II. 1-d elastic models - curve detection

II.1 Introduction

- Many imaging problems involve detecting the closed boundary of an object: tumor in xray or MRI, heart ventricle in ultrasound.
- Difficult to obtain through standard bottom up edge detection and contour tracing. See e.g. edge map of left heart ventricle in figure II.1.
- Introduce global constraints: 'ideal' contour closed and continuous.
- Place a closed contour on image surface and let it deform continuously to adjust to the shape of the contour of the desired object.
- Prior knowledge on shape represented in initial contour, and constraints on deformations.
- Grenander (1970), Grenander et al. (1991), Kass et al. (1987).

II.2 Variational formulation

Most of the literature on these models in not formulated from a statistical point of view. Moreover the corresponding statistical assumptions are rather questionable. We therefore start with a non-statistical formulation.

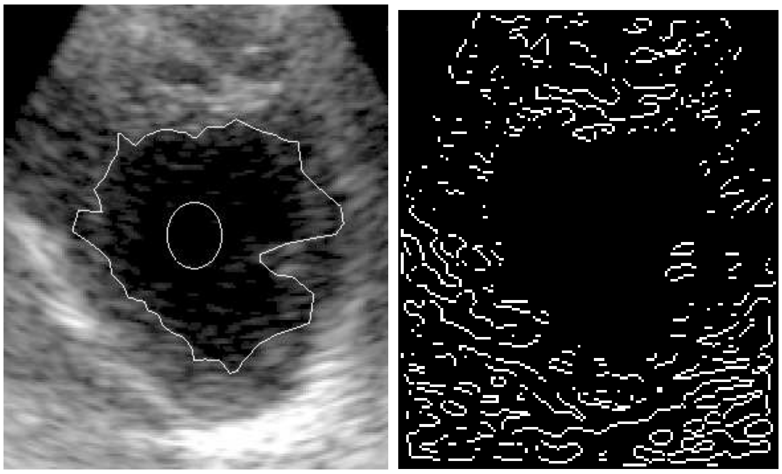

Figure II.1: Left: An ultrasound image of the left heart ventricle with the outcome of an elastic contour initialized at the small circle shown in the center, Right: detected edges.

- Z - closed contour in R.

  Θ - smooth perturbations of Z.

  In the continuum: $Z(t), t \in [0,1]$ closed contour in unit square R.

  $\gamma(t,\theta) = \theta Z(t), t \in [0,1]$ - family of contours (counter clockwise.)

  $\gamma(t,\theta) = [\gamma_1(t,\theta_1), \gamma_2(t,\theta_2)]$.

- Cost function:

$$J(\theta) = \mathcal{E}(\theta) + \mathcal{D}(\gamma(\theta), I), \tag{II.1}$$

  $\mathcal{E}$ - *penalty term* on values of θ.

  $\mathcal{D}$ - *data term*: fit between $\gamma(\theta)$ and image I.

II.2.1 Inside-Outside model

- Data term:

$$\mathcal{D}(\theta) = \left(\int_{\gamma(\theta)_{in}} G_{in}(x)dx + \int_{\gamma(\theta)_{out}} G_{out}(x)dx \right), \tag{II.2}$$

$G_{in}(x), G_{out}(x)$ - functions of image data in neighborhood of x. $G_{in}(x)$ - small if data around x consistent with the model for the data inside the object. Similarly for G_{out}.

- Example: $G_{in}(x) = (I(x) - \mu_{in})^2$, $G_{out}(x) = (I(x) - \mu_{out})^2$.

  μ_{in} mean value inside, μ_{out} mean value outside.

II.2.2 Curve parameterization and penalty

- Expand components of curve in basis of functions on interval $[0, 1]$, truncated at some level d.

  θ_1, θ_2 - basis coefficients. (Fourier, wavelet, wavelet packet bases, splines)

- Finite expansion:

$$\gamma_j(t, \theta_j) = \sum_{k=1}^{d} \theta_{j,k} \psi_k(t) \tag{II.3}$$

  for $j = 1, 2$.

- Penalty:

$$\mathcal{E}(\theta) = \alpha \sum_k \lambda_k (\theta_{1,k}^2 + \theta_{2,k}^2).$$

  Basis functions of higher index - high frequency or less smooth - λ_k larger.

- Penalty can be viewed as Gaussian prior on the space of curves.

II.2.3 Gradient of cost function

- (For derivation see appendix)

$$\frac{\partial \mathcal{D}}{\partial \theta_{1,k}} = \int_0^1 \Big(G_{out}(\gamma(t, \theta)) - G_{in}(\gamma(t, \theta)) \Big) \frac{\partial \gamma_1(t, \theta)}{\partial \theta_{i,k}} \dot{\gamma_2} dt \tag{II.4}$$

$$\frac{\partial \mathcal{D}}{\partial \theta_{2,k}} = \int_0^1 \Big(G_{in}(\gamma(t,\theta)) - G_{out}(\gamma(t,\theta)) \Big) \frac{\partial \gamma_2(t,\theta)}{\partial \theta_{i,k}} \dot{\gamma}_1 dt. \tag{II.5}$$

- For basis function expansion:

$$\frac{\partial \mathcal{D}}{\partial \theta_{1,k}} = \int_0^1 \Big(G_{out}(\gamma(t,\theta)) - G_{in}(\gamma(t,\theta)) \Big) \dot{\gamma}_2 \psi_k(t) dt \tag{II.6}$$

$$\frac{\partial \mathcal{D}}{\partial \theta_{2,k}} = \int_0^1 \Big(G_{in}(\gamma(t,\theta)) - G_{out}(\gamma(t,\theta)) \Big) \dot{\gamma}_1 \psi_k(t) dt. \tag{II.7}$$

- $\frac{\partial \mathcal{D}}{\partial \theta_{1,k}}, k = 1, \ldots, d$ - coefficients of $(G_{in} - G_{out})(\gamma(t,\theta))\dot{\gamma}_2$,

  $\frac{\partial \mathcal{D}}{\partial \theta_{2,k}}, k = 1, \ldots, d$ - coefficients of $(G_{in} - G_{out})(\gamma(t,\theta))\dot{\gamma}_1$,

  in the basis $\psi_k, k = 0, \ldots, d$.

- Assume ψ_k of local support.

  $\theta_{1,k}, \theta_{2,k}$ determine curve around t: $-\dot{\gamma}_2, \dot{\gamma}_1$ - *inward* pointing normal.

  If $G_{in} > G_{out}$ minimization pulling point $\gamma(\theta, t)$ *outward* along normal.

  Locally inside of curve fits data better than outside.

II.2.4 Adding an edge term to the data model

- Consistency of curve with edges in its neighborhood.

$$\mathcal{D}(\theta) = -\int_0^1 \left(-I_2(\gamma)\dot{\gamma}_1 + I_1(\gamma)\dot{\gamma}_2\right)\left(1 + I_1^2 + I_2^2\right)^{-1/2}(\gamma) dt, \tag{II.8}$$

I_1, I_2 are the partial derivatives of I.

Integrand is positive if the normalized gradient of the image has similar direction to the outward normal.

- The gradient of $\mathcal{D}$:

$$\frac{\partial \mathcal{D}}{\partial \theta_{1,k}} = \int_0^1 K(\gamma)\dot{\gamma_2}\psi_k dt \quad \frac{\partial \mathcal{D}}{\partial \theta_{2,k}} = -\int_0^1 K(\gamma)\dot{\gamma_1}\psi_k dt, \tag{II.9}$$

where $K(x)$ is the mean curvature of the surface defined by I namely

$$K(x) = \Big(I_{22}(1+I_1^2) + I_{11}(1+I_2^2) - 2I_1 I_2 I_{12}\Big)\,(1+I_1^2+I_2^2)^{-3/2}.$$

$I_{\alpha\beta}$ second order partial derivatives.

- Again gradient involves forward transform of a function.

II.3 Computation - gradient descent

- $\theta_{j,T}, j = 1,2, T = 0,1\ldots$ - iterates of the coefficients of the curve.

  Forward transform of $f(t)$ on $[0,1]$ -

  $\Psi(f) = \{\int_0^1 f\psi_k dt, k = 0,\ldots,d\}$ - coefficients of f in $\psi_k, k = 0,\ldots,d$.

  For coefficients $c = (c_0,\ldots,c_d)$ -

  Ψ^{-1} - backwards transform - $\Psi^{-1}(c) = \sum_{k=0}^d c_k\psi_k$.

- **Algorithm 1:**

  1. Initialize $\theta_{j,0} = \{\theta_{j,k,0},\ k = 0,\ldots,d\}, j = 1,2$. Set $T = 0$.
  2. Calculate $\gamma_{1,T} = \Psi^{-1}(\theta_{1,T}), \gamma_{2,T} = \Psi^{-1}(\theta_{2,T})$, and $\dot{\gamma}_{1,T}, \dot{\gamma}_{2,T}$.
  3. Calculate

$$\begin{aligned} v_1 &= \Psi\Big((G_{out} - G_{in})(\gamma)\dot{\gamma}_{2,T}\Big), \\ v_2 &= \Psi\Big((G_{in} - G_{out})(\gamma)\dot{\gamma}_{1,T}\Big). \end{aligned}$$

  4. Set $\theta_{j,k,T+1} = \theta_{j,k,T} - \Delta\cdot(\lambda_k\theta_{j,k,T} + v_{j,k})$ for $j = 1,2,\ k = 1,\ldots,d$.

5. If a stopping criterion is satisfied exit, otherwise goto 2.

- The computationally intensive steps in the algorithm are the forward and backward transforms Ψ, Ψ^{-1}.

- Discretization:

  Curve discrete: $\gamma(\theta, z_i), i = 1, \ldots, n$.

  $\dot{\gamma}$ - approximated using differences.

  Discrete version of basis vectors: fast discrete forward and backwards transforms - FFT, DWT.

- Time step: Inverse of upper bound on trace of Hessian: Set $G = G_{out} - G_{in}$.

$$\frac{\partial^2 \mathcal{D}}{\partial \theta_{j,k}} = \int (G_j \psi_k + G\psi_k^2)\dot{\gamma} dt.$$

  Assuming G smooth so that first term in integrand is small, taking $|G| < M$, and taking say ψ_k to be Fourier basis trace is bounded by $2dM \cdot l(\gamma) + \sum_k \lambda_k$, - $l(\gamma)$ length of γ.

- Allowing curve to freely change size and location - set $\lambda_k = 0$ for small k.

- If curve is initialized at mean shape γ_0, re-parameterize $\gamma = \gamma_0 + \gamma_r$, and solve for γ_r.

II.3.1 Coarse to fine

- Low index elements of basis - smooth - large support: Global smooth variations.

  High indices - higher frequency or local variations.

- Minimize first on low index coefficients and gradually add higher indices.

  Avoid local minima near initial curve.

- Choose $d_1 < d_2 < \ldots < d_m = d$. Apply algorithm 1 but in item (4) update only $k = 1, \ldots, d_1$.

  After convergence, initialize $\theta_{j,k}, k = d_1 + 1, \ldots, d_2$. Update in (4) on $\theta_{j,k}, k = 1, \ldots, d_2$. etc.

- For wavelet basis use $d = 2^m$ and $d_l = 2^l, l = 1, \ldots, m$. The result in figure II.1 was obtained in this fashion.

II.4 The statistical interpretation

- Statistical formulation - details: Grenander et al. (1991), Grenander & Miller (n.d.).

- Given the curve $\gamma(\theta)$: data inside i.i.d $N(\mu_{in}, \sigma)$, outside i.i.d $N(\mu_{out}, \sigma)$.

- Likelihood:

$$\begin{aligned} P(I|\theta) = & \prod_{x \in \gamma(\theta)_{in}} \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu_1)^2/2\sigma^2] \\ & \prod_{x \in \gamma(\theta)_{out}} \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu_2)^2/2\sigma^2] \end{aligned} \tag{II.10}$$

- Coefficients $\theta_{i,k}$ in basis are independent with log-densities $l_{i,k}(\theta_{i,k})$.

- Log-posterior:

$$\log P(\theta|I) = \sum_{j,k} l_{j,k}(\theta_{j,k}) + \sum_{x \in \gamma(\theta)_{in}} (x-\mu_1)^2 + \sum_{x \in \gamma(\theta)_{out}} (x-\mu_2)^2 + C, \tag{II.11}$$

  where C does not depend on θ.

  In cost function above $\theta_{j,k}$ are $N(0, 1/\lambda_k)$.

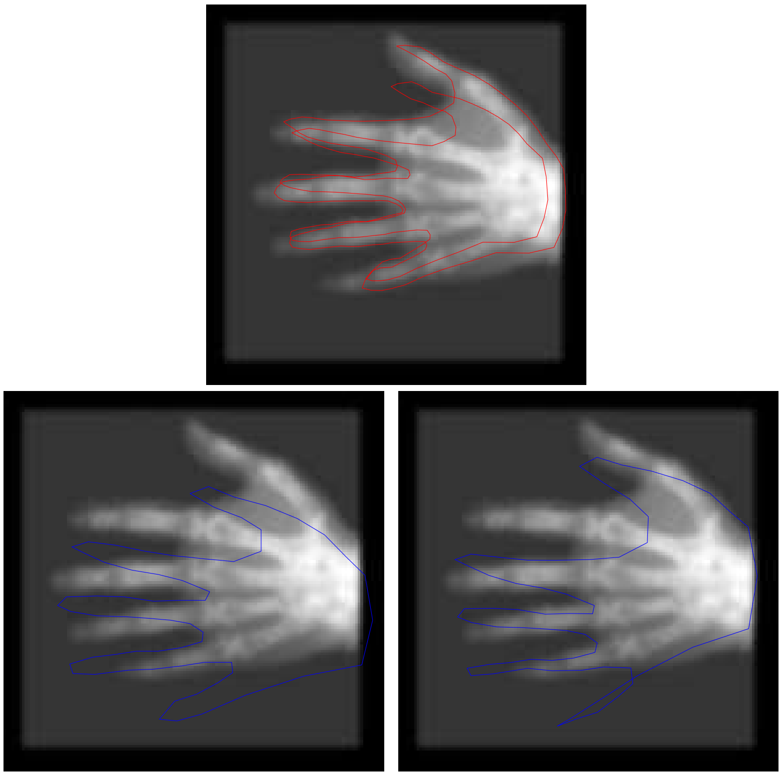

Figure II.2: Top: The initial and final curve of a successful match of a 'hand' snake to a hand xray. Bottom Right: Initialization with wrong fingers aligned. Bottom Left: Final match for middle initialization.

II.5 Discussion

- Bayesian formulation enables simulation from posterior: express uncertainties of the solution.

  Requires serious consideration of relative weights of the various components of the model.

- If the pixel intensities do not have simple statistical model of equation II.10, hard to write an explicit formula for the normalizing constants.

  Note: normalizing constant C typically *will* depend on the parameter θ.

Only meaningful statistical model is for *joint* distribution of the curve and the data:

$$P(\theta, I) = \frac{1}{C}\exp[\mathcal{E}(\theta) + \alpha_{in}\sum_{x\in\gamma(\theta)_{in}} G_1(x) + \alpha_{out}\sum_{x\in\gamma(\theta)_{out}} G_2(x)]. \qquad \text{(II.12)}$$

- The inside outside model is the only case in these notes where the data term can not be written as a function of the data around the points of $\theta Z = \gamma(\theta)$. However the optimization algorithm using gradient descent only involves data near the curve.

- Literature: Variations on data models - mostly edge type terms (equation II.2.4), different forms of curve parameterization and different forms of imposing smoothing penalties on the curves: Chuang & C. (1996), Cohen (1991), Cohen et al. (1992), Cohen & Cohen (1993), Figueiredo & Leitao (1992).

- When initialized with small contour - can be viewed as smoothed region growing algorithms. Region growing is a very old methodology in pattern recognition see for example Duda & Hart (1973) or Haralick & Shapiro (1992), and Zhu & Yuille (1996) for further discussion.

- Problems encountered:

  Initialization of the contour - getting stuck in local minima. See figure II.2.

  Difficult to incorporate robust statistics of image near the curve.

  Difficult to incorporate photometric invariance.

  Difficult to represent typical deformations of the initial shape.

  Some attempts at learning deformations in Cootes & Taylor (1992).

II.6 Appendix: derivation of variational formulas

- This derivation is presented in Zhu & Yuille (1996).

  For $\gamma(t,\theta)$ let
  $$\mathcal{D}(\theta) = \int_0^1 L(\gamma_1(t,\theta), \gamma_2(t,\theta), \dot{\gamma}_1(t,\theta), \dot{\gamma}_2(t,\theta))dt,$$
  for some function L.

  $L_i, i = 1,2,3,4$ - partial derivatives.

- Chain rule yields:

$$\frac{\partial \mathcal{D}(\theta)}{\partial \theta_k} = \int_0^1 L_1 \frac{\partial \gamma_1(t,\theta)}{\partial \theta_k} + L_2 \frac{\partial \gamma_2(t,\theta)}{\partial \theta_k} + L_3 \frac{\partial \dot{\gamma}_1(t,\theta)}{\partial \theta_k} + L_4 \frac{\partial \dot{\gamma}_2(t,\theta)}{\partial \theta_k} dt.$$

- Interchanging the derivatives in the third and fourth terms, and integrating by parts

$$\begin{aligned}\frac{\partial \mathcal{D}(\theta)}{\partial \theta_k} &= \int_0^1 (L_1 - \frac{d}{dt}L_3)(\gamma(t,\theta), \dot{\gamma}(t,\theta)) \frac{\partial \gamma_1(t,\theta)}{\partial \theta_k} + \\ &\quad (L_2 - \frac{d}{dt}L_4)(\gamma(t,\theta), \dot{\gamma}(t,\theta)) \frac{\partial \gamma_2(t,\theta)}{\partial \theta_k} dt.\end{aligned}$$

- L of the form $L(\gamma, \dot{\gamma}) = P(\gamma)\dot{\gamma}_1 + Q(\gamma)\dot{\gamma}_2$ and γ as in equation II.3,

$$\begin{aligned}\frac{\partial \mathcal{D}(\theta)}{\partial \theta_{1,k}} &= \int_0^1 (\partial Q/\partial x_1 - \partial P/\partial x_2)(\gamma)\dot{\gamma}_2 \psi_k dt \\ \frac{\partial \mathcal{D}(\theta)}{\partial \theta_{2,k}} &= -\int_0^1 (\partial Q/\partial x_1 - \partial P/\partial x_2)(\gamma)\dot{\gamma}_1 \psi_k dt.\end{aligned} \tag{II.13}$$

- $\mathcal{D}$ as in equation II.2 - subtract integral of G_{out} over entire domain,

$$\mathcal{D}(\theta) = \int_{\gamma(\theta)_{in}} (G_{in} - G_{out})(x)dx \equiv \int_{\gamma(\theta)_{in}} G(x)dx. \tag{II.14}$$

- Define

$$Q(x_1,x_2) = 1/2 \int_0^{x_1} G(t,x_2)dt, \; P(x_1,x_2) = -1/2 \int_0^{x_2} G(x_1,t)dt.$$

Then $\partial Q/\partial x_1 - \partial P/\partial x_2 = G(x_1, x_2)$, and by Green's theorem

$$\int_{\gamma_{in}} G(x)dx = \int_0^1 P(\gamma(t))\dot{\gamma}_1(t) + Q(\gamma(t))\dot{\gamma}_2(t)dt.$$

- Equation II.13 leads to equation II.6.
- For $\mathcal{D}$ of equation II.8 $P = -(1 + I_1^2 + I_2^2)^{-1/2} I_2$, and $Q = (1 + I_1^2 + I_2^2)^{-1/2} I_1$.

  Substituting in equation II.13 obtain equation II.9.