognition

The DM-based approach has been found to be an effective tool for both shape reconstruction (or recovery) and shape recognition. However, due to their different ultimate goals, these two tasks impose contradictory requirements for model representation. Shape reconstruction concerns the *physical* quality of the reconstructed contours/surfaces/volumes for later visualization or measurement. So, the required modeling scheme should have high representational power for capturing local shape details. This implies the need for representations with only simple local constraints (e.g., smoothness) plus a relatively larger number of "distributed" parameters. Shape recognition emphasizes on the discriminative quality of the salient features extracted by the model. Thus, related representations require some "*just-enough*" representation power so that intra-class variations of some global shape properties can be effectively described while outliers from irrelevant classes can be discriminated. These requirements imply the need for representations capable of characterizing global shapes using a small set of parameters. Related concerns are the representation uniqueness and their efficiency [17,18]. Although reconstruction-oriented representations can also provide a large amount of shape information for classification, they are often too redundant and hard to be interpreted, easily leading to bad classification. Having said this, it should be noted that model representation is just one (though important) factor affecting DM-based

[3] In computer graphics, such physically based models are also very useful for producing realistic animation of the interaction of non-rigid natural objects. Interested readers are referred to Ref. [9] for further details.

[4] It includes model representation and criterion formulation of DMs.

[5] For some cases, however, some considerations of the three steps are combined to a certain extent.

		Local Shape Parameterization	*Global Shape Parameterization*
Prototype-based		-	**Deformation Transform of a Prototype** - Different transform bases [19,5] - Deformable intensity models [20] **Mixture of Multiple Prototypes** - Mixture of boundaries [21] or images [22]
Description–based	*Fixed Topology*	**Boundary-based:** - Active contour models [8] - B-spline [23,4]; Bézier [24] - Local affine transform [25] **Grid-based:** - GSnake [26] - Labeled graphs [27,28]	**Analytical:** - Parametric curves [29,30] **Decomposition-based:** - Fourier decomposition [3] - Thin-plate spline [31] - Eigen decomposition: - Active shape models [2] - Modal matching [32]
	Adaptive Topology	**Split-and-Merge:** - T-Snakes [33]	**Split-and-Merge:** - Blended deformable models [34] **Topology-Free:** - Level Sets: - Geodesic active contours [35] - Implicit polynomials [36] - Pedal curves: Snake pedals [37]

Fig. 1. A classification of DMs with citation of representative works.

recognition. There still exist successful cases where reconstruction-oriented representations are adopted for recognition, but then the discriminant measures to be used in the classification step (see Section 5) have to be carefully designed.

In the following, a taxonomy of DM representations is provided. Representative works under each of the categories are then surveyed. We will end the discussion by revisiting this representation dilemma regarding the different categories of DMs.

2.2. A taxonomy of DM representations

Here, we categorize DMs proposed in the literature using two orthogonal parameterization characteristics: (i) locally vs. globally parameterized; (ii) description-based vs. prototype-based. Local DMs adopt parameterization schemes to model local shape characteristics while global DMs model overall shape characteristics. Description-based DMs model shape explicitly (shape descriptive) with deformations modeled as perturbations in the shape parameter space. On the other hand, prototype-based DMs parameterize shape deformations directly and the parameters are not shape descriptive.

Also, based on the adaptability of their topology, description-based DMs are further categorized according to whether the topology is fixed or adaptive. Fig. 1 shows a 3×2 matrix derived based on the aforementioned characteristics with representative DMs tabulated accordingly.

2.3. Description-based DMs with fixed topology

2.3.1. Local parameterization

Local description-based DMs are also called *free-form* models, *physics-based* models, or *distributed parameter* models. Some of them explicitly model object boundaries (*boundary-based*) while some model objects using the local spectral properties at different locations of the objects with their spatial relationships maintained by a grid structure (*grid-based*).

(*a*) *Boundary-based.* Polygonal representation is one of the simplest ways to represent a shape, say the silhouette of an object, by a linear interpolation of an ordered set of points. It is locally parameterized, where each model point can move freely without involving movement of the others. Such a polygonal representation is used in Kass et al.'s active contour model [8] and has been widely studied by many other researchers in the field [19–23], mainly due to its simplicity.

Piecewise parametric curves are good candidates for representating smooth objects. Using a set of local basis functions, smooth shapes can be described with a relatively small set of parameters when compared with the polygonal representation. Examples include Bézier [24] and B-spline curves [4,25], which can guarantee C^1 and C^2 continuities, respectively. [6] They are still local since movement of a related control point only affects some immediately preceding and following curve segments.

Local affine transform (LAT) is proposed by Wakahara for handwriting recognition [26] and can be viewed as a piecewise linear approximation of the overall deformation using a set of localized affine transforms. The scheme starts with a polygonal representation and neighboring model points are grouped together, where each group is associated with a single local affine transform. Deformations are then modeled by perturbations in affine transform parameters.

(*b*) *Grid-based.* Labeled graphs with 2D regular structures were first proposed by Lades et al. [27] for face recognition and the corresponding matching process is known as *labeled graph matching*. The grid representation also includes the responses of multi-scale, multi-orientation Gabor wavelets applied to a reference face as grid points' attributes (*landmarks*) to represent the local spectral properties distributed over the face, with their spatial relationship maintained by the grid structure. The use of attributes relaxes the requirement of the shape modeling accuracy for the grid as far as the important "landmarks" of the interested objects can be accurately identified and represented. The regular grid representation was later enhanced to a *fiducial graph*, a specially designed graph triangulation with grid points (graph node) located at anatomically identifiable points of the face for a more accurate representation [28].

2.3.2. Global parameterization

Global description-based DMs are also called *parameteric* models, *geometric* models or *lumped parameter* models. They are normally used when some prior knowledge on what shapes to be represented is known. Then, the particularly adopted parameterization scheme determines the family of shapes that can be represented. Various related parameterizations have been proposed in the literature, which can be further categorized into *analytical* or *decomposition-based.*

(*a*) *Analytical.* Parametric curves here refer to 2D curves with simple analytical forms so that their shapes can be controlled by a compact set of parameters $\boldsymbol{w}$ (e.g., $y - w_1 = w_2(x - w_3)^2$ for parabolas). They are global shape models, as any change in one of the elements in $\boldsymbol{w}$ results in changes in the overall shape. In general, parametric curves have restricted modeling capability, which however may be good enough for specific applications where only some restricted classes of shapes are of interest and at the same time gain computational advantage by reducing the search space for the subsequent matching step. Besides, in many cases the extracted model parameters can provide meaningful shape information for subsequent image analysis. A good example is the deformable template used by Yuille et al. [29] to locate the eyes in a human face image. The template consists of two parabolic curve segments modeling the upper and lower eye lids and one circle modeling the iris. The model is thus parameterized by the coefficients of the two parabolas together with the center and radius of the circle. The coefficients of the parabolas define their convexities and the circle's radius provides dimensional information about the iris.

(*b*) *Decomposition-based.* The shape decomposition approach represents shapes of similar types via combination of a set of basis functions with global support, where the basis can be either pre-defined in a priori manner or obtained via training. Compared with analytical models, decomposition-based models are in general less restrictive with respect to their representational power, which is controlled by the corresponding basis adopted.

Fourier decomposition of contours has been adopted in Ref. [3], where sinusoidal functions of different harmonics are used to form the basis. A parametric contour $\{\boldsymbol{v}(\boldsymbol{t})\}$ is represented by the following formula:

$$\boldsymbol{v}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \simeq \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{B} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \sin(kt) \\ \cos(kt) \end{bmatrix},$$

where B denotes the number of harmonics used. The model parameters are $\{a_0, c_0, a_k, b_k; 1 \leqslant k \leqslant B\}$, which control the contribution of different harmonics. For representing smooth objects, B can normally be restricted to a small number so that only the low-frequency components are retained and the parameter set remains to be a compact one. Also, the sinusoidal basis is orthogonal, which makes the representation unique and thus favorable for recognition purpose.

Thin-plate spline kernel has also been proposed as the basis [30], where a curve $\{\boldsymbol{v}(t)\}$ is represented accordingly as $\boldsymbol{v}(t) = \boldsymbol{v}(t) \cdot \boldsymbol{A} + \boldsymbol{\Phi}(\boldsymbol{v}(t)) \cdot \boldsymbol{w}$, assuming that $\boldsymbol{v}(t)$ (1×3) is in homogeneous coordinate system, $\boldsymbol{A}$ (3×3) is the affine transform, $\boldsymbol{w}$ $(K \times 3)$ is the warping coefficients (model parameters), $\boldsymbol{\Phi}(\boldsymbol{v})$ is the thin-plate spline kernel $(1 \times K)$ with its element $\{\Phi_i(\boldsymbol{v})\} = \{c||\boldsymbol{v} - d_i||^2 \log||\boldsymbol{v} - d_i||\}$, d_i is the ith data point and K is the total number of data points. The kernel contains information about the model point-set's internal structural relationship. Again, the thin-spline kernel is chosen in a priori manner.

Eigen decomposition is another technique commonly used to derive a basis for representing a shape family. Dimension reduction can be effectively achieved by retaining only the significant eigenvectors as the basis. Mathematically, a shape represented by an ordered set of point $\boldsymbol{v}$ can be approximated by a linear mixture of

[6] A C^n continuous piecewise curve means that the nth derivative of the curve respect to the parameter is equal at the join points of the curve segments.

eigenvectors $\{\boldsymbol{q}_i; 1 < i < N_q\}$, given as $\boldsymbol{v} = \boldsymbol{v_{ref}} + \sum_{i=1}^{N_q} w_i \boldsymbol{q}_i$ where the mixture coefficients $\{w_i\}$ are the model parameters. While Fourier and thin-plate spline decomposition are originated from the fields of signal processing and data analysis, eigen decomposition tries to derive the basis according to the physical properties of objects. A good example is the *active shape model* (ASM) proposed by Cootes et al. [31] (also dubbed as the point distribution model in Ref. [2]) for face recognition, where eigenvectors are extracted from the covariance matrix of the model points. For situations where adequate training data are not available for computing the covariance matrix, Sclaroff et al.'s [32] *modal matching*, according to the FEM formulation, performs the decomposition on a matrix equal to the product of the mass matrix's inverse and the stiffness matrix to form the deformation basis. Also, Wang et al. [33] proposed a smoothness matrix which forces the neighboring model points to be correlated throughout the model.

2.4. Description-based DMs with adaptive topology

All the aforementioned DMs are capable of representing non-rigid shapes to different extents. However, they all assume that the topology of the object of interest is fixed, which limits their applications to cases with intra-class topology variations.

2.4.1. Local parameterization

For the local boundary-based DMs, the need for modeling topology changes arises when (1) there are an unknown number of boundaries to be extracted (e.g., unknown number of cells in a medical image) or, (2) when boundaries split and merge at junction points. Although the problem can be partially solved by using multiple contour models, the exact number of models required is unknown in most of the cases and hence accurate model initialization is still hard to achieve.

So far, the *split-and-merge* paradigm is the only approach that has been proposed for local DMs to achieve topology adaptation. For example, *T-Snake* was proposed by McInerney et al. [34] as a topology-adaptive extension of the original active contour model. The evolution of a T-snake is defined together with a cell decomposition which partitions the image into regular triangular grids. Each grid vertex is assigned a state value of "on" if the vertex is inside a closed contour or "off" if it is outside. The change of the contour topology is triggered when some neighboring vertices of the grid change from being in the same state to different states (occurs when a contour shrinks and eventually breaks up) or from being in different states to the same state (occurs when two contours meet and eventually merge together). With a similar spirit, Perera et al. [35] also proposed a topology-adaptive active contour model. But instead of relying on a particular grid decomposition, they proposed some algorithms to detect contours intersections directly and determine whether any topology change is needed.

It should be noted that the topology transformation process required by this approach is not part of the common formulation described in Section 1.2 and additional procedures are needed for the detection and structure modification.

2.4.2. Global parameterization

Topology adaptation for global DMs can be achieved either using the *split-and-merge* approach or via *topology-free representation*.

(*a*) *Split-and-merge*. Other than local DMs, the split-and-merge approach can also be applied to global DMs for topology adaptation. The difference lies in the shape primitives involved. The primitives for local DMs are simply line segments while those for global DMs are geometric models with fixed topology. Shapes with various topologies are then represented via primitive blending. The blended DMs proposed by DeCarlo et al. [36,37] requires first going through a splitting process of a single initial model where the splitting locations are detected through analysis of the "force" distribution of the converged models (see Section 4). After splitting, the separated primitives are allowed to deform independently but are joined together via a blending function. The shape parameters of the individual primitives together with the parameters of the blending function form the parameter set of the blended DM.

(*b*) *Topology-free representation*. An implicit way to address the topology adaptation issue is to adopt a topology-free representation. [7] As mentioned above, the split-and-merge approach requires an additional procedure to detect and modify the representation. The use of topology-free representations is attractive as the topology adaptation mechanism is implicitly embedded in the representation and thus no additional processes are required. However, in general those representations are relatively more complicated.

Level set is one of the topology-free examples where a 2D contour is represented by the zero set of a function $z = f(x, y; \boldsymbol{w})$. By varying the coefficients $\boldsymbol{w}$, the topology of the zero set changes accordingly (but not the topology of $z = f(x, y; \boldsymbol{w})$). This representation has been adopted in geodesic active contours [38,39] and is highly related to shape modeling with front propagation [40]. Also, this representation is very similar to implicit polynomials, studied by Subrahmonia et al. as well as some other researchers from the perspective of data fitting [41].

[7] Note that similar techniques do not exist for locally parameterized DMs as it is hard to couple the overall shape topology with some localized parameterizations.

Pedal curve is another tool borrowed from differential geometry for deriving topology-free shape representations. It is defined with respect to a predefined planar curve α (generator) and a predefined pedal point $\boldsymbol{p}$, and is formed by tracing the locus of the intersection points between the tangents of α and the perpendicular line drawing from $\boldsymbol{p}$ to the tangents. Vemuri et al. [42] modified the pedal curve by associating the generator α with a corresponding snake which replaces the role of $\boldsymbol{p}$ and proposed *snake pedals.* [8] By using discontinuous (or multiple open) snakes, curves of different topologies can be represented. However, it is not mentioned how the splitting of the snake can be achieved.

2.4.3. Hybrid parameterization

To gain the advantages of both local and global parameterizations, DMs using hybrid parameterization schemes have been proposed. One can pick one global and one local parameterization schemes and combine them to form hybrid models. The snake pedal is one of the examples, which consists of global shape parameters forming the planar curve α (α being an ellipse in Ref. [42]) and local shape parameters driving the shape of the snake. By combining B-spline curves and thin-plate splines, Amini et al. proposed a DM for analyzing 2D tissue deformation in medical images [43]. Mignotte et al. [44] integrate parameteric displacement fields with local perturbations. Similar ideas have also been widely used in 3D DMs. One of the interesting examples is the *deformable superquadric* (DS) [45]. The DS consists of a superquadric which is parameterized by a compact set of parameters so that it can deform from an ellipsoid to a cube continuously. A linear combination of some basis functions is superimposed on it for capturing local deformation details. Further extensions of the DS include [36,46,47]. The idea of superimposing local deformations has also been applied to global parameteric models like *general cylinders* [48] and *hyperquadric* [49] for increasing their local representational power.

2.5. Prototype-based DMs

One major characteristic of description-based DMs is to use explicit shape abstractions for representing objects. In contrast, prototype-based DMs are fundamentally different in that the building blocks of the models are simply images or shape prototypes (cf. non-parametric) and the underlying deformation process is parameterized instead. Using the prototype-based paradigm, the model construction step is unnecessary. This, however, implies that the global shape information cannot be easily captured without further post-processing steps.

2.5.1. Deformation transform of a single prototype

One way to model the deformation process is by using background displacement fields (deformation transform). Denote $I_{prototype}(x, y)$ as the intensity profile of a prototype, where each pixel takes the intensity value of 0–255 for gray-level images or 0–1 for binary images. A particular prototype can be deformed according to the equation $I_{deform} = I_{prototype}(x + u(x, y), y + v(x, y))$ where $u(.,.)$ and $v(.,.)$ denote the horizontal and vertical displacement fields, respectively. To represent smooth deformations, two sets of eigenfunctions $\{\varphi_i^u(x, y)\}$ and $\{\varphi_i^v(x, y)\}$ can be used to define the displacement fields, where $\Delta\varphi_i^u(x, y) \propto w_i^u \varphi_i^u(x, y)$ and $\Delta\varphi_i^v(x, y) \propto w_i^v \varphi_i^v(x, y)$. The displacement fields $u(x, y)$ and $v(x, y)$ can then be expressed as linear combinations of the eigenfunctions such that $u(x, y) = \sum_i w_i^u \varphi_i^u(x, y)$ and $v(x, y) = \sum_i w_i^v \varphi_i^v(x, y)$. The parameters of the corresponding DM are the set of weights $\{w_i^u\}$ and $\{w_i^v\}$. How well deformations can be modeled depends on the choices of the eigenfunctions. This approach to model shape deformations has been proposed by Grenander et al. under their *pattern theory* developed for studying biological shapes [50–52]. Similar ideas have also been adopted in Refs. [5,53,54]. So far, to the best of our knowledge, all the displacement fields adopted in the literature are global ones and there does not exist any work using some kind of localized displacement fields (which accounts for the empty upper left cell in Fig. 1).

By adding one more dimension (intensity) to the 2D displacement fields, Moghaddam et al. [55] proposed a deformable intensity model for image matching where a reference image is *warped* onto an input image via the mapping $h(x, y, I) = (x + u(x, y, I), y + v(x, y, I), I + l(x, y, I))$. This modification can allow the model to capture also the variations due to pose changes and lighting conditions.

2.5.2. Mixture of multiple prototypes

Instead of explicitly modeling the deformation process, one can use a mixture of prototypes $\boldsymbol{p}_i$ to represent a shape $\boldsymbol{v}$ such that $\boldsymbol{v} = \sum_i w_i \boldsymbol{p}_i$. [9] The prototypes are basically deformed versions of a reference shape and the model parameters are the mixing coefficients ($\sum_i w_i = 1$). Thus, v is bounded within the simplex with the prototypes being its vertices. The corresponding representational power is determined accordingly. This approach is originated from Ullman et al.'s work [56] which uses linear combinations of object canonical views ($\boldsymbol{p}_i$) to represent an arbitrary view of the object due to rigid 3D transformations. Tanaka et al. [57] used images of carefully chosen deformed patterns for $\boldsymbol{p}_i$ and their mixtures for representing arbitrarily deformed patterns. The mixing process can

[8] Note that with the introduction of the snake, the snake pedal is no longer a global DM but should instead be classified as a hybrid one. See Section 2.4.3.

[9] Note that the correspondence between the prototypes is assumed but not the correspondence of the model and the input image, which is to be established in the matching step.

be understood as stacking up the prototype images and summing up their pixel values weighted by corresponding w_i, whose values are determined via matching.

2.6. Representation dilemma revisited

Generally speaking, global DMs are more often proposed for shape recognition while local ones are found to be more applicable to shape reconstruction. Whether hybrid DMs are needed depends on specific applications. If the nature of the recognition task requires detailed examination of the local shape properties or an application first requires a recognition step and then a reconstruction step (e.g., model-based compression [58]), hybrid models seem to be inevitable. For the issue of prototype-based versus description-based approaches, both have been adopted for recognition as well as reconstruction. Although it is beyond the scope of this paper to argue which approach is more suitable for recognition or reconstruction, a general comment is that the prototype-based approach is worth trying for applications, like signature recognition where description-based representations are hard to construct accurately.

2.7. Geometric invariance

The geometric invariance property, though largely ignored in boundary detection, is vital to viewpoint-independent object recognition. If an object is moved along a plane perpendicular to the optical axis of the camera, along that axis or is allowed to be tilted, the corresponding shape change can be approximated by an *affine transform* which includes scaling, shifting, rotation and shearing [59]. By factoring out the underlying transformation or using a representation invariant to affine transformation, the DM becomes viewpoint-independent, or called *geometrically invariant*. [10]

2.7.1. Affine transform incorporation

For a contour $\boldsymbol{v}$ parameterized by $\boldsymbol{w}$ such that $\boldsymbol{v}=\mathscr{H}(\boldsymbol{w})$, one can incorporate the affine transform $\boldsymbol{A}$ such that $\boldsymbol{v}=\boldsymbol{A}\mathscr{H}(\boldsymbol{w})$, where $\boldsymbol{A}$ and $\boldsymbol{w}$ are to be determined in the matching step. The deformation effect due to $\boldsymbol{A}$ can be factored out by using only $\boldsymbol{w}$ for computing the amount of shape deformation. An additional advantage is that the estimated affine transform sometimes may be needed for further analysis. This approach normally involves a higher computational complexity due to the increased number of parameters and possibly more complicated (non-linear) formulations. Under certain representations, however, this approach can still be effectively applied to achieve affine invariance [4,25,29].

[10] But, it should be noted that, for some applications, not all affine transforms are equally common and the rare ones should be restricted via a proper prior [60].

2.7.2. Affine invariant representations

Choosing a geometrically invariant parameterization is an implicit alternative to achieve affine invariance. The approach can in general enjoy computational gain as no additional affine transform parameters are required. One example is the G-Snake (a contour model) proposed by Lai et al. [23], which uses a shape matrix for shape representation. The matrix is created by first substracting the coordinates of all the contour points by an arbitrarily fixed reference to form a set of vectors $\{\boldsymbol{u}_i\}$. Then, each $\boldsymbol{u}_i$ can be expressed as $\alpha_i\boldsymbol{u}_{i-1}+\beta_i\boldsymbol{u}_{i+1}$ (except for the boundary cases). The shape matrix is formed based on $\{\alpha_i\}$ and $\{\beta_i\}$ and can be shown to be rotation. Scale invariance can also be achieved by normalization. Another example is the implicit polynomial adopted by Subrahmonia et al. [41]. It uses algebraic invariants, which are functions of the implicit polynomial coefficients and are derived to be invariant to affine transformations. The trade-off is the loss of the estimated affine transform.

3. Criteria formulation

Given a model constructed based on a chosen model representation, the model deformation process is mainly controlled by formulating a criterion function which combines a *model deformation* criterion and a *data mismatch* criterion, which is to be optimized for model matching. The criterion functions can be interpreted as *soft* constraints (penalties) for restricting the resultant model shape. [11]

3.1. Model deformation criterion

A model deformation criterion measures the degree of model deformation for a DM. The related criterion in Kass et al.'s active contour models [8] is called *internal energy*, [12] given as

$$\int_0^1 \alpha_1(s)\left|\frac{\partial \boldsymbol{v}}{\partial s}\right|^2+\alpha_2(s)\left|\frac{\partial^2 \boldsymbol{v}}{\partial s^2}\right|^2 \mathrm{d}s, \tag{1}$$

where $\boldsymbol{v}(s)$ denotes the contour parameterized by $s\in[0,1]$. Spatial derivatives of the criterion function can be interpreted as some internal forces exerted on the model to restore it back to its reference (or undeformed) shape. Although a relatively large portion of the discussion in the following section is related to the active contour model or its variants, most of them can be easily generalized by plugging in some other representations. The only difference is the associated physical interpretation for the criteria.

[11] Hard constraints can also be incorporated for controlling the deformation. See Section 4.2.

[12] Terms with similar meanings are also called deformation energy in Ref. [4], strain energy in Ref. [32], etc.

3.1.1. A summary of deformation criteria

Based on Eq. (1), the finite difference technique can be used to discretize s and the integral can be approximated by a quadrative form [8].

(*a*) *Inter-point distance*

$$E_{def_1}(\boldsymbol{w};\mathscr{H}_j,\bar{\boldsymbol{w}})=\sum_{i=1}^{k}||\boldsymbol{v}_i-\boldsymbol{v}_{i+1}||^2 \tag{2}$$

is a discrete version of the first term in Eq. (1) where $\boldsymbol{w}^t=(\boldsymbol{v}_0^t,\boldsymbol{v}_1^t,\boldsymbol{v}_2^t,\ldots,\boldsymbol{v}_{k/2}^t)$ and $||\cdot||^2$ denotes an L_2 norm. It can be expressed using a quadratic form $\boldsymbol{w}^t\boldsymbol{A}\boldsymbol{w}$ where $\boldsymbol{A}$ is a tri-diagonal matrix. $\Delta_w E_{def_1}=2\boldsymbol{A}\boldsymbol{w}$ act as some internal forces pulling the contour points together.

(*b*) *Curvature-related measure*

$$E_{def_2}(\boldsymbol{w};\mathscr{H}_j,\bar{\boldsymbol{w}})=\sum_{i=1}^{k}||\boldsymbol{v}_{i-1}-2\boldsymbol{v}_i+\boldsymbol{v}_{i+1}||^2 \tag{3}$$

is a discrete version of the second term in Eq. (1) which can be expressed as $\boldsymbol{w}^t\boldsymbol{B}\boldsymbol{w}$ where $\boldsymbol{B}$ is a penta-diagonal matrix. $\Delta_w E_{def_2}$ act as some internal forces preventing the model from being bent too much.

(*c*) *Normalized curvature* The criteria in Eqs. (2) and (3) favor a shorter curve to a longer one with the same curvature. To achieve scale invariance, a new criterion can be defined as the ratio of E_{def_2} to E_{def_1} [61].

(*d*) *Distance from reference* For recognition purpose, instead of measuring the intrinsic properties of the model, it is more appropriate to define model deformation as some distance between the current model parameter vector $\boldsymbol{w}$ and its reference value $\bar{\boldsymbol{w}}$ in the corresponding parameter space. It is typical to use a quadratic form for the distance, such that

$$E_{def_3}(\boldsymbol{w};\mathscr{H}_j,\bar{\boldsymbol{w}})=(\boldsymbol{w}-\bar{\boldsymbol{w}})^t\boldsymbol{\Sigma}(\boldsymbol{w}-\bar{\boldsymbol{w}}), \tag{4}$$

where $\boldsymbol{\Sigma}$ could be a diagonal matrix (resulting in a weighted Euclidean distance in the parameter space), or the inverse of the covariance matrix of $\boldsymbol{w}$ learned from a set of training data (resulting in the Mahalanobis distance [4]), or a manually defined stiffness or smoothness matrix [2,32,62]. For physical interpretation, the use of diagonal $\boldsymbol{\Sigma}$ assumes that the model points can be moved independently, while non-diagonal $\boldsymbol{\Sigma}$ introduces interaction among the model points.

3.2. Data mismatch criterion

The data mismatch criterion measures the data discrepancy given the current model. The spatial derivatives of this criterion can be interpreted as some external forces exerted on the model to deform it to match with some regions of interest of the input data. Based on different types of preprocessing on the input, various data mismatch criteria have been proposed. To further improve the matching accuracy, criteria which combine different types of extracted features can also be used [63–65].

3.2.1. A summary of mismatch criteria based on features

The following provides some criteria which have been proposed in the literature for describing data mismatch. The list is by no mean exhaustive but to help readers to understand how different types of extracted features can be used.

(*a*) *Image-based*

$$E_{img_1}(\boldsymbol{w};\mathscr{H}_j,\boldsymbol{D})=-\sum_{i=1}^{k}\mathscr{I}(\boldsymbol{v}_i(\boldsymbol{w});\boldsymbol{D}) \tag{5}$$

was proposed in Ref. [8] for matching bright pixels in a gray-level image, where $\mathscr{I}(\boldsymbol{v}_i;\boldsymbol{D})$ denotes the intensity of the image $\boldsymbol{D}$ at $\boldsymbol{v}_i$ and k denotes the number of elements in $\{\boldsymbol{v}_i\}$.

$$E_{img_2}(\boldsymbol{w};\mathscr{H}_j,\boldsymbol{D}) = -\sum_{l=1}^{N}\log\frac{1}{k}\sum_{i=1}^{k}\frac{\beta}{2\pi}\exp\frac{-\beta||\boldsymbol{v}_i(\boldsymbol{w})-\boldsymbol{y}_l||^2}{2} \tag{6}$$

was proposed in Ref. [4] for matching black pixels in a binary image, where $\boldsymbol{y}_l$ is the location of an individual black pixel, β is a signal-strength parameter, and N is the number of black pixels.

(*b*) *Edge-based*

$$E_{edge_1}(\boldsymbol{w};\mathscr{H}_j,\boldsymbol{D})=-\sum_{i=1}^{k}||\mathscr{G}(\boldsymbol{v}_i(\boldsymbol{w});\boldsymbol{D})|| \tag{7}$$

was proposed in Ref. [8] for matching edges, where $\mathscr{G}(\boldsymbol{v}_i;\boldsymbol{D})$ denotes the gradient vector of $\boldsymbol{D}$ at $\boldsymbol{v}_i$.

$$E_{edge_2}(\boldsymbol{w};\mathscr{H}_j,\boldsymbol{D}) = -\sum_{i=1}^{k}|(\cos\phi\cos\psi\ \mathscr{G}_x,\sin\phi\sin\psi\ \mathscr{G}_y)| \tag{8}$$

was proposed in Ref. [66] for utilizing also the gradient direction information, where $\mathscr{G}_x$ and $\mathscr{G}_y$ are the x and y components of $\mathscr{G}(\boldsymbol{v}_i;\boldsymbol{D})$, respectively, ϕ is the normal angle of the contour at $\boldsymbol{v}_i$ and ψ is the gradient angle there. Such a formulation tries to avoid matching to some undesirable edges caused by irrelevant objects in the neighborhood. Similar ideas have also been adopted in Ref. [61,67].

$$E_{edge_3}(\boldsymbol{w};\mathscr{H}_j,\boldsymbol{D})=-\frac{1}{k}\sum_{i=1}^{k}|\,||\mathscr{G}(\boldsymbol{v}_i(\boldsymbol{w});\boldsymbol{D})||-\bar{\mathscr{G}}| \tag{9}$$

was proposed in Ref. [68] to favor boundaries with a constant gradient, where $\bar{\mathscr{G}}=\frac{1}{L}\sum_{i=1}^{k}||\mathscr{G}(\boldsymbol{v}_i(\boldsymbol{w});\boldsymbol{D})||$ is the average gradient magnitude.

(*c*) *Region-based*

$$E_{reg_1}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}) = -\sum_{i=1}^{k} \mathscr{F}(\boldsymbol{v}_i(\boldsymbol{w}); \boldsymbol{D}) \tag{10}$$

was adopted in Ref. [69,70], where $\mathscr{F}$ is an indicator function for foreground pixels and can be computed using some standard region-based segmentation algorithms. This criterion tends to direct the model to enclose regions with homogeneous grey level, which however relies heavily on the accuracy of $\mathscr{F}$.

$$E_{reg_2}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}) = -\sum_{\boldsymbol{s} \in \boldsymbol{R}(\boldsymbol{w})} \left[\left(\frac{\mathscr{I}(\boldsymbol{s}; \boldsymbol{D}) - \mu}{\sigma} \right)^2 - f_s \right] \tag{11}$$

was used in Ref. [71] to tolerate noisy homogeneous regions, where $\boldsymbol{R}(\boldsymbol{w})$ denotes the set of pixels enclosed by the model contour, $\mathscr{I}(\boldsymbol{s}; \boldsymbol{D})$ is the pixel intensity at $\boldsymbol{s}$, μ is the mean of the foreground intensity, σ is the standard deviation and f_s is some adaptively estimated offset.

(*d*) *Motion-based*

$$E_{mot}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}) = -\sum_{i} \mathscr{M}_v(\boldsymbol{v}_i(\boldsymbol{w}); \boldsymbol{D}) \tag{12}$$

was proposed in Refs. [64,72] to facilitate object segmentation using motion information, where $\mathscr{M}_v$ is an indicator function for moving pixels computed based on consecutive image frames.

(*e*) *Landmark-based*

$$E_{landmark}(\boldsymbol{w}, \boldsymbol{f}; \mathscr{H}_j, \boldsymbol{D}) = -\frac{\sum_i f_i \hat{f}(v_i(\boldsymbol{w}); \boldsymbol{D})}{\sqrt{\sum_i f_i^2 \sum_i \hat{f}(v_i(\boldsymbol{w}); \boldsymbol{D})^2}} \tag{13}$$

was proposed in Ref. [27] for their landmark-based deformable grids, [13] where f_i denotes the landmark feature associated to v_i and $\hat{f}(v_i(\boldsymbol{w}); \boldsymbol{D})$ denotes the landmark feature extracted at the location of v_i.

3.2.2. Noise modeling and robust statistics

As mentioned in Section 1.2.4, the data mismatch criterion can be interpreted as the likelihood of the input data based on some noise model. For example, we can modify Eq. (5) by adding the value 255 to all the terms in the summation so that each term can be interpreted as measuring the deviation (or noise) of the actual intensity from the ideal value 255. The corresponding noise model then follows a Laplacian distribution such that

$$p_{img_1}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}) \propto \exp\left(-\sum_{i=1}^{k} |255 - \mathscr{I}(\boldsymbol{v}_i(\boldsymbol{w}); \boldsymbol{D})| \right). \tag{14}$$

A similar argument applies to Eq. (11), which corresponds to a Gaussian noise model for the intensities of the pixels enclosed by the contour model. Also, for Eq. (6), the noise model for the location of the black pixels is a mixture of Gaussians. The introduction of probabilistic interpretation implies that the data mismatch criterion can in fact be formulated from the perspective of modeling the underlying noise process of the problem, which, under the context of computer vision, is related to the imaging environment. For example, Poisson distribution has been reported to be better for describing noise in images at low photon levels, like astronomical images. Following this line of thinking, noise models of different members in the exponential family (e.g., Poisson, Rayleigh, Bernoulli, etc.) have been studied in Ref. [73] for formulating data mismatch criteria.

Even though there exist many different sophisticated noise models, there are still cases where the use of any noise models may not be sufficient, especially when the input data contain outliers. Robust statistics [74] can be used to modify the data mismatch criterion to discount the effect caused by outliers [75]. In particular, for the application of extracting characters from handwritten cursive script using DMs, it is quite unreasonable to treat the outlier data using, say a simple uniform noise model. Using a robust statistical technique called M-estimation, Cheung et al. [76] has showed that non-rigid characters can be correctly extracted even in the presence of outliers.

3.3. Regularization

Model deformation and data mismatch are two often conflicting criteria. The most common way to achieve a trade-off is to define a combined criterion function by a (possibly adaptively) weighted sum of the two. Such a technique, called *regularization* [77], has long been used for computer vision tasks [78]. Let α denote the regularization parameter. The combined criterion function is

$$E_{tot}(\boldsymbol{w}, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}}) = \alpha E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}}) + E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}). \tag{15}$$

Properly setting the value of α is vital for good matching, especially when the imagery data is noisy. Either too small or too large the value of α will result in bad matching (Fig. 2). Finding an input-independent optimal value of α either through training or by trial and error is one solution. Better matching can also be achieved by carefully decreasing the value of α from large to small, or in other words, allowing the model to be rigid at the beginning to obtain a rough match and then increasingly flexible towards the end for a fine match [1,4,29,79]. To achieve adaptive regularization, cross-validation has been adopted by Shahraray et al. [80], which however is rather computationally expensive. Another approach

[13] Here a negative sign is added to the original similarity measure to make a *mismatch* criterion.

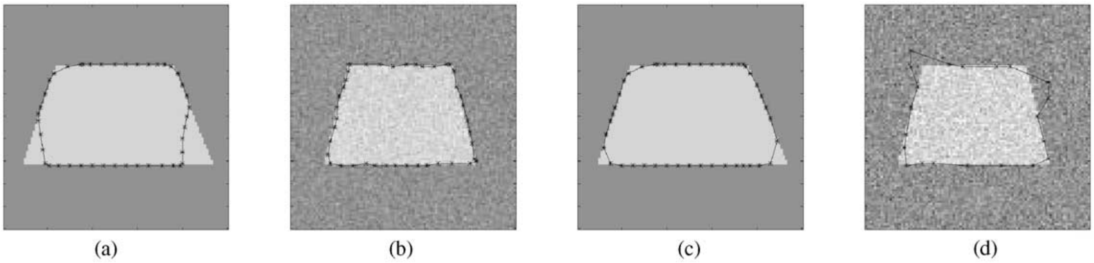

Fig. 2. Effects of different values of regularization parameter α. For the noiseless cases (a) ($\alpha = 0.02$) and (c) ($\alpha = 0.001$), too large the value of α is less preferable because sharp corners, which in many situations serve as salient features, cannot be detected. For the noisy cases (b) ($\alpha = 0.02$) and (d) ($\alpha = 0.001$), too small the value of α is less preferable as the model then cannot smooth out the noise and becomes very sensitive to it.

used by Cheung et al. [81] is to treat α as the hyperparameter and use Bayesian techniques to obtain its MAP estimate.

Besides the standard way of regularization, it is interesting to note that a minimax criterion can also be used [61,82] with $E_{tot}(\boldsymbol{w}, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}}) = \max\{E_{mis}(\boldsymbol{w}; \mathscr{H}_j, \boldsymbol{D}), E_{def}(\boldsymbol{w}; \mathscr{H}_j, \bar{\boldsymbol{w}})\}$ instead of a weighted combined criterion. It has been shown that under certain convexity assumptions, minimizing such a criterion (which implements the matching process to be discussed in the next section) is equivalent to minimizing the weighted sum criterion with optimal regularization.

Before ending this section, it should be stressed that if global parameterization is adopted for model representation, the corresponding regularization will become relatively easier. This is mainly due to the reduced degree of freedom by introducing more a priori knowledge which implicitly results in a better control over the resultant model shape [83].

4. Matching

Matching a DM to the data is performed by minimizing a combined criterion function E_{tot}. From the point of view of a dynamic system, the matching process can be understood as finding an equilibrium point in the presence of internal and external forces derived from the deformation criterion E_{def} and the data mismatch criterion E_{mis} (Fig. 3).

4.1. Optimization

For the minimization of E_{tot}, it is often the case that E_{tot} is a highly non-linear function and hence contains many spurious local minima. In order for the model to finally converge to the optimal solution, i.e., $\boldsymbol{w}^f = \boldsymbol{w}^*$ where $\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} E_{tot}(\boldsymbol{w}, \alpha; \mathscr{H}_j, \boldsymbol{D}, \bar{\boldsymbol{w}})$, either a local minimizer with good initialization or otherwise a global minimizer is required. Furthermore, coarse-to-fine multi-resolution optimization forms a good trade-off between global and local optimization methods regarding the matching quality and efficiency. This section provides a summary of various related issues.

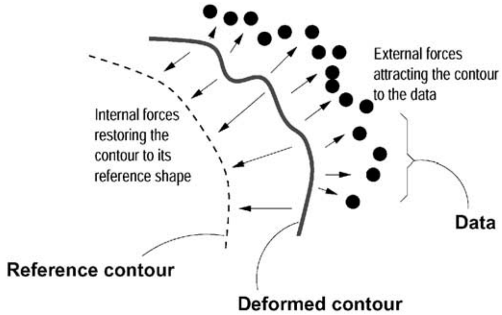

Fig. 3. Illustration of DM matching as force balancing in a physical system.

4.1.1. Initialization

To achieve optimal matching, good model initialization is known to be important (Fig. 4). If user interaction is allowed, the simplest way is to manually place the initial model close enough to the region of interest, e.g., the object boundary. To semi-automate the initialization step, Berger et al. [84], for example, proposed to initialize an active contour model by starting with a short snake placed on the desired boundary. The snake then "grows" towards the two ends to extract the whole boundary. With a similar idea, Neuenschwander et al.'s approach [85] starts with a complete snake whose end points are accurately placed at some desired positions. Then, image forces due to the data come into effect progressively, starting from the two end points towards the middle portion of the snake. Such semi-automatic techniques, though useful for some specific applications, consider only a small part of the model at a time and fail when the noise level is high or when subjective contours exist. For robust and fully automatic initialization,

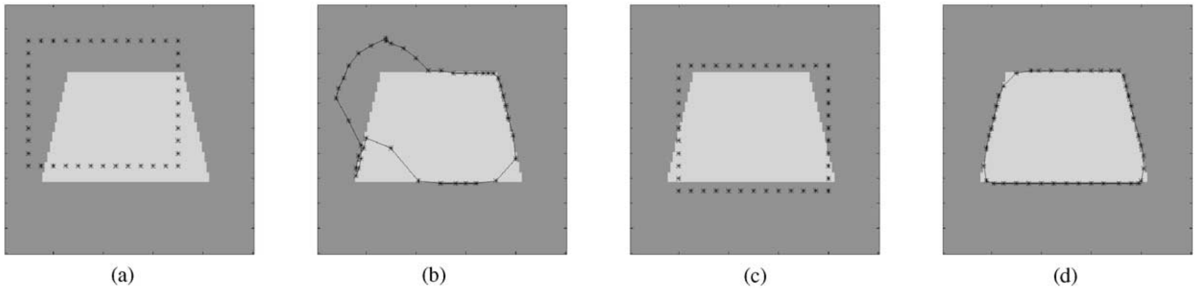

Fig. 4. Effects of different initializations. The object boundary can be successfully extracted only when initialization is good. (a) Bad initialization. (b) Final match of (a). (c) Good initialization. (d) Final match of (b).

rigid feature-based model matching can be used [27,86]. Also, MacCormick et al. [87] demonstrated using random sampling together with a feature-based model to locate objects of interest efficiently, but with the assumption of small degree of scale variation. Other than the feature-based approach, the use of invariant moments has been suggested by Blake et al. [88]. Lai [61] proposed using generalized Hough transform with promising results. Garrido et al. [89] used a modified Hough transform which can tolerate a higher degree of deformation. Williams et al. [90] and Rowley et al. [91] succeeded in training artificial neural networks to achieve reasonably good initialization.

4.1.2. Optimization in continuous domain

Among different continuous optimization methods, steepest descent in the continuous parameter space [3,8,29] is one of the most common ones. Based on Kass et al.'s formulation, it is equivalent to solving the Euler equation [8]. To improve the convergence rate, more complex gradient-based methods like conjugate gradient [3], Newton's method and Levenberg–Marquardt method [92] can also be used. Besides, Powell's direction set method [3] and iterated condition mode [71,72] have also been proposed. The aforementioned methods are for local optimization, where only sub-optimal solutions can be guaranteed. To achieve optimal matching, methods, like simulated annealing [5,72,93] and genetic algorithms [94] have been used.

4.1.3. Optimization in discrete grid domain

If the model parameters are simply a set of x–y coordinates in the image plane, matching algorithms can also be derived based on the discrete grid of the sampled image. By placing search windows at different locations of the model, some discrete local search methods, e.g., greedy search algorithm [20], can be used for criterion function optimization. The advantages of searching in the discrete grid domain include numerical stability and the ease of incorporating hard constraints. Besides, optimizing a criterion with a very complex non-differential functional form in the continuous domain is often a nightmare, but can still be effectively done by some search methods in the discrete grid domain.

To achieve optimal matching, techniques like dynamic programming [21,95] and the A^* algorithm [10] can be used. However, the optimality is limited to the search space defined by the search windows. True optimality can only be obtained if each search window can be extended to the whole image and the resolution of the search window grid is down to pixel level. However, these requirements make the algorithm too computationally expensive to be practical. Good model initialization can solve part of the problem, but the solution optimality and the computational efficiency are still two conflicting factors in determining the size of the search windows.

4.1.4. Multi-resolution optimization

Coarse-to-fine multi-resolution optimization has been found to be more appealing in practice, as on one hand it can escape from many of the local minima and on the other hand the total computational cost required is much less than that of global optimization. The standard way of a multi-resolution optimization is first to create an image pyramid by consecutive sub-sampling and perform optimization from the coarsest level to the finest level where the optimal solution at a particular level is passed to the next level for model initialization. Examples of using the coarse-to-fine paradigm can be found in Refs. [5,61,96,97]. Other than using the standard pyramid, Akgul et al. [98] proposed to use the external energy to segment the images into levels using some efficient algorithm. The discrete grid for the optimization is then formed by the centroids of the segments with similar external energy as the grid points. Superior performance in term of matching optimality has been reported when applied to medical image analysis.

4.2. Constraint incorporation using prior knowledge

The main goal of incorporating (hard) constraints is to limit a DM from deforming to some irrelevant shapes.

From an optimization point of view, adding correct constraints into the search space can avoid a lot of spurious local minima. Deriving such constraints normally requires a priori knowledge specific to applications. In fact, one can consider that carefully choosing a global shape parameterization based on the application-specific knowledge is one *implicit* way to incorporate the related constraints. [14] For explicit ways to incorporate constraints, the deformable template proposed by Yuille et al. [29], which uses parametric curve components for composing an eye template, restricts the upper lid parabola to be always convex upwards and the lower lid to be always convex downwards. Dubuisson et al. [64] used a set of hand-drafted constraints to explicitly restrict the deformation of a deformable polygon for representing cars of different types. Besides, Olstad et al. [99] used syntactical approaches to incorporate constraints, where a priori information about object shape can be encoded into an active contour model as a set of grammar rules. Also, Fua et al. [100] confined the model parameter search in the orthogonal subspaces of some given hard constraint surfaces.

Although adding constraints to avoid sub-optimal matching is well understood to be important for DM-based pattern recognition, there exists another subtle reason for incorporating constraints for classification applications, which may affect the steps for deriving constraints. See Section 5.2 for more details.

5. Classification

5.1. A summary of discriminant measures

Different discriminant measures have been adopted for DM-based classification. In general, they can be categorized as follows.

(*a*) *Ad hoc distance measures*. Using prior knowledge about specific problems, ad hoc distance measures can sometimes be derived with reasonably good discriminating power. For example, Burr's elastic matching [1] for line drawings used a distance measure based on directional and positional incompatibilities between model and data for classification. Sclaroff et al. [32] defined a strain energy, which measures model deformation (equivalent to a model deformation criterion) for classifying the adopted FEM models. A similar idea has also been adopted by Cootes at el. for face recognition [101].

(*b*) *Combined criterion functions*. There is a major difference between an ad hoc distance measure and a combined criterion function. Although both of them can be seen as distance measures, a combined criterion function, besides for classification, also serves as the optimization criterion for matching. This tightly integrates pattern matching and classification into a single step. In fact, it is quite natural and common to use combined criterion functions as discriminant measures [5,27,64], though it is not fully theoretically justified.

(*c*) *Class posterior probabilities*. A theoretically sound and disciplined classification method is by comparing the posterior probabilities $Pr(\mathscr{H}_j|\boldsymbol{D})$ computed for different classes [61,102]. If all the model candidates are assumed to be equally probable, then maximizing $Pr(\mathscr{H}_j|\boldsymbol{D})$ is equivalent to maximizing the likelihood $p(\boldsymbol{D}|\mathscr{H}_j)$. Using the Bayes rule, $p(\boldsymbol{D}|\mathscr{H}_j)$ can be expressed as

$$p(\boldsymbol{D}|\mathscr{H}_j) = \int \frac{p(\boldsymbol{D}|\boldsymbol{w},\mathscr{H}_j)\,p(\boldsymbol{w}|\alpha,\mathscr{H}_j)}{p(\boldsymbol{w}|\boldsymbol{D},\alpha,\mathscr{H}_j)}\,p(\alpha|\mathscr{H}_j)\,\mathrm{d}\alpha. \tag{16}$$

As the exact computation of Eq. (16) is very often either too difficult or too computationally demanding, approximation based on optimal matching and regularization results is generally required [12].

(*d*) *Measures based on discriminative classifiers*. For model-based classification, any inaccuracies in the modeling and criterion function formulation steps can lead to serious performance degradation. Rather than correcting the inaccuracies which sometimes may not be that easy, the values of E_{def} and E_{mis} for different models can be considered as some (possibly noisy) high-level features to be fed into an artificial neural network (ANN) [4,103] or a statistical classifier [104] for subsequent classification. The classification performance is optimized by training the classifier so that the effect of the inaccuracies can be reduced. The limitations are the requirement of large quantity of training data for the classifier and the need for re-training whenever there is a new class.

5.2. Classification accuracy and efficiency

No matter what sophisticated approaches are adopted for pattern classification, the ultimate concerns remain unchanged, namely, accuracy and efficiency.

5.2.1. Accuracy

Using DMs, classification is done by measuring the dissimilarity of an input pattern to all the *optimally deformed* candidate models and identifying the one with the smallest dissimilarity as the output. Difficulties arise if the model of a particular class is "too" flexible and

[14] Global parameterization transforms the original parameter space of a higher dimension (concatenation of coordinates of model instances in the image plane) to a space of a much lower dimension (spanned by a compact set of basis functions) for representation. Such a transformation limits the shapes in the image plane to be within the sub-space or manifold defined by the global shape basis.

can be deformed to shapes that resemble those of the other classes. Taking handwritten digit classification as example, the "5" model can be readily deformed to the shape of "6". So, controlling the flexibility of DMs with the objective to minimize inter-class confusion is one of the most crucial issues to achieve highly accurate classification systems.

The three most important steps related to the flexibility control are: (1) model training, (2) regularization, and (3) constraint incorporation. Model training is used together with a set of training data such that a priori knowledge on the possible variations can be captured (e.g., in the form of a covariance matrix [4]). ML estimation techniques are often used for training, where each model is trained independently. For the case with a limited amount of training data so that accurate parameter estimation cannot be achieved, *discriminative* training methods have been proved to be superior, at least when applied to speech recognition [105,106] and image retrieval [107]. For regularization, adaptive techniques sometimes result in exceedingly high model flexibility for some of the candidate models. Either a proper prior distribution can be imposed on the regularization parameters under a Bayesian framework or some constraints on the value of the regularization parameter obtained from the training data can be used [102]. Using prior knowledge on the easily confused classes, hard constraints can be incorporated accordingly into models in a bruce-force manner. However, generating those constraints sometimes can be very time-consuming, especially when there are a large number of models.

5.2.2. Efficiency

The efficiency of a DM-based recognizer is another major challenge to face. Even though a very efficient matching algorithm can be derived, the recognition time is still linear to the size of the model base as long as sequential machines are used, hindering the practical use of this approach for real-time applications. There are at least two different ways to improve the situation. In Ref. [81], it was shown that significant speedup without much performance degradation can be achieved if some competition process is introduced at the early stage of matching to eliminate some unlikely model candidates. Such a competition process is believed to be a promising direction to eliminate the unnecessary computational power used for matching irrelevant models to an input image. Another approach is to use some indexing algorithms, such as Hough transform [108] and geometric hashing [109]. The approach is known to be efficient for model-based recognition and the general idea is to gain computational advantage by using additional memory to store precomputed values. The development of indexing algorithms has mostly been based on rigid models, where all the object transformations can be solely described by affine transformation. Rigoutsos et al. [110] proposed a modified version which can allow a tiny amount of deformation. For some applications involving non-rigid shapes that can be parameterized by a compact set of parameters, the indexing approach may also apply.

6. Discussions

6.1. Current applications in pattern recognition

DMs are known to be capable of accurately locating and tracking non-rigid objects in noisy images through an image sequence [22,46,72,111–115]. Such properties are found to be especially useful for medical image analysis, e.g., locating and tracking a shape-varying biological object in a sequence of echocardiographic (ECG), magnetic resonance (MR) or X-ray images [47,71,116–119]. Other related applications include precise measurement of some shape-related parameters for medical diagnosis [120] and industrial automatic inspection [121], image registration [122,123], etc.

Object correspondence (also called signal matching) is a non-trivial but important component in many computer vision tasks, like stereopsis, motion analysis, and constructing 3D models from 2D image slices. The difficulties in obtaining the correspondence between a pair of images lie in the fact that the objects of interest can be non-rigid, occluded, or very noisy. Using two 2D DMs for the left and right images, which interact either through a true 3D model [124], affine epipolar geometry [125] or a smoothness constraint on the disparity between them [8,126], promising results on stereo matching have been demonstrated. Other examples of applying DMs to object correspondence include Witkin et al. [96] and Sclaroff et al. [32].

DMs have been applied to the recognition of different non-rigid objects with promising results. The domains include human faces [27,58,101,127,128], gestures [50,72], handwriting [1,4,26,54,81,104,129,130], etc. See Fig. 5 for a handwriting recognition illustration.

6.2. Research problems

McInerney et al.'s survey paper [13] provides very good discussions on research issues related to the continuous development of the DM approach. Although the context is specific to medical image analysis, most of the issues discussed there are common to DM-based pattern recognition regarding the modeling and matching steps. Without repeating them, here we try to focus only on the issues related to the classification step and discuss some related open research problems that are worth future exploration.

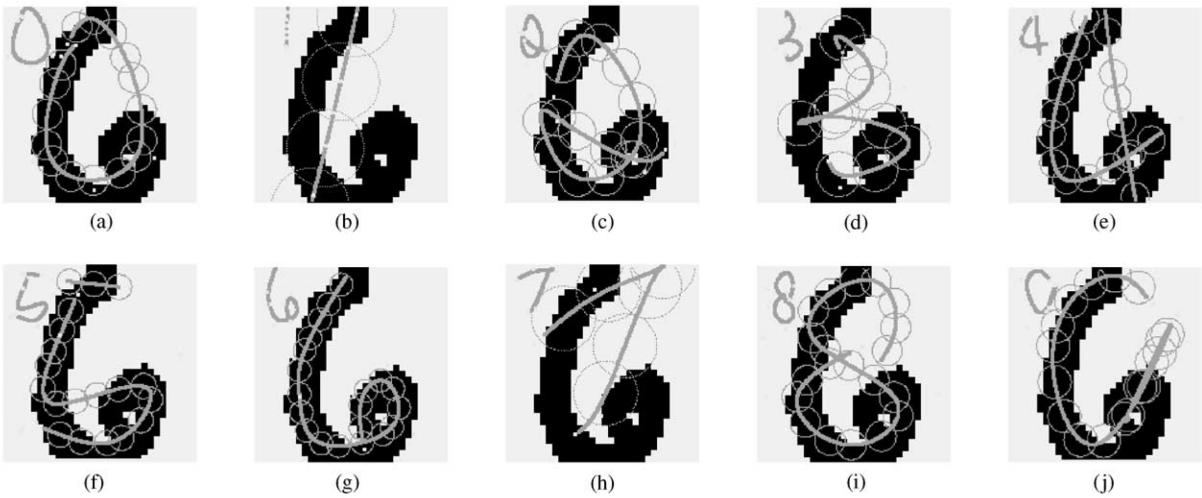

Fig. 5. Handwritten digit recognition using deformable models. The caption MX/ZZZZ under each sub-figure is read as: X-model class and ZZZZ-mismatch measure. For this example, model M6 is found to have the lowest mismatch value and thus the input is classified to be "6". See Ref. [81] for more details. (a) M0/6.39. (b) M1/7.03. (c) M2/6.49. (d) M3/6.76. (e) M4/6.30. (f) M5/6.19. (g) M6/6.06*. (h) M7/7.19. (i) M8/6.48. (j) M9/6.60.

6.2.1. *Automatic model construction and training*

In most DM-related works, the models involved are usually assumed to be constructed manually with special care. For extracting a pre-defined shape from the input image, such a manual process is still tolerable, even though the model is very complicated. However, for classification, at least one model has to be constructed for each class. Such a construction process may not be too trivial. To automate the process, an algorithm for constructing the shape abstraction for each training example is first needed. Also, we need another algorithm for summarizing the set of constructed abstractions based on certain criteria. Related research problems include:

- how to automatically construct the abstraction for an input with unknown topology (examples with different topologies can exist within a class, e.g., digit recognition [102]), where the topology adaptive representation should play an important role;
- how model complexity should be taken into account for constructing optimal shape abstractions with "just-enough" representational power to avoid over- or under-fitting;
- how to automatically summarize the constructed abstractions with possibly different topologies, where some special clustering algorithms should be needed [131];
- what criteria should be used for the summarization (or clustering) such that the resultant set of reference models is optimal for classification;
- can the whole construction process be implemented in an incremental manner so that abstractions created earlier can facilitate the construction process that follows.

6.2.2. *Efficient implementations*

As mentioned in Section 5.2.2, efficient implementations for alleviating the scale-up problem can follow two approaches: (1) using a compete-and-reject process at earlier iterations, and (2) using indexing techniques, where the two approaches are non-exclusive to each other. Related research issues include:

- can the competitive process be enhanced by considering inter-class dependency to achieve improvement in both speed and accuracy (say inter-related classes are grouped for the competition first and then followed by another competition among the classes within the winner group);
- how can the DM-based pattern recognition be implemented using some indexing schemes, where the discretization of the shape parameter space should be a major issue to be studied;
- can some randomized algorithms similar to those used in Ref. [132] be adopted for further boosting the system efficiency.

7. Conclusion

An extensive review of various DM methods for modeling, matching and classifying non-rigid shapes has been presented and compared. Future research directions of the field, particularly on the classification step, are discussed in detail. As of today, the state of the art of DM-based pattern matching is close to full maturity for real-world applications in some confined problem domains, while much more research and experience is still required for the development of DM-based pattern classification.

Acknowledgements

This research work was supported in part by the Hong Kong Research Grants Council (RGC) under Competitive Earmarked Research Grants HKUST 746/96E and HKUST 6081/97E.

References

[1] D. Burr, Elastic matching of line drawings, IEEE Trans. Pattern Anal. Mach. Intell. 3 (6) (1981) 708–713.
[2] T.F. Cootes, C.J. Taylor, A. Lanitis, D.H. Cooper, J. Graham, Building and using flexible models incorporating grey-level information, in: Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany, April 1993, pp. 242–246.
[3] L.H. Staib, J.S. Duncan, Boundary finding with parametrically deformable models, IEEE Trans. Pattern Anal. Mach. Intell. 14 (11) (1992) 1061–1075.
[4] M. Revow, C.K.I. Williams, G.E. Hinton, Using generative models for handwritten digit recognition, IEEE Trans. Pattern Anal. Mach. Intell. 18 (6) (1996) 592–606.
[5] A.K. Jain, Y. Zhong, S. Lakshmanan, Object matching using deformable templates, IEEE Trans. Pattern Anal. Mach. Intell. 18 (3) (1996) 267–278.
[6] B. Widrow, The "rubber-mask" technique—I. Pattern measurement and analysis, Pattern Recognition 5 (1973) 175–197.
[7] M.A. Fischler, R.A. Elschlager, The representation and matching of pictorial structures, IEEE Trans. Comput. 22 (1) (1973) 67–92.
[8] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vision 1 (4) (1987) 321–331.
[9] D. Terzopoulos, Deformable models, Visual Comput. 4 (6) (1988) 306–331.
[10] D.B. Cooper, Maximum likelihood estimation of Markov-process blob boundaries in noisy images, IEEE Trans. Pattern Anal. Mach. Intell. 1 (4) (1979) 372–384.
[11] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum-likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc. Ser. B 39 (1977) 1–38.
[12] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (3) (1992) 415–447.
[13] T. McInerney, D. Terzopoulos, Deformable models in medical image analysis: a survey, Med. Image Anal. 1 (2) (1996) 91–108.
[14] A.K. Jain, Y. Zhong, M.-P. Dubuisson-Jolly, Deformable template models: a review, Signal Process. 71 (1998) 109–129.
[15] W.J. Christmas, J. Kittler, M. Petrou, Structural matching in computer vision using probabilistic relaxation, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 749–764.
[16] C.Y. Liou, H.C. Yang, Handprinted character recognition based on spatial topology distance measurement, IEEE Trans. Pattern Anal. Mach. Intell. 18 (9) (1996) 941–945.
[17] A.R. Pope, Model-based object recognition: a survey of recent research, Technical Report 94-04, Computer Science Department, University of British Columbia, January 1994.
[18] A. Pentland, S. Sclaroff, Closed-form solutions for physically based shape modeling and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 13 (7) (1991) 715–729.
[19] L.D. Cohen, I. Cohen, Finite element methods for active contour models and balloons for 2D and 3D images, IEEE Trans. Pattern Anal. Mach. Intell. 15 (11) (1993) 1131–1147.
[20] D.J. Williams, M. Shah, A fast algorithm for active contours and curvature estimation, CVGIP: Image Understanding 55 (1) (1992) 14–26.
[21] A.A. Amini, T.E. Weymouth, R.C. Jain, Using dynamic programming for solving variational problems in vision, IEEE Trans. Pattern Anal. Mach. Intell. 12 (9) (1990) 855–867.
[22] F. Leymarie, M.D. Levin, Tracking deformable objects in the plane using an active contour model, IEEE Trans. Pattern Anal. Mach. Intell. 15 (6) (1993) 617–634.
[23] K.F. Lai, R.T. Chin, Deformable contours: modeling and extraction, IEEE Trans. Pattern Anal. Mach. Intell. 17 (11) (1995) 1084–1090.
[24] T.R. Chou, W.T. Chen, A stochastic representation of cursive Chinese characters for on-line recognition, Pattern Recognition 30 (6) (1997) 903–919.
[25] Z. Huang, F.S. Cohen, Affine-invariant B-spline moments for curve matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 490–495.
[26] T. Wakahara, Shape matching using LAT and its application to handwritten numeral recognition, IEEE Trans. Pattern Anal. Mach. Intell. 16 (6) (1994) 618–629.
[27] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C.V.D. Malsburg, R.P. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput. 42 (3) (1993) 300–311.
[28] L. Wiskott, J.M. Fellous, N. Krüger, C.V.D. Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.
[29] A.L. Yuille, D.S. Cohen, P. Hallinan, Feature extraction from faces using deformable templates, Int. J. Comput. Vision 8 (2) (1992) 99–111.
[30] H. Chui, A. Rangarajan, A new algorithm for non-rigid point matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 44–51.

[31] T.F. Cootes, C.J. Taylor, Active shape models—Smart snakes, in: Proceedings of the Third British Machine Vision Conference, Leeds, UK, September 1992, pp. 266–275.
[32] S. Sclaroff, A. Pentland, Modal matching for correspondence and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 17 (6) (1995) 545–561.
[33] Y. Wang, L.H. Staib, Boundary finding with correspondence using statistical shape models, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998, pp. 338–345.
[34] T. McInerney, D. Terzopoulos, Topologically adaptable snakes, in: Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, June 1995, pp. 840–845.
[35] A. Perera, C.-L. Tsai, R.Y. Flatland, C.V. Stewart, Maintaining valid topology with active contours: theory and application, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 496–503.
[36] D. DeCarlo, D. Metaxas, Blended deformable models, IEEE Trans. Pattern Anal. Mach. Intell. 18 (4) (1996) 443–448.
[37] D. DeCarlo, D. Metaxas, Shape evolution with structural and topological changes using blending, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1186–1205.
[38] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, Int. J. Comput. Vision 22 (2) (1997) 61–79.
[39] V. Caselles, R. Kimmel, G. Sapiro, C. Sbert, Minimal surfaces based object segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 19 (4) (1997) 394–398.
[40] R. Malladi, J.A. Sethian, B.C. Vemuri, Shape modeling with front propagation: a level set approach, IEEE Trans. Pattern Anal. Mach. Intell. 17 (2) (1995) 158–175.
[41] J. Subrahmonia, D.B. Cooper, D. Keren, Practical reliable Bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants, IEEE Trans. Pattern Anal. Mach. Intell. 18 (5) (1996) 505–519.
[42] B.C. Vemuri, Y. Guo, Snake pedals: geometric models with physics-based control, in: Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, January 1998, pp. 427–432.
[43] A.A. Amini, Y. Chen, R.W. Curwen, V. Mani, J. Sun, Coupled B-snake grids and constrained thin-plate splines for analysis of 2D tissue deformations from tagged MRI, IEEE Trans. Med. Imaging 17 (3) (1998) 344–356.
[44] M. Mignotte, J. Meunier, Deformable template and distribution mixture-based data modeling for the endocardial contour tracking in an echographic sequence, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, June, 1999, pp. 225–230.
[45] D. Terzopoulos, D. Metaxas, Dynamic 3D models with local and global deformations: deformable superquadrics, IEEE Trans. Pattern Anal. Mach. Intell. 13 (7) (1991) 703–714.
[46] D. Metaxas, D. Terzopoulos, Constrained deformable superquadrics and nonrigid motion tracking, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, HI, June 1991, pp. 337–343.
[47] J. Park, D. Metaxas, A. Young, Deformable models with parameter functions: application to heart-wall modeling, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 437–442.
[48] T. O'Donnel, T.E. Boult, S.F. Xi, A. Gupta, The extruded generalized cylinder: a deformable model for object recovery, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 174–181.
[49] I. Cohen, L.D. Cohen, A hybrid hyperquadric model for 2D and 3D data fitting, Comput. Vision Image Understanding 63 (3) (1996) 527–541.
[50] U. Grenander, Y. Chow, D.M. Keenan, Hands: A Pattern Theoretic Study of Biological Shapes, Springer, Berlin, 1991.
[51] M.I. Miller, G.E. Christensen, Y. Amit, U. Grenander, Mathematical textbook of deformable neuroanatomies, Proc. Nat. Acad. Sci. USA 90 (24) (1993) 11 944–11 948.
[52] U. Grenander, Elements of Pattern Theory, John Hopkins University Press, Baltimore, 1996.
[53] G.E. Christensen, R.D. Rabbitt, M.T. Miller, Deformable templates using large deformation kinematic, IEEE Trans. Image Process. 5 (10) (1996) 1435–1447.
[54] A.K. Jain, D. Zongker, Representation and recognition of handwritten digits using deformable templates, IEEE Trans. Pattern Anal. Mach. Intell. 19 (12) (1997) 1386–1390.
[55] B. Moghaddam, C. Nastar, A. Pentland, Bayesian face recognition using deformable intensity surfaces, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 1996, pp. 638–645.
[56] S. Ullman, R. Basri, Recognition by linear combinations of models, IEEE Trans. Pattern Anal. Mach. Intell. 13 (10) (1991) 992–1005.
[57] K. Tanaka, M. Sano, S. Ohara, M. Okudaira, A parametric template method and its application to robust matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 620–627.
[58] A. Lanitis, C.J. Taylor, T.F. Cootes, Automatic interpretation and coding of face images using flexible models, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 743–756.
[59] K. Arbter, Affine-invariant Fourier descriptors, in: J.C. Simon (Ed.), From Pixels to Features, Bonas, France, August 1989, North-Holland, Amsterdam.
[60] M.B. Skouson, Z.-P. Liang, Template deformation constrained by shape priors, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 511–516.
[61] K.F. Lai, Deformable contours: modeling, extraction, detection and classification, Ph.D. Thesis, Electrical and Computer Engineering Department, University of Wisconsin-Madison, 1994.

[62] H. Luo, Q. Lu, R.S. Acharya, R. Gaborski, Robust snake model, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 452–457.

[63] S.C. Zhu, A. Yuille, Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 18 (9) (1996) 884–900.

[64] M.P. Dubuisson, S. Lakshmanan, A.K. Jain, Vehicle segmentation and classification using deformable templates, IEEE Trans. Pattern Anal. Mach. Intell. 18 (3) (1996) 293–308.

[65] A. Chakraborty, M. Worring, J.S. Duncan, On multi-feature integration for deformable boundary finding, in: Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, June 1995, pp. 846–851.

[66] P. Padeva, J. Serrat, E. Martí, A snake for model-based segmentation, in: Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, June 1995, pp. 816–821.

[67] M. Worring, W.M. Smeulders, L.H. Staib, J.S. Duncans, Parameterized feasible boundaries in gradient vector fields, Comput. Vision Image Understanding 63 (1) (1996) 135–144.

[68] P. Fua, Y.G. Leclerc, Model driven edge detection, Mach. Vision Appl. 3 (1) (1990) 45–56.

[69] R. Ronfard, Region-based strategies for active contour models, Int. J. Comput. Vision 13 (2) (1994) 229–251.

[70] A. Chakraborty, L.H. Staib, J.S. Duncan, Deformable boundary finding influenced by region homogeneity, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 624–627.

[71] I.L. Herlin, A deformable region model using stochastic processes applied to echocardiographic image, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, June 1992, pp. 534–539.

[72] C. Kervrann, F. Heitz, A hierarchical statistical framework for the segmentation of deformable objects in image sequences, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 724–728.

[73] C. Chesnaud, P. Réfrégier, V. Boulet, Statistical region snake-based segmentation adapted to different physical noise models, IEEE Trans. Pattern Anal. Mach. Intell. 21 (11) (1999) 1145–1157.

[74] P.J. Huber, Robust Statistics, Wiley, New York, 1981.

[75] A. Yuille, C. Peterson, K. Honda, Deformable templates, robust statistics and Hough transforms, in: Proceedings of the SPIE: Geometric Methods in Computer Vision, Vol. 1570, 1991, pp. 166–173.

[76] K.W. Cheung, D.Y. Yeung, R.T. Chin, Robust deformable matching for character extraction, in: Proceedings of the Sixth International Workshop on Frontiers in Handwriting Recognition, Taejon, Korea, August 1998.

[77] A.N. Tikhonov, Regularization of incorrectly posed problems, Soviet Math. DOKL 4 (1963) 1624–1627.

[78] T. Poggio, V. Torre, C. Koch, Computational vision and regularization theory, Nature 317 (1985) 314–319.

[79] R. Durbin, D. Willshaw, An analogue approach to the travelling salesman problem using an elastic net method, Nature 326 (1987) 689–691.

[80] B. Shahraray, D.J. Anderson, Optimal estimation of contour properties by cross-validated regularization, IEEE Trans. Pattern Anal. Mach. Intell. 11 (6) (1989) 600–610.

[81] K.W. Cheung, D.Y. Yeung, R.T. Chin, Competitive mixture of deformable models for pattern classification, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 1996, pp. 613–618.

[82] M.A. Gennert, A.Y. Yuille, Determining the optimal weights in multiple objective function optimization, in: Proceedings of the Second International Conference on Computer Vision, Tampa, FL, 1988, pp. 87–89.

[83] A.K. Klein, L. Forester, A.A. Amini, Quantitative coronary angiography with deformable spline models, IEEE Trans. Image Process. 16 (5) (1997) 468–482.

[84] M. Berger, R. Mohr, Towards autonomy in active contour models, in: Proceedings of the 10th IAPR International Conference on Pattern Recognition, Atlantic City, NJ, June 1990, pp. 847–851.

[85] W. Neuenschwander, P. Fua, G. Székely, O. Kübler, Initializing snakes, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 658–663.

[86] Y. Amit, A. Kong, Graphical templates for model registration, IEEE Trans. Pattern Anal. Mach. Intell. 18 (3) (1997) 225–236.

[87] J. MacCormick, A. Blake, A probabilistic contour discriminant for object localization, in: Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, January 1998, pp. 390–395.

[88] A. Blake, M. Isard, Active Contour, Springer, London, 1998.

[89] A. Garrido, N. Pérez de la Blanca, Applying deformable templates for cell image segmentation, Pattern Recognition 33 (2000) 821–832.

[90] C.K.I. Williams, M. Revow, G.E. Hinton, Instantiating deformable models with a neural net, Comput. Vision Image Understanding 68 (1) (1997) 120–126.

[91] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 1996, pp. 203–208.

[92] X. Xie, R. Subhakar, H. Zhuang, On improving eye feature extraction using deformable templates, Pattern Recognition 27 (6) (1994) 791–799.

[93] R.P. Grzeszczuk, D.N. Levin, "Brownian strings": segmenting images with stochastically deformable contours, IEEE Trans. Pattern Anal. Mach. Intell. 19 (10) (1997) 1100–1114.

[94] A. Lanitis, Locating facial features using genetic algorithms, in: Proceedings of the International Conference on Digital Signal Processing, Limassol, Cyprus, 1995, pp. 520–525.

[95] D. Geiger, A. Gupta, L.A. Costa, J. Vlontzos, Dynamic programming for detecting, tracking, and matching

deformable contours, IEEE Trans. Pattern Anal. Mach. Intell. 17 (3) (1995) 294–302.

[96] A. Witkin, D. Teropoulos, M. Kass, Signal matching through scale space, Int. J. Comput. Vision 1 (4) (1987) 133–144.

[97] R.P. Würtz, Object recognition robust under translations, deformations, and changes in background, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 769–775.

[98] Y.S. Akgul, C. Kambhamettu, A new multi-level framework for deformable contour optimization, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, June 1999, pp. 465–470.

[99] B. Olstad, A.H. Torp, Encoding of a priori information in active contour models, IEEE Trans. Pattern Anal. Mach. Intell. 18 (9) (1996) 863–872.

[100] P. Fua, C. Brechbühler, Imposing hard constraints on deformable models through optimization in orthogonal subspaces, Comput. Vision Image Understanding 65 (2) (1997) 148–162.

[101] T.F. Cootes, C.J. Taylor, Combining point distribution models with shape models based on finite element analysis, Image Vision Comput. 13 (5) (1995) 403–409.

[102] K.W. Cheung, D.Y. Yeung, R.T. Chin, A Bayesian framework for deformable pattern recognition with application to handwritten character recognition, IEEE Trans. Pattern Anal. Mach. Intell. 20 (12) (1998) 1382–1388.

[103] A.W. Senior, A.J. Robinson, An off-line cursive handwriting recognition system, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 309–321.

[104] T. Hastie, R. Tibshirani, Handwritten digit recognition via deformable prototypes, Technical Report Draft, Statistics and Data Analysis Research Department, AT&T Bell Laboratories, Murray Hill, NJ, USA, July 1992.

[105] W. Reichl, G. Ruske, Discriminative training for continuous speech recognition, in: Proceedings of European Conference on Speech Communication and Technology, Madrid, Spain, September 1995, pp. 537–540.

[106] D. Willett, C. Neukirchen, J. Rottland, Dictionary-based discriminative HMM parameter estimation for continuous speech recognition systems, in: Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, April 1997.

[107] Y. Wu, Q. Tian, T.S. Huang, Discriminant-EM algorithm with application to image retrieval, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 2000, pp. 222–227.

[108] D.H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 13 (1981) 137–145.

[109] Y. Lamdan, H.J. Wolfson, Geometric hashing: a general and efficient model-based recognition scheme, in: Proceedings of the Second International Conference on Computer Vision, Tampa, FL, December 1988, pp. 238–249.

[110] I. Rigoutsos, R. Hummel, A Bayesian approach to model matching with geometric hashing, Comput. Vision Image Understanding 62 (1) (1996) 11–26.

[111] S.J. Dickinson, P. Jasiobedzki, Qualitative tracking of 3D objects using active contour networks, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, Washington, June 1994, pp. 812–817.

[112] S. Benayoun, N. Ayache, I. Cohen, Adaptive meshes and nonrigid motion computation, in: Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 1, Jerusalem, Israel, October 1994, pp. 730–732.

[113] C. Nastar, N. Ayache, Frequency-based non-rigid motion analysis: application to four dimensional medical images, IEEE Trans. Pattern Anal. Mach. Intell. 18 (11) (1996) 1067–1079.

[114] J. Luettin, N.A. Thacker, Speechreading using probabilistic models, Comput. Vision Image Understanding 65 (2) (1997) 163–178.

[115] S. Basu, N. Oliver, A. Pentland, 3D modeling and tracking of human lip motions, in: Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, January 1998, pp. 337–343.

[116] S. Kumar, D. Goldgof, Automatic tracking of SPAMM grid and the estimation of deformation parameters from cardiac MR images, IEEE Trans. Med. Imaging 13 (1) (1994) 122–132.

[117] A. Young, C. Chang, L. Axel, Semi-automatic tracking of myocardial motion in MR tagged images, IEEE Trans. Med. Imaging 14 (3) (1995) 422–433.

[118] G.F. Lea, A. Gupta, The use of hybrid models to recover cardiac wall motion in tagged MR images, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 1996, pp. 625–630.

[119] Y. Kita, Elastic-model driven analysis of several views of a deformable cylindrical object, IEEE Trans. Pattern Anal. Mach. Intell. 18 (12) (1996) 1150–1162.

[120] Y.L. Fok, J. Chan, R.T. Chin, Automated analysis of nerve-cell images using active contour models, IEEE Trans. Med. Imaging 15 (3) (1996) 353–368.

[121] S. Hirai, H. Wakamatsu, K. Iwota, Modeling of deformable thin parts for their manipulation, in: Proceedings of the International Conference on Robotics and Automation, San Diego, CA, May 1994, pp. 2955–2960.

[122] D. Burr, A dynamic model for image registration, Comput. Graphics Image Process. 15 (1981) 102–112.

[123] J. Feldmar, N. Ayache, Locally affine registration of free-form surfaces, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, June 1994, pp. 496–501.

[124] B. Bascle, R. Deriche, Stereo matching, reconstruction and refinement of 3D curves using deformable contours, in: Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany, April 1993, pp. 421–430.

[125] T.J. Cham, R. Cipolla, Stereo coupled active contours, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 1997, pp. 1094–1099.

[126] S. Menet, P. Saint-Marc, G. Medioni, B-snakes: implementation and application to stereo, in: Proceedings of the Seventh Israeli Conference, Ramat Gan, Israel, December 1990, pp. 223–236.

[127] K. Matsuno, S. Tsuji, C.W. Lee, Recognition of human facial expressions using 2D physical model, in: Proceedings of the IEEE Workshop on Visualization and Machine Vision, Seattle, WA, June 1994, pp. 74–82.
[128] C. Nastar, A. Pentland, Matching and recognition using deformable intensity surfaces, in: Proceedings of the IEEE International Symposium on Computer Vision, Coral Gables, FL, November 1995, pp. 223–228.
[129] J.M. Bertille, An elastic matching approach applied to digit recognition, in: Proceedings of the Second International Conference on Document Analysis and Recognition, Tsukuba Science City, Japan, October 1993, pp. 82–85.
[130] W.T. Chen, T.R. Chou, A hierarchical deformation model for on-line cursive script recognition, Pattern Recognition 27 (2) (1994) 205–219.
[131] N. Duta, A.K. Jain, M.-P. Dubuisson-Jolly, Learning 2D shape models, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, June 1999, pp. 8–14.
[132] L. Xu, E. Oja, P. Kultanen, A new curve detection method: randomized Hough transform, Pattern Recognition Lett. 11 (1990) 331–338.

About the Author—KWOK-WAI CHEUNG received his B.Sc. and M.Phil degrees both in Electronic Engineering from the Chinese University of Hong Kong, and his Ph.D. degree in Computer Science from the Hong Kong University of Science and Technology. He was the recipient of the Croucher Foundation Studentship and the Edward Youde Memorial Scholarship in 1991 and 1993, respectively. He is currently an assistant professor in the Department of Computer Science at Hong Kong Baptist University. His current research interests include pattern recognition, machine learning, data mining and intelligent agents. He has served as a paper reviewer for the *IEEE Transactions on System*, *Man and Cybernetic*, *the International Journal of Pattern Recognition and Artificial Intelligence* as well as some other refereed journals and conferences.

About the Author—DIT-YAN YEUNG received his B.Sc. (Eng.) degree in Electrical Engineering and M.Phil. degree in Computer Science from the University of Hong Kong, and his Ph.D. degree in Computer Science from the University of Southern California in Los Angeles. From 1989 to 1990, he was an assistant professor at the Illinois Institute of Technology in Chicago. He is currently an associate professor in the Department of Computer Science at the Hong Kong University of Science and Technology. His current research interests are in the theory and applications of pattern recognition, machine learning, and neural networks. He frequently serves as a paper reviewer for a number of international journals and conferences, including *Pattern Recognition*, *Pattern Recognition Letters*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Image Processing*, and *IEEE Transactions on Neural Networks*.

About the Author—PROFESSOR CHIN received the B.S. with honors and the Ph.D., in Electrical Engineering from the University of Missouri, Columbia. From 1979 to 1981, he was a researcher at NASA Goddard Space Flight Center, Maryland. He was on the faculty of Electrical and Computer Engineering at the University of Wisconsin, Madison from 1981 to 1995; become a full Professor in 1989 and served as Associate Department Chair from 1986 to 1990. Since 1992, he has been a professor of the Computer Science Department of the Hong Kong University of Science and Technology, and is now Department Head. Professor Chin was the recipient of the NSF Presidential Young Investigator Award in 1984. He has served on the editorial board of of the IEEE Transactions on Image Processing and the IEEE Transactions on Pattern Analysis and Machine Intelligence.