Deformable Template Methods for Object Detection

Yali Amit[1]

Department of Statistics

University of Chicago

Email: amit@marx.uchicago.edu.

November 5, 1998

[1]Supported in part by the Army Research Office under grant DAAH04-96-1-0061 and MURI grant DAAH04-96-1-0445,

I. Introduction

Numerous computer vision problems consist of finding a certain object in an image. Find the face in a photograph, find the road in a satellite image, find the ventricles in an axial MRI image of the brain. Often it is not guaranteed that the object is present in the image, or there may be more than one instance. The notion of 'find' is also very vague. One could require a precise outline of the object in the image, one could require the detection of a certain number of well defined landmarks on the object. One could define 'finding' as identifying a deformation from a model or prototype of the object into the image. The deformation could be a simple 2d affine map or a high dimensional warp. The object itself may have different degrees of variability. It may be a rigid 2d object, such as a fixed computer font, or it may be a highly deformable object such as the left ventricle of the heart.

In these notes we describe 4 different families of algorithms: One dimensional elastic contours (also called active contours or snakes) in section II; Two dimensional elastic image matching in section III; Curve detection with semi-global methods in section IV; Sparse global model matching in sections V, and VI. The sections are self contained and can be read independently. The next section provides some notation and an attempt at an abstract unifying formulation of of the algorithms.

I.1 Abstract statistical formulation

The deformable template model

- Reference grid R centered at the origin.

- Image lattice L centered at the origin. $R \subset L$.
- Image - $I(x), x \in L$.
- **Transformed image:**

$$\begin{aligned} FI(x) &= (X_1(x), \ldots, X_k(x)) \\ X_k(x) &= X_k(I_{N(x)}),\ k = 1, \ldots, K \end{aligned} \tag{I.1}$$

  $I_{N(x)}$ is the image data in the neighborhood of x.

  $X_k(I_{N(x)})$ - some function of the image data in $N(x)$.

- **Template:** $T(x) = (T_1(x), \ldots, T_K(x)), x \in R$

  'Mean' of transformed images of the object class (equation I.1) - 'registered' to reference grid.

- **Deformations:** $\mathcal{G}$ - set of smooth one to one mappings of R into L - 'admissible' geometric deformations of template.

  gR - image of reference grid in image lattice L under $g \in \mathcal{G}$.

- The object is present in the image - there exists $g \in R$ such that the image data is distributed as follows:
  - In gR: $[FI(x); x \in gR]$ has a joint distribution $D_{obj}\Big(g, T(g^{-1}x), x \in gR\Big)$

    FI on gR is a random degradation of the

    deformed template - $T(g^{-1}x), x \in gR$.
  - In $(gR)^c$: $[FI(x); x \notin gR]$ has a joint distribution depending on g and generic characteristics of non-object images. $D_{bgd}(g, FI(x), x \notin gR)$.
  - Assume the two parts are independent.

Object is **present at deformation** $g \in \mathcal{G}$.

- Models for multiple objects can be formulated similarly.

- Hard to give precise formulation of $\mathcal{G}$. Hard to provide reliable models for D_{obj} and in particular for D_{bgd}, which are computationally tractable.

  Easier to deal with subsets of R for defining deformations and data models.

Detection as instantiation

- Use $Z = \{z_1, \ldots, z_n\} \subset R$ instead of all of R.

  z_i - 'landmarks' on 'registered' images of object.

- Θ - set of possible **instantiation** parameters. $\theta \in \Theta$ defines a mapping of Z into the image grid L.

  θZ - denotes vector $(\theta z_1, \ldots, \theta z_n) \subset L^n$.

- If object is present at deformation $g \in \mathcal{G}$ then $\theta z_i = g z_i, i = 1, \ldots, n$ -

  Object is **present at instantiation** θ.

  The n 'landmarks' are present at $\theta z_i, i = 1, \ldots, n$.

  Could be more than one such $\theta \in \Theta$.

- Possible to use $\Theta = \mathcal{G}$.

  At times more convenient to use 'larger' Θ which is easier to define and to compute with.

  $\{gZ : g \in \mathcal{G}\} \subset \{\theta Z : \theta \in \Theta\}$.

  T restricted to Z is now the **restricted template**.

  It is **deformed** through elements in Θ.

- A landmark $z_i \in Z$ may be associated with some labeled component of the object or may simply have some desirable statistical properties in terms of the data around θz_i for an instantiation θ of the object. Other labeled components of the object may be identified *relative* to the points in the set θZ.

- **Detection:** finding $\theta^{(1)}, \dots, \theta^{(k)} \in \Theta$, ($k$ may be 0), which map Z into the image: k instantiations of the object in the image.

The Bayesian model

- $I(x), x \in L$ denotes the image data. FI local transform on image as above.

- Define $L(\theta) \subset L$ a union of neighborhoods of the points $\theta z_i, i = 1, \dots, n$.

- Given an instantiation θ of the object define

$$\begin{aligned} P(FI(x), x \in L|\theta) &= \prod_{x \in L(\theta)} P(\theta, FI(x)) \prod_{x \notin L(\theta)} P_0(FI(x)). \\ &\propto \prod_{x \in L(\theta)} \frac{P(\theta, FI(x))}{P_0(FI(x))} \end{aligned} \tag{I.2}$$

$P(\theta, FI(x))$ - conditional distribution of $FI(x)$ given the instantiation θ for $x \in L(\theta)$.

$P_0(FI(x))$ - conditional distribution of $FI(x)$ given $x \notin L(\theta Z)$. This term does not depend on θ.

Underlying assumptions: *Conditional on the presence of an object at instantiation θ, the transformed data at the different pixels is independent.*

Conditional on no object present the transformed data at the different pixels is independent and identical.

- Together with a prior on Θ - posterior distribution:

$$P(\theta|FI) \propto P(\theta) \prod_{x \in L(\theta Z)} \frac{P(\theta, FI(x))}{P_0(FI(x))} \tag{I.3}$$

- Prior can be uniform distribution on Θ; or can incorporate a penalty on deviations from the model configuration Z.

- When one object is known to exist in the image - maximize the posterior in θ.

- When more than one instance - find peeks of the posterior at points in Θ. Any point above some threshold flags a candidate instantiation. Further processing determines if indeed an object is present or not.

Related work

Work relating to each one of the specific algorithms will be mentioned throughout the notes.

The generic model formulation is closely related to 'Pattern Theory' as presented in the first chapters of Grenander (1993) although the computational and statistical emphasis is very different.

Instantiation as matching of points in the model is also related to the *alignment* methods described in Ullman (1996), and other computer vision literature and the *inexact consistent-labeling problem* of Haralick & Shapiro (1992) although here the emphasis is much more statistical and the algorithms are entirely model driven. There is no apriori bottom up processing or 'segmentation' which is unrelated to the specific object being detected.

I.1.1 Overview of the models

The different models involve variations in the definition of the set Z, the set of deformations Θ^0 and the prior distribution defined on that set; the image transforms F together with the conditional probability of the data given an element $\theta \in \Theta$ and finally the computational algorithm for maximizing the posterior. These components are very tightly interlinked. The type of algorithm chosen may constrain the types of data models as well as the definitions of the set Z and Θ.

User initialization

- **1-d elastic models: contour detection.** Section - II.

  Z is a closed curve in R.

  Θ - family of smooth perturbations of Z.

  Initial point θ - provided by the user.

  Optimization - gradient descent methods,

  Data term - 'Inside-outside' model, edge model.

- **contour detection: semi-global methods.** Section - IV.

  Z - open or closed curve.

  $\Theta \subset L^n$ - defined directly in terms of constraints on locations of θz_i.

  Prior: Penalize geometric deviation of $(\theta z_1, \ldots, \theta z_n)$ from $(z_1, \ldots, z_n)$.

  FI - edge type point process: X_k - binary features.

  Initialization - θ_1, θ_n.

  Optimization - dynamic programming. For a very efficient tree based algorithm (see Geman & Jedynak (1996)).

- **2-d elastic models: image matching.** Section - III.

  $Z = R$.

  Θ - smooth deformations of the reference grid.

  Initial θ provided by user.

  FI - the image itself.

  Data model - only 'on object': additive i.i.d noise.

  Optimization is again done with gradient descent methods.

Decomposable models. Section - V

Z - small set of points.

$\Theta \subset L^n$ - defined directly in terms of admissible n-tuples $(\theta z_1, \ldots, \theta z_n)$, denoted $(\theta_1, \ldots, \theta_n)$.

Prior - defined in terms of triples of points in Θ - decomposable.

FI - small number of low density point processes, in terms of relational local operators.

Optimization - dynamic programming.

Detection of multiple instances in the presence of occlusion. Section - VI

Z - small set of points.

Collection of *sub-instantiations*: Defined directly as $\Theta_m \subset L^{n_m}, m = 1, \ldots, M$.

FI - low density point processes: estimated from training data.

Optimization - 'Efficient Coarse to fine' implementation of the generalized Hough transform

Merging the models. Section - VI.4.1

View these models and algorithms as coarse to fine computational components of one global model.

I.2 Important issues

I.2.1 Invariance

- Photometric transformations: smooth monotone variations in gray level values.
- Geometric transformations: variations in instantiation parameters. *Scale, translation, partial rotation, and other types of non-linear deformations.*

  Translation and scale - important components of object description.

Often regarded as nuisance parameters which need somehow to be estimated.

Computer vision literature often suggests apriori ‘image segmentation’.

In this manuscript these parameters are taken as an integral part of the detection problem and the model. The degree to which an algorithm can *efficiently* estimate these parameters is explicitely discussed.

There are several ways to deal with invariance to possible transformations of the data and the object.

- Normalization:
  - Generic normalization of data: scaling, gray level normalization etc: transform data into reference pose. Assumes object is isolated in image, computationally intensive.
  - Model based normalization: **Deformable templates** - find the instantiation, then normalize using instantiation information. This is the approach implemented here.
- Full invariance: find fully invariant data transforms which identify object. Virtually impossible due to the immense variablity and noisy nature of the images.
- Partial invariance: find data transforms which are partially invariant, or invariant in ‘distribution’, in large subregions of instantiation space.

I.2.2 Statistical model:

- Statistical models: systematic way of estimating parameters of model, merging of multiple sources of information.

  Note: some models are too complex to allow for reliable and robust estimation.

- Most cost functions have a statistical interpretation. Every posterior distribution leads to a cost function which needs to be maximized.
- Often independence assumptions are incorporated in model to make it tractable.

  In some cases these assumptions are more credible than in others.

I.2.3 Computation

- Is *user* initialization required?

  How sensitive is the algorithm to initialization?
- Relaxation techniques: problematic in the face of multiple modes of cost function. Intensive computationally.
- Global optimization: typically computationally intractable.

  For certain types of models efficient approximations exist: dynamic programming, Hough transform.
- Coarse to fine: computations are usually made more efficient with some form of coarse to fine procedure.