# Image Segmentation for ISH embryos

**Fanny Yang**
University of California
Berkeley, CA 94720, USA
fanny-yang@eecs.berkeley.edu

**Siqi Wu**
University of California
Berkeley, CA 94720, USA
siqi@stat.berkeley.edu

## 1 Previous works

**Deformable models** Active contours: [8, 7, 9, 16] [23, 12, 19] Review papers: [20, 2]

Level based: [24, 4] Review paper [11]

Level based with shape, intensity prior: Cremers [10, 3, 5]

**Graphical models** MRF: CRF: [17, 13] GraphCut: [18] Sth completely different: [1]

**Hybrid models** [14, 6, 25, 22]

## 2 Model

We are using a MRF closely linked to Huang. As a difference however we do not use contours $c$ which carry the burden of parameterization. But we use the continuous counterpart: level set functions, or more specifically, signed distance functions $\phi(i)$.

Let's denote $y$ as our hidden labels, $x$ as our observed image. We index pixels by $i$ and our signed distance function is $\phi$. Our graphical model looks like in Huang, and thus our joint distribution with potentials $\psi$ reads

$$\begin{aligned} p(x, y, \phi) &= \prod_i \psi(x_i, y_i) \prod_{i,j} \psi(y_i, y_j) \prod_i p(y_i|\phi) p(\phi) \\ &=: \prod_i p(x_i|y_i) p(y_i, y_j) p(y_i|\phi) p(\phi) \end{aligned} \tag{1}$$

We model $p(x_i|y_i)$ as Gaussians with parameters $\mu_{y_i}$ and $\sigma_{y_i}$, i.e.

$$p(x_i|y_i) = \frac{1}{\sqrt{2\pi}\sigma_{y_i}} \exp(-\frac{(x_i - \mu_{y_i})^2}{\sigma_{y_i}^2})$$

Our pairwise potentials

$$p(y_i, y_j) \propto \exp(-c|y_i - y_j|)/\theta_1 1_{y_i \neq y_j} + \theta_2 1_{y_i = y_j}$$

For the consistency measure of the labels $y$ on $\phi$ we use the logistic function

$$p(y_i = 1|\phi) = \frac{1}{1 + \mathrm{e}^{-\lambda\phi(i)}}$$

where $\phi(i)$ is basically a signed distance function (possibly other distances) and $\lambda$ controls the sharpness of the slope. Note: Schlesinger uses $\psi(y|\phi) = \mathrm{e}^{\sum_i y_i \phi(i)}$.

### 2.1 Level Set segmentation

The term $p(\phi)$ enables us to incorporate priors and regularizations. In our model we use length and area as in [10, 4, 21]. Note that instead of $p(x|\phi)$ in level set formulations we now have $p(x|y)$ and $p(y|\phi)$.

$$
\begin{aligned}
-\log p(\phi|\tilde{\phi}) &= E_{int}(\phi) + E_{prior}(\phi) \\
&= \lambda_1 \int_\Omega \delta_\epsilon(\phi(x,y)) \|\nabla\phi(x,y)\| \mathrm{d}x\mathrm{d}y + \lambda_2 \int_\Omega H_\epsilon(\phi(x,y)) \mathrm{d}x\mathrm{d}y \\
&- \log \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2} d^2(H(\phi), H(\phi_i))\right)
\end{aligned}
$$

where $H$ is the Heaviside function and $\sigma^2$ is the width of the Gaussian kernel. $E_{int}$ denotes the regularization on length and area of the region (boundary), $E_{prior}$ is the energy given by the shape prior which is measured as a distance $d^2(H(\phi), H(\phi_i)) = \int_\Omega (H(\phi) - H(\phi_i))^2 \mathrm{d}x$ is basically doing a nonparametric kernel density estimation for the shape distribution using given samples $\phi_i$. We choose for the kernel width $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d^2(H(\phi_i), H(\phi_j))$.

The task is now to maximize $\log Q_D(\phi|b) \propto \log p(b_i|\phi)p(\phi)$ which is equivalent to minimizing the energies $E_{int}, E_{prior}$. By the multivariate Euler Lagrange equations we know that $F(x) = \int \mathcal{L}(x, \phi(x), \nabla\phi(x)) \mathrm{d}x$ is minimized if

$$
\frac{\partial}{\partial\phi} \mathcal{L}(x, \phi(x), \nabla\phi(x)) + \operatorname{div} \frac{\partial}{\partial\nabla\phi} \mathcal{L}(x, \phi(x), \nabla\phi(x)) = 0
$$

We now add time for the evolution of the function and minimize the energy with respect to $\phi$ using the update equation

$$
\begin{aligned}
\frac{\partial\phi}{\partial t} &= -\frac{\partial E_{int}(\phi)}{\partial\phi} - \frac{\partial E_{prior}(\phi)}{\partial\phi} \\
&= \delta_\epsilon(\phi)(\lambda_1 \left[\operatorname{div}(\frac{\nabla\phi}{\|\nabla\phi\|})\right] + \lambda_2) - \frac{\sum_i \alpha_i \frac{\partial}{\partial\phi} d^2(H(\phi), H(\phi_i))}{2\sigma^2 \sum_i \alpha_i}
\end{aligned} \tag{2}
$$

with $\alpha_i = \exp(-\frac{d^2(H\phi, H\phi_i)}{2\sigma^2})$. This is basically doing a gradient descent. For the second term since $E_{prior}$ is a direct function of $\phi$, we can readily take the partial derivative with respect to $\phi$. $\frac{\partial}{\partial\phi} d^2(H(\phi), H(\phi_i))$ is a Gateaux derivative and yields

$$
\frac{\partial}{\partial\phi} d^2(H(\phi), H(\phi_i)) = \delta_\epsilon(\phi)(H(\phi) - H(\phi_i))
$$

The question is how we can discretize the continuous first term in (2). If we expand it we obtain

$$
\operatorname{div} \frac{\nabla\phi}{\|\nabla\phi\|} = \frac{\frac{\partial^2\phi}{\partial x^2}(\frac{\partial\phi}{\partial y})^2 - 2\frac{\partial^2\phi}{\partial x \partial y}\frac{\partial\phi}{\partial x}\frac{\partial\phi}{\partial y} + \frac{\partial^2\phi}{\partial y^2}(\frac{\partial\phi}{\partial x})^2}{(\frac{\partial\phi}{\partial x}^2 + \frac{\partial\phi}{\partial y}^2)^{\frac{3}{2}}}
$$

In our implementation we simply approximate the continuous derivatives by finite differences (this is exactly the formula they use in the Chan Vese Matlab package).

The overall formula for the update thus reads

$$
\begin{aligned}
\frac{\partial\phi}{\partial t} &= \delta_\epsilon(\phi)(\lambda_1 \frac{\frac{\partial^2\phi}{\partial x^2}(\frac{\partial\phi}{\partial y})^2 - 2\frac{\partial^2\phi}{\partial x \partial y}\frac{\partial\phi}{\partial x}\frac{\partial\phi}{\partial y} + \frac{\partial^2\phi}{\partial y^2}(\frac{\partial\phi}{\partial x})^2}{(\frac{\partial\phi}{\partial x}^2 + \frac{\partial\phi}{\partial y}^2)^{\frac{3}{2}}} + \lambda_2) \\
&- \frac{\sum_i \alpha_i \delta_\epsilon(\phi)(H(\phi) - H(\phi_i))}{2\sigma^2 \sum_i \alpha_i}
\end{aligned}
$$

### 2.2 Variational inference

The inference task is to maximize $p(y,\phi|x) \propto p(x,y,\phi)$. However finding the maximum of (1) exactly is computationally intractable. We use structural variational inference [15, 26] to approximate the problem, in which we look for a $Q(y,\phi|x,a,b)$ with a certain structure and parameters $a,b$ and minimize the Kullback Leibler divergence $KL(Q||P) = \sum_\phi \sum_y Q(y,\phi|x,a,b) \log \frac{Q(y,\phi|x,a,b)}{p(y,\phi|x)}$ Our ansatz for the structure of $Q$ is the following

$$Q(y,\phi|x,a,b) = Q_M(y|x,a)Q_D(\phi|b)$$

with $Q_M(y|x,a) = \prod_{i,j} \psi(y_i,y_j) \prod_i \xi(x_i,y_i)p(y_i|a_i)$ and $Q_D(\phi|b) \propto \prod_i p(b_i|\phi)p(\phi)$ so that we have factorized the distribution with respect to the two unknown vectors $y,\phi$. We are hence assuming independence of the two, given $a,b$ respectively.

We now need to minimize the KL divergence for which we follow the easy derivation in Blei: Using the independence we can write out the KL

$$\begin{aligned}
&\mathbb{E}\ \log Q_M(y|x,a) + \mathbb{E}\ \log Q_D(\phi|b) - \mathbb{E}\ \log p(y,\phi|x) \\
&= \mathbb{E}\ [\log Q_M(y|x,a)|\phi] + \mathbb{E}\ [\log Q_D(\phi|b)|y] - \mathbb{E}\ [\mathbb{E}\ \log p(y|\phi)|\phi] - \mathbb{E}\ \log p(\phi) \\
&= \mathbb{E}_{Q_M}(\log Q_M(y|x,a) - \mathbb{E}_{Q_D} \log p(y|\phi)) + \mathbb{E}_{Q_D}(\log Q_D(\phi|b) - \log p(\phi)) \\
&= \mathbb{E}_{Q_M}(\log Q_M(y|x,a) - \mathbb{E}_{Q_D} \log p(y|\phi)) + \mathbb{E}_{Q_D}(\log p(b|\phi) + c)
\end{aligned}$$

where the first equality follows from the conditional independence of $y,\phi$ given $a,b$ and the second uses $Q_D(\phi) = p(\phi)$.

By Lagrange equations of first kind to incorporate the constraint of $\mathbb{E}\, Q_D(\phi|b)|b = 1$ and $\mathbb{E}\, Q_M(y|x,a)|x,a = 1$ we obtain that the Lagrange multiplier $\lambda = 1$ and the following must hold for optimal $a,b$:

$$\begin{aligned}
&- \mathbb{E}_{Q_D} \log p(y|\phi) + \log Q_M(y|x,a) = 0 \\
\implies &\sum_i (\mathbb{E}_{Q_D} \log p(y_i|\phi) - \log p(y_i|a_i)) = \sum_{j\in N(i)} \log p(y_i,y_j) - \mathbb{E}_{Q_D} \log(p(y_i,y_j)) \\
\implies &\log p(y_i|a_i) = \mathbb{E}_{Q_D} \log p(y_i|\phi)
\end{aligned}$$

and similarly

$$\begin{aligned}
&- \mathbb{E}_{Q_M} \log p(y|\phi) + \log p(b|\phi) = 0 \\
\implies &\log p(b_i|\phi) = \mathbb{E}_{Q_M^i} \log p(y_i|\phi)
\end{aligned}$$

where $Q_M^i = Q_M(y_i|x,a)$.

These are the fixed point equations and we will aim at solving them iteratively using alternating minimization.

### 2.3 Algorithm

We now turn to the algorithm pipeline .

**Algorithm** Initialize level set function $\phi$. Step $k$.

1. HOW DO WE do this first step - we only cover a region of $\phi$ for the steps?
2. Calculate $\phi^* = \text{argmax}_\phi Q_D(\phi|b)$ using the current $p(b|\phi)$ and approximate by $Q'_D(c|b) = \delta(\phi^*)$
3. Calculate $p(y_i|a_i) = \exp(\mathbb{E}_{Q'_D} \log(y_i|a_i)) = p(x_i|\phi^*)$.
4. In order to calculate $Q_M(y_i|x_i,a)$ estimate the parameters of $p(x_i|\theta_i)$ which are $\mu_1,\mu_0,\sigma_1,\sigma_0$ using the latent variable $y_i$ and the EM algorithm
5. Calculate $p(b_i|\phi) = \mathbb{E}_{Q_M^i} \log p(y_i|\phi)$.

### 2.4 EM algorithm

We now turn to the step of estimating the parameters of the Gaussians $p(x_i|y_i)$ for $Q_M(y|x,a)$ for which we will use an EM algorithm. The aim is to calculate

$$\max_\theta p(x|\theta,a) = \sum_y p(x,y|\theta) = \sum_y p(x|y,\theta)p(y|a)$$

For the EM algorithm we have the E Step

$$\begin{aligned}
\mathcal{L}(q^{(t+1)},\theta) &= \sum_y p(y|x,\theta^{(t)})(\log p(x|y,\theta) + \log p(y|a)) \\
&= \sum_i \sum_{y_i} \log p(x_i|y_i,\theta) \sum_{y_j:j\neq i} p(y|x,\theta^{(t)},a) + c(\theta) \\
&= \sum_i \sum_{y_i} p(y_i|x,\theta^{(t)},a) 1_{y_i=l}(-\log\sigma_l - \frac{(x_i-\mu_l)^2}{2\sigma_l^2}) + c(\theta) \\
&= \sum_i \sum_l p(y_i=l|x,\theta^{(t)},a)(-\log\sigma_l - \frac{(x_i-\mu_l)^2}{2\sigma_l^2}) + c(\theta)
\end{aligned}$$

For the maximization step we maximize $\mathcal{L}(q^{(t+1)},\theta)$ over $\theta$ so that we have

$$\begin{aligned}
\mu_l^{(t+1)} &= \frac{\sum_i p(y_i=l|x,\theta^{(t)},a)x_i}{\sum_i p(y_i=l|x,\theta^{(t)})} \\
\sigma_l^{2(t+1)} &= \frac{\sum_i p(y_i=l|x,\theta^{(t)},a)(x_i-\mu_l)^2}{\sum_i p(y_i=l|x,\theta^{(t)},a)}
\end{aligned}$$

which intuitively just computes the average intensity/variance within one label.

### 2.5 Belief Propagation

As we see we need to compute the marginal likelihood

$$p(y_i|x,\theta^{(t)},a) \propto \sum_{y_j:j\neq i} \prod_i p(x_i|y_i) \prod_{i,j} p(y_i,y_j) \prod_i p(y_i|a_i)$$

This can be readily computed using a loopy belief propagation algorithm/junction tree algorithm (slow?).

### 2.6

## References


[1] Bjoern Andres, Jörg H Kappes, Thorsten Beier, Ullrich Kothe, and Fred A Hamprecht. Probabilistic image segmentation with closedness constraints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2611–2618. IEEE, 2011.

[2] Dr A Govardhan Baswaraj and D P Premchand. Active contours and image segmentation: The current state of the art. *Global Journal of Computer Science and Technology*, 12(11-F), 2012.

[3] Tony Chan and Wei Zhu. Level set based shape prior segmentation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1164–1170. IEEE, 2005.

[4] Tony F Chan and Luminita A Vese. Active contours without edges. *Image processing, IEEE transactions on*, 10(2):266–277, 2001.

[5] Siqi Chen and Richard J Radke. Level set segmentation with both shape and intensity priors. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 763–770. IEEE, 2009.

[6] Xinjian Chen, Jayaram K Udupa, Ulas Bagci, Ying Zhuge, and Jianhua Yao. Medical image segmentation by combining graph cuts and oriented active appearance models. *Image Processing, IEEE Transactions on*, 21(4):2035–2046, 2012.

[7] Timothy F Cootes, Gareth J Edwards, Christopher J Taylor, et al. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[8] Timothy F Cootes and Christopher J Taylor. Active shape modelssmart snakes. In *BMVC92*, pages 266–275. Springer, 1992.

[9] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Training models of shape from sets of examples. In *BMVC92*, pages 9–18. Springer, 1992.

[10] Daniel Cremers, Stanley J Osher, and Stefano Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 69(3):335–351, 2006.

[11] Daniel Cremers, Mikael Rousson, and Rachid Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision*, 72(2):195–215, 2007.

[12] Ayman El-Baz and Georgy Gimelfarb. Robust medical images segmentation using learned shape and appearance models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 281–288. Springer, 2009.

[13] Xuming He, Richard S Zemel, and MA Carreira-Perpindn. Multiscale conditional random fields for image labeling. In *Computer vision and pattern recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE computer society conference on*, volume 2, pages II–695. IEEE, 2004.

[14] Rui Huang, Vladimir Pavlovic, and Dimitris N Metaxas. A graphical model framework for coupling mrfs and deformable models. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–739. IEEE, 2004.

[15] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[16] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[17] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, p.282-289, June 28-July 01, 2001*, 2001.

[18] Victor Lempitsky, Andrew Blake, and Carsten Rother. Branch-and-mincut: global optimization for image segmentation with high-level priors. *Journal of Mathematical Imaging and Vision*, 44(3):315–329, 2012.

[19] Michael E Leventon, W Eric L Grimson, and Olivier Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 316–323. IEEE, 2000.

[20] Tim McInerney and Demetri Terzopoulos. Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108, 1996.

[21] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.

[22] Dmitrij Schlesinger. A continuous shape prior for mrf-based segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 350–361. Springer, 2013.

[23] Stan Sclaroff and Lifeng Liu. Deformable shape detection and description via model-based region grouping. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(5):475–489, 2001.

[24] Andy Tsai, Anthony Yezzi Jr, William Wells, Clare Tempany, Dewey Tucker, Ayres Fan, W Eric Grimson, and Alan Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Transactions on*, 22(2):137–154, 2003.

[25] Mustafa Gökhan Uzunbaş, Chao Chen, Shaoting Zhang, Kilian M Pohl, Kang Li, and Dimitris Metaxas. Collaborative multi organ segmentation by integrating deformable and graphical models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 157–164. Springer, 2013.

[26] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305, 2008.

# Appendices