Deformable Models

Thomas Albrecht, Marcel Lüthi, Thomas Vetter

University of Basel, Switzerland
{thomas.albrecht, marcel.luethi, thomas.vetter}@unibas.ch

Synonyms

Statistical Models, PCA Models, Active {Contour, Shape, Appearance} Models, Morphable Models

Definition

The term Deformable Model describes a group of computer algorithms and techniques widely used in computer vision today. They all share the common characteristic that they *model* the variability of a certain class of objects (In biometrics this could be the class of all faces, or of all hands, or all eyes, etc.). Today, different representations of the object classes are commonly used. Earlier algorithms modeled shape variations only. The shape, represented as curve or surface, is *deformed* in order to match a specific example in the object class. Later, the representations were extended to model texture variations in the object classes as well as imaging factors such as perspective projection and illumination effects. For biometrics, deformable models are used for image analysis such as face recognition, image segmentation or classification. The image analysis is performed by fitting the deformable model to a novel image, thereby parametrizing the novel image in terms of the known model.

Main Body Text

Introduction

Deformable models denote a class of methods that provide an abstract model of an object class [1] by modeling separately the variability in shape, texture or imaging conditions of the objects in the class. In its most basic form, deformable models represent the shape of objects as a flexible 2D curve or a 3D surface that can be deformed to match a particular instance of that object class. The deformation a model can undergo is not arbitrary, but should satisfy some problem specific constraints. These constraints reflect the prior knowledge about the object class to be modeled. The key considerations are the way curves or surfaces are represented and the different form of prior knowledge to be incorporated. The different ways of representing the curves range from parametrized curves in 2D images, as in the first successful method introduced as Snakes in 1988 [2], to 3D surface meshes in one of the most sophisticated approaches, the 3D Morphable Model (3DMM) [3], introduced in 1999. In the case of Snakes, the requirement on the deformation is that the final deformed curve should be smooth. In the 3DMM, statistical information about the object class (such as *e.g.* the class of all faces) is used as prior knowledge. In other words, the constraint states that the deformed surface should with high probability belong to a valid instance of the object class that is modeled. The required probability distributions are usually derived from a set of representative examples of the class.

All algorithms for matching a deformation model to a given data set are defined as an energy minimization problem. Some measure of how well the deformed model matches the data has to be minimized. We call this the *external energy* that pushes the model to match the data set as good as possible. At the same time the *internal* energy, representing the prior knowledge, has to be kept as low as possible. The internal energy models the object's resistance to be pushed by the external force into directions not coherent with the prior knowledge. The optimal solution constitutes an equilibrium of internal and external forces. For instance, in the case of Snakes, this means that a contour is pushed to an image feature by the external force while the contour itself exhibits resistance to be deformed into a non-smooth curve. In the case of the 3DMM, the internal forces become strong when the object is deformed such that it does not belong to the correct object class.

This concept can be expressed in a more formal framework. In each of the algorithms, a model $\mathcal{M}$ has to be deformed in order to best match a data set $\mathcal{D}$. The optimally matched model $\mathcal{M}^*$ is sought as the minimum of the energy functional E, which is comprised of the external and internal energies E_{ext} and E_{int}:

$$E[\mathcal{M}] = E_{\text{ext}}[\mathcal{M}, \mathcal{D}] + E_{\text{int}}[\mathcal{M}] \tag{1}$$

$$\mathcal{M}^* = \arg\min_{\mathcal{M}} E[\mathcal{M}]. \tag{2}$$

Snakes

In their landmark paper, Kaas et al. [2] introduced Snakes, also known as the Active Contour Model. Here, the deformable model $\mathcal{M}$ is a parametrized curve and the goal is to segment objects in an image $\mathcal{D}$ by fitting the curve to object boundaries in the image. The external energy $E_{\text{ext}}[\mathcal{M}, \mathcal{D}]$ measures how well the snake matches the boundaries in the image. It is expressed in form of a feature image, such as for instance an edge image. If an edge image I with low values on the edges of the image is used, the external energy is given as:

$$E_{\text{ext}}[\mathcal{M}, \mathcal{D}] = E_{\text{ext}}[v, I] = \int_0^1 I(v(s))\, ds, \tag{3}$$

where $v : [0, 1] \to \mathbb{R}^2$ is a suitable parametrization of the curve $\mathcal{M}$ and $I : \mathbb{R}^2 \to \mathbb{R}$ is the edge image of the input image $\mathcal{D}$. If a point $v(s)$ of the curve lies on a boundary, the value of the edge image $I(v(s))$ at this point is low. Therefore, the external energy is minimized if the curve comes to lie completely on a boundary of an image.

The internal energy ensures that the curve always remains a smooth curve. For the classical snakes formulation, it is defined as the spline bending energy of the curve:

$$E_{\text{int}}[\mathcal{M}] = E_{\text{int}}[v] = (\alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2)/2, \tag{4}$$

where α and β control the weight of the first and second derivative terms.

By finding a minimum of the combined functional $E[\mathcal{M}]$, the aim is to find a smooth curve $\mathcal{M}$, which matches the edges of the image and thereby segments the objects present in the image.

The Snake methodology is the foundation for a large number of methods based on the same framework. There are three main lines of development:

- Flexible representation of curves and surfaces
- Incorporation of problem specific prior knowledge from examples of the same object class.
- Use of texture to complement the shape information

Level Set Representation for Curves and Surfaces

The idea of snakes was to represent the curve $\mathcal{M}$ as a parametric curve. While such a representation is simple, it is topologically rigid, i.e. it cannot represent objects that are comprised of a variable number of independent parts. Caselles et al. [4] proposed to represent the curve $\mathcal{M}$ as a level set, *i.e.* the contour is represented as the zero level set of an auxiliary function ϕ:

$$\mathcal{M} = \{\phi = 0\}. \tag{5}$$

A typical choice for ϕ is the distance function to the model $\mathcal{M}$.

This representation offers more topological flexibility, because contours represented by level sets can break apart or join without the need of reparametrization. Additionally, the level set formulation allows a treatment of surfaces and images in any dimension, without the need of reformulating the methods or algorithms. The idea of representing a surface by level-set has led to a powerful framework for image segmentation, which is referred to as *level-set segmentation*.

Example Based Shape Priors

Before the introduction of Active Shape Models [5], the internal energy or prior knowledge of the Deformable Model has been very generic. Independently of the object class under consideration, the only prior knowledge imposed was a smoothness constraint on the deformed model. Active Shape Models or "Smart Snakes" and the 3D Morphable model [3] incorporate more specific prior knowledge about the object class by learning the typical shapes of a specific object class.

The main idea of these methods is to assume that all shapes in the object class are distributed according to a *multivariate normal distribution*. Let a representative training set of shapes $\mathcal{M}_1, \ldots, \mathcal{M}_m$, all belonging to the same object class be given. Each shape $\mathcal{M}_i$ is represented by a vector $\mathrm{x_i}$ containing the coordinates of a set of points. For 2D points (x_j, y_j), such a vector x takes the form $\mathrm{x} = (x_1, y_1, \ldots, x_n, y_n)$. For the resulting example vectors $\mathrm{x}_1, \ldots, \mathrm{x}_m$, we can estimate the mean $\overline{\mathrm{x}}$ and covariance matrix Σ. Thus, the shapes are assumed to be distributed according to the multivariate normal distribution $\mathcal{N}(\overline{\mathrm{x}}, \Sigma)$. To conveniently handle this normal distribution, its main modes of variation, which are the eigenvectors of Σ, are calculated via Principal Components Analysis (PCA) [6]. The corresponding eigenvalues measure the observed variance in the direction of an eigenvector. Only the first k most significant eigenvectors $\mathrm{v}_1, \ldots, \mathrm{v}_k$ corresponding to the largest eigenvalues are used, and each shape is modeled as:

$$\mathrm{x} = \overline{\mathrm{x}} + \sum_{i=1}^{k} \alpha_i \mathrm{v}_i, \tag{6}$$

with $\alpha_i \in \mathbb{R}$. In this way, the estimated normal distribution $\mathcal{N}(\overline{\mathrm{x}}, \Sigma)$ and the prior knowledge it contains about the object class is used to define the internal energy. Indeed, looking at Equation (6), we see that the shape can only be deformed by the principal modes of variation of the training examples.

Furthermore, the coefficients α_i are usually constrained, such that deformations in a direction of v_i are not much larger than those observed in the training data. For the Active Shape Model, this is achieved by introducing a threshold D_{max} on the mean squares of the coefficients α_i, scaled by the corresponding standard deviation σ_i of the training data. The internal force of the Active Shape Model is given by:

$$E_{\text{int}}[\mathcal{M}] = E_{\text{int}}[\alpha_1, \ldots, \alpha_k] = \begin{cases} 0 & \text{if } \sum_{i=1}^{k} (\frac{\alpha_i}{\sigma_i})^2 \leq D_{\text{max}} \\ \infty & \text{else.} \end{cases} \tag{7}$$

In contrast, the 3D Morphable Model [3] does not strictly constrain the size of these coefficients. Rather, the assumed multivariate normal distributions $\mathcal{N}(\overline{x}, \Sigma)$ is used to model the internal energy of a deformed model $\mathcal{M}$ as the probability of observing this model in the normally distributed object class:

$$E_{\text{int}}[\mathcal{M}] = E_{\text{int}}[\alpha] = -\ln P(\alpha) = -\ln e^{-\frac{1}{2}\sum\limits_{i=1}^{k}(\alpha_i/\sigma_i)^2} = \frac{1}{2}\sum_{i=1}^{k}(\alpha_i/\sigma_i)^2. \tag{8}$$

Correspondence and Registration

All deformable models using prior knowledge in form of statistical information presented here assume the example data sets to be *in correspondence*. All objects are labeled by the same number of points and corresponding points always label the same part of the object. For instance in a shape model of a hand, a given point could always label the tip of the index finger in all the examples. Without this correspondence assumption, the resulting statistics would not capture the variability of features of the object but only the deviations of the coordinates of the sampled points. The task of bringing a set of examples of the same object class *into correspondence* is known as the *Registration Problem* and constitutes another large group of algorithms in computer vision.

Incorporating Texture Information

One shortcoming of the classical Snake model is that the information of the data set $\mathcal{D}$ is only evaluated at contour points of the model $\mathcal{M}$. In level-set segmentation, new external energy terms have been introduced in [7] and [8]. Instead of measuring the goodness of fit only by the values of the curve $\mathcal{M}$ on a feature image, in these new approaches the distance between the original image and an approximation defined by the segmentation is calculated. Typical approximations are images with constant or smoothly varying values on the segments. This amounts to incorporating the prior knowledge that the appearance or texture of the shape outlined by the deformable model is constant or smooth.

By incorporating more specific prior knowledge about the object class under consideration, the appearance or texture can be modeled much more precisely. This can be done in a similar fashion to the shape modeling described in the previous section. The appearance or texture $\mathcal{T}$ of a model $\mathcal{M}$ is represented by a vector T. All such vectors belonging to a specific object class are assumed to be normally distributed. For instance, it is assumed that the texture images of all faces can be modeled by a multivariate normal distribution. Similar to the shapes, these texture vectors need to be *in correspondence* in order to permit a meaningful statistical analysis.

Given m example textures $T_1, \ldots, T_m$, which are in correspondence, their mean $\overline{T}$, covariance matrix Σ_T, main modes of variation $t_1, \ldots, t_k$, and eigenvalues ρ_i can be calculated. Thus, the multivariate normal distribution $\mathcal{N}(\overline{T}, \Sigma_T)$ can be used to model all textures of the object class, which are then represented as:

$$T = \overline{T} + \sum_{i=1}^{k} \beta_i t_i. \tag{9}$$

A constraint on the coefficients β_i analogous to Equation (7) or (8) is used to ensure that the model texture stays in the range of the example textures. In this way, not only the outline or shape of an object from the object class but also its appearance or texture can be modeled. The Active Appearance Models [9, 10, 1] and the 3D Morphable Model [3] both use a combined model of shape and texture in order to model a specific object class. A complete object is modeled as a shape given by Equation (6) with texture given by Equation (9). The model's shape and texture are deformed by choosing the shape and texture coefficients $\alpha = (\alpha_1, \ldots, \alpha_k)$ and $\beta = (\beta_1, \ldots, \beta_k)$. The external energy of the model is defined by the distance between the input data set $\mathcal{D}$ and the modeled object (S, T), measured with a distance measure which does not only take the difference in shape but also that in texture into account. The internal energy is given by Equation (7) or (8) and the analogous equation for the β_i .

2D versus 3D Representation

While the mathematical formalism describing all previously introduced models is independent of the dimensionality of the data, historically the Active Contour, Shape, and Appearance Models were only used on 2D images, whereas the 3D Morphable Model was the first model to model an object class in 3D. The main difference between 2D and 3D modeling is in the expressive power and the difficulty of building the deformable models. Deformable models, when incorporating prior knowledge on the objects class, are derived from a set of examples of the object class. In the 2D case these examples are usually registered images showing different instances of the object class. Similarly, 3D models require registered 3D examples of the object class. As an additional difficulty, 3D examples can only be obtained with a complex scanning technology, *e.g.* CT, MRI, laser, or structured light scanners. Additionally, when applied to images the 3D models require a detailed model for the imaging process such as the simulation of occlusions, perspective, or the effects of variable illumination.

While building 3D models might be difficult, 3D models naturally offer a better separation of object specific parameters from parameters such as pose and illumination that originate in the specific imaging parameters. For 2D models these parameters are often extremely difficult to separate. For instance, with a 2D model, 3D pose changes can only be modeled by shape parameters. Similarly, 3D illumination effects are modeled by texture variations.

Applications

Deformable Models have found a wide range of applications in many fields of computer science. For biometrics, the most obvious and well-researched applications are certainly face tracking ([10], Figure 1) and face recognition, [11]. For face recognition, the fact is exploited that an individual face is represented by its shape and texture coefficients. Faces can be compared for recognition or verification by comparing these coefficients.

Another important area in which Deformable Models have found application is in medical image analysis, most importantly medical image segmentation, see [12], Figure 2 for instance.

Recent Developments

While the level-set methods allowing for greater topological flexibility, the Active Appearance Model and the 3D Morphable model in turn provide an internal energy term representing prior knowledge about the object class. It is natural to combine the advantages of all these methods by using the level-set representation and its resulting external energy term together with the internal energy term incorporating statistical prior knowledge. In [12], Leventon et al. propose such a method that relies on the level-set representation of snakes introduced by Caselles et al. [4]. The internal energy is given by statistical prior knowledge computed directly from a set of level-set functions (distance functions) representing the curves using a standard PCA approach.

Summary

Deformable models provide a versatile and flexible framework for representing a certain class of objects by specifying a model of the object together with its variations. The variations are obtained by deforming the model in accordance to problem specific constraints the deformation has to fulfill. These constraints represent the prior knowledge about the object and can range from simple smoothness assumption on the deformed object to the requirement that the resulting object still belongs to the same object class. The analysis of novel objects is done by fitting the deformable model to characteristics of a new object. The fitting ranges from simple approaches of matching the object's boundary in an image, to optimally matching the object's full texture. Because of their flexibility, deformable models are used for many applications in biometrics and the related fields of computer vision and medical image analysis. Among others, the most successful use of these models are in automatic segmentation and image analysis and synthesis (see figure 3).

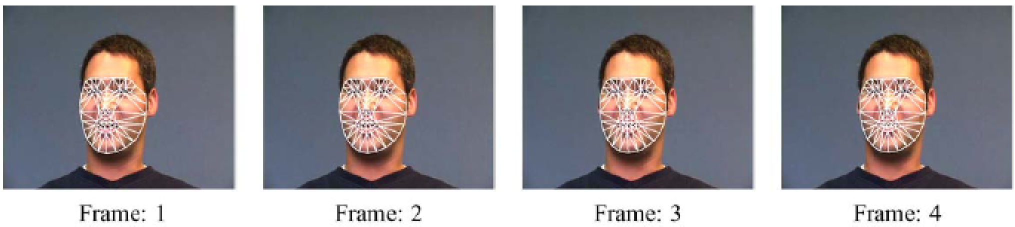

Fig. 1. Tracking a face with the active appearance model. Image from [10].

Related Entries

- Face Alignment
- Face Recognition, Overview
- Image Pattern Recognition
- Image Preprocessing

References

1. Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7) (1997) 733–742
2. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision **1**(4) (1988) 321–331
3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press (1999) 187–194
4. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. International Journal of Computer Vision **22**(1) (1997) 61–79
5. Cootes, T., Taylor, C.: Active shape models-'smart snakes'. Proc. British Machine Vision Conference **266275** (1992)
6. Bishop, C.: Pattern recognition and machine learning. Springer (2006)
7. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2) (2001) 266–277
8. Mumford, D., Shah, J., for Intelligent Control Systems (US, C.: Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. Center for Intelligent Control Systems (1988)
9. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6) (2001) 681–685

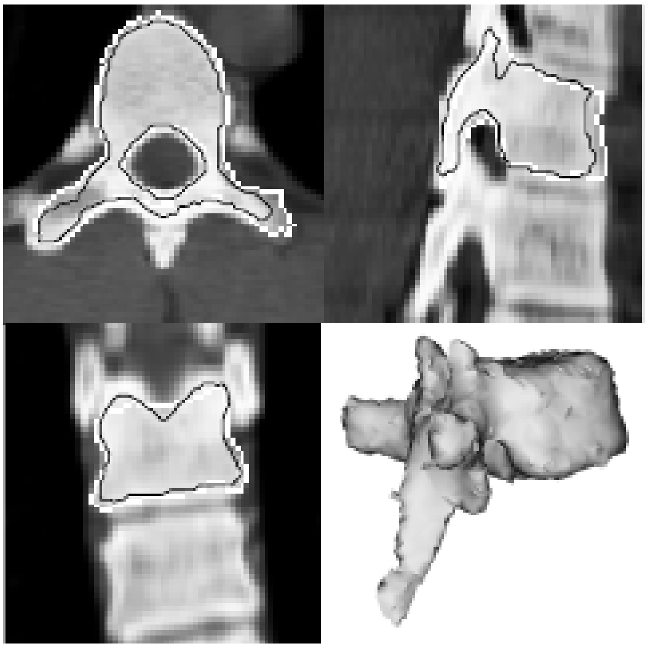

Fig. 2. 3D level set segmentation with shape prior of a vertebrae. Image from [12].

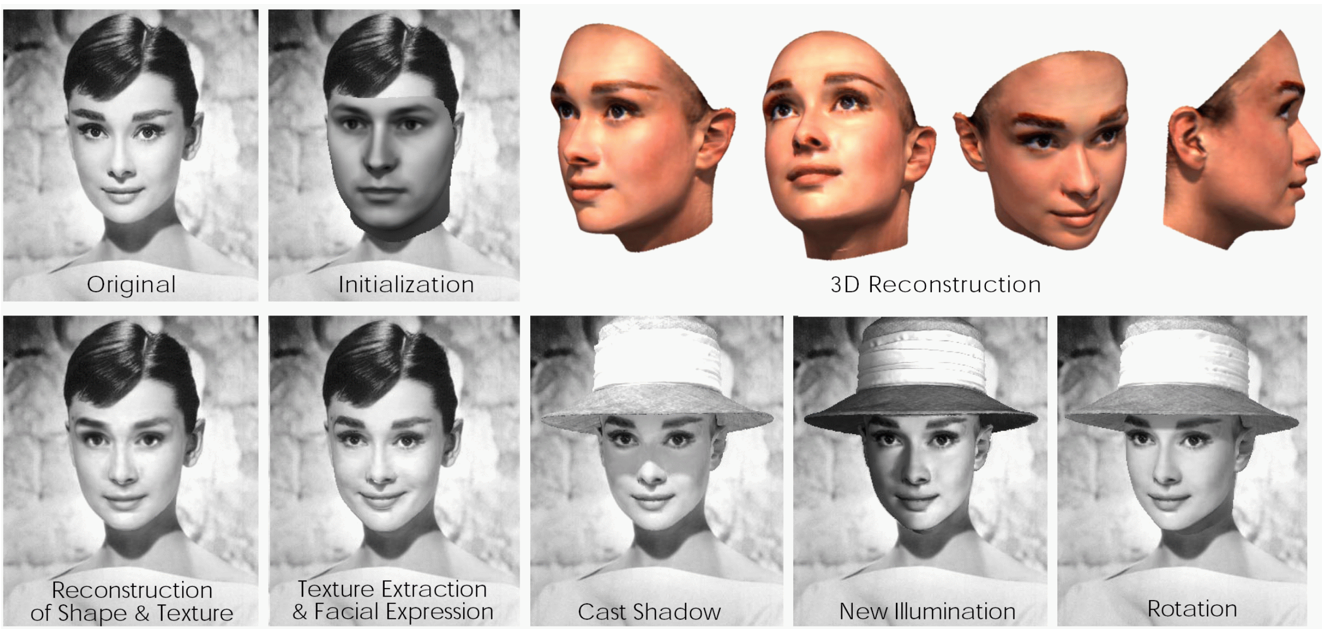

Fig. 3. The 3D Morphable Model: A 3D reconstruction of a 2D image is performed using the 3D Morphable Model. Parameters of the model, such as facial expression, illumination and perspective are modified and the result is rendered back into the image. Image from [3].

10. Matthews, I., Baker, S.: Active Appearance Models Revisited. International Journal of Computer Vision **60**(2) (2004) 135–164
11. Blanz, V., Vetter, T.: Face recognition based on fitting a 3 D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(9) (2003) 1063–1074
12. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on **1** (2000)

Definitional Entries

Principal Components Analysis

Principal Component Analysis (PCA) is a common technique in data analysis. PCA performs an orthogonal basis transform in a linear space, such that the data in the transformed space becomes uncorrelated. A crucial property of PCA is that after such a transformation, the largest variance of the projection of the data in any possible direction is explained by the coordinate vector corresponding to the first basis vector. Similarly, the second largest variance is explained by the coordinate vector corresponding to the second basis vector, etc. It is this property that makes PCA a common tool for dimensionality reduction and data compression.

Segmentation

Segmentation refers to the process of partitioning a data set into several parts, where each part corresponds to a semantic entity of the data set. Semantic entities might be word boundaries in natural language processing, lexical units in scripts, or objects in images. Segmentation is usually performed as a pre-processing step, to simplify the understanding and interpretation of the data. In Biometrics, a common task is to perform a segmentation of a face image into regions for skin, eyes, mouth and background, before further analysis of the individual parts is carried out.

Registration

The registration problem consists of identifying corresponding points in two different objects, such that comparison of the objects can be performed. Given two objects $\mathcal{M}^1, \mathcal{M}^2 \subset \Omega \subset \mathbb{R}^n$, it is usually formulated as finding a mapping $\phi : \Omega \to \Omega$ between the coordinate systems in which the objects $\mathcal{M}_1$ and $\mathcal{M}_2$ are defined. More formally, given a distance measure $\mathcal{D}$, the mapping ϕ is sought which minimized $\mathcal{D}(\mathcal{M}_1, \mathcal{M}_2 \circ \phi)$. The correspondence point in object $\mathcal{M}_1$ becomes the closest point in the mapped object $\mathcal{M}_2 \circ \phi$. The object domain Ω is usually chosen to be a coordinate vector of regularly sampled points of the objects.

As an example assume that two images of human faces are given. It is tried to find a mapping such that the corner of the mouth, the tip of the nose as well as other facial features are identified by the same coordinate location and hence comparison of these features in the two images becomes feasible.