

# A Stochastic-Variational Model for Soft Mumford-Shah Segmentation \*

Jianhong (Jackie) Shen

School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA

Lotus Hill Institute for Computer Vision and Information Science, E'Zhou, Wuhan 436000, China

## Abstract

In contemporary image and vision analysis, stochastic approaches demonstrate great flexibility in representing and modeling complex phenomena, while variational-PDE methods gain enormous computational advantages over Monte-Carlo or other stochastic algorithms. In combination, the two can lead to much more powerful novel models and efficient algorithms. In the current work, we propose a stochastic-variational model for *soft* (or *fuzzy*) Mumford-Shah segmentation of mixture image patterns. Unlike the classical *hard* Mumford-Shah segmentation, the new model allows each pixel to belong to each image pattern with some probability. We show that soft segmentation leads to hard segmentation, and hence is more general. The modeling procedure, mathematical analysis, and computational implementation of the new model are explored in detail, and numerical examples of synthetic and natural images are presented.

**Keywords:** Segmentation, soft, mixture, pattern, ownership, probability simplex, Modica-Mortola, phase-field, symmetry, supervision, Egorov's theorem, Poincaré inequality, existence, AM algorithm.

## Contents

<b>1</b>	<b>Introduction: Soft vs. Hard Segmentation</b>	<b>2</b>
<b>2</b>	<b>Bayesian Rationale to the New Model and Gaussian Mixture</b>	<b>4</b>
2.1	Bayesian Rationale . . . . .	4
2.2	Gaussian Mixture with Smooth Mean Fields . . . . .	5
<b>3</b>	<b>Modica-Mortola's Phase-Field Model for Ownership Energy</b>	<b>6</b>
<b>4</b>	<b>Soft Mumford-Shah (SMS) Segmentation</b>	<b>8</b>
4.1	The Model and Admission Space . . . . .	8
4.2	Breaking the Hidden Symmetry via Weak Supervision . . . . .	9
4.3	Existence Theorems for Non-Supervision and Supervision . . . . .	10
4.4	Mixture of Homogeneous Gaussians . . . . .	13
<b>5</b>	<b>Euler-Lagrange Equations and Computation on <math>(K - 1)</math>-Simplex</b>	<b>13</b>
5.1	Euler-Lagrange Equations on $(K - 1)$ -Simplex . . . . .	13
5.2	Computation of the Euler-Lagrange Equations . . . . .	15
<b>6</b>	<b>Computational Examples</b>	<b>17</b>

---

\*This work has been partially supported by the NSF (USA) under grant number DMS-0202565. Email: jhshen@math.umn.edu; Tel: (USA) (612) 625-3570.

# 1 Introduction: Soft vs. Hard Segmentation

Segmentation is the key step towards high-level vision modeling and analysis, including object characterization, detection, and classification. There have been some recent developments indicating that certain high-level visual tasks such as global scene interpretation might be able to bypass segmentation [24, 23]. Nevertheless, segmentation still remains perhaps *the* most important and inspiring task to date in low- or middle-level vision analysis and image processing.

The segmentation problem can be formulated as follows. Given an image  $I$  on a 2-dimensional (2D) domain  $\Omega$  (assumed to be bounded, smooth, and open), one seeks out a *closed* “edge set”  $\Gamma$ , and all the connected components  $\Omega_1, \dots, \Omega_K$  of  $\Omega \setminus \Gamma$ , such that by certain suitable visual measure (e.g., textural or photometric), the image  $I$  is discontinuous along  $\Gamma$  while smooth or homogeneous on each segment  $\Omega_i$ . Each image patch  $I_i = I|_{\Omega_i}$  is also called a *pattern*, and  $\Omega_i$  its *support*.

We shall call this most common practice “hard” segmentation. A hard segmentation partitions the image domain  $\Omega$  along a *definitive* edge set  $\Gamma$ , and outputs *non-overlapping* pattern supports  $\Omega_1, \dots, \Omega_K$ .

The present work introduces the notion of “soft” segmentation. Mathematically, a hard segmentation amounts to the partition of the unit using indicator functions:

$$1_{\Omega}(x) = \sum_{i=1}^K 1_{\Omega_i}(x), \quad a.e. \quad x = (x_1, x_2) \in \Omega.$$

A soft segmentation seeks out instead a softer partition of the unit:

$$1_{\Omega}(x) = \sum_{i=1}^K p_i(x), \tag{1}$$

where  $p_i$ ’s are continuous or smoother functions. Formally, each  $p_i$  could be considered as the mollified version of  $1_{\Omega_i}(x)$ .

In the stochastic literature of image analysis and modeling, the above notion of soft segmentation is closely connected to *mixture image models* (e.g., [20]). Suppose a given image  $I$  is composed from  $K$  unknown patterns:

$$\omega = 1, \quad \omega = 2, \quad \dots, \quad \omega = K,$$

where  $\omega$  denotes the pattern label variable. At each pixel  $x \in \Omega$ ,  $\omega(x) \in \{1, \dots, K\}$  becomes a random variable. Then the  $p_i$ ’s in (1) carry the natural stochastic interpretation:

$$p_i(x) = \text{Prob}(\omega(x) = i), \quad i = 1 : K.$$

For this reason, each  $p_i$  shall be called the *ownership* of pattern  $i$ . Instead of the repulsive ownership in a hard segmentation, a soft one allows each pattern to “own” a pixel with some likelihood.

Soft segmentation is more general since it can lead to natural hard segmentation under the *maximum likelihood* (ML) principle. Given a soft segmentation  $\{p_i(x) : i = 1 : K\}$ , one can define for each pixel  $x \in \Omega$  its unique owner  $\omega_*(x)$  by:

$$\omega_*(x) = \text{argmax}_{\omega \in 1:K} p_{\omega}(x), \tag{2}$$

and if the maxima are non-unique, accept the largest index from the argmax pool. The segments are then defined by

$$\Omega_i = \omega_*^{-1}(i) = \{x \in \Omega \mid \omega_*(x) = i\}, \quad i = 1 : K, \tag{3}$$

which leads to a natural hard segmentation. (2) and (3) are called the *hardening formulae*.

Soft segmentation has been motivated by practical analysis of natural images. Patterns in natural scenes often do not have clear-cut boundaries. In Figure 1, for example, there does not seem to exist a “hard” boundary between the grass and sand areas. If one draws an oriented line as shown in the figure, it makes more sense to state that along the arrow the pattern transits from being “more” sand like to being “more” grass like. Such consideration favors the following stochastic view that along the arrow, the ownership

$\text{Prob}(\omega(x) = \text{grass})$  increases,    while     $\text{Prob}(\omega(x) = \text{sand})$  decreases.



Figure 1: Natural images often do not have clear-cut “hard” boundaries between different patterns. Along the arrow, for example, one only observes that the sand pattern *gradually* becomes a grass pattern. Such a “soft” view is the stochastic view on the segmentation problem.

In the present work, we propose a new stochastic-variational soft-segmentation model for the following celebrated Mumford-Shah model [31, 33]:

$$\min_{\Gamma, u} E[u, \Gamma \mid I] = \mathcal{H}^1(\Gamma) + \alpha \int_{\Omega \setminus \Gamma} |\nabla u|^2 + \lambda \int_{\Omega} (u - I)^2, \quad (4)$$

where  $\mathcal{H}^1$  stands for the 1D Hausdorff measure, which is simply the length when  $\Gamma$  is regular enough. *For notational conciseness, the default area-element symbol  $dx = dx_1 dx_2$  will be omitted in most integral formulae.*

As stated in the abstract, the stochastic softness induces more flexibility and universality in modeling, while the variational-PDE approach facilitates rigorous mathematical analysis as well as more efficient computational implementations compared with purely stochastic approaches including, e.g., the Monte-Carlo method or Gibbs’ sampling [4, 15, 43, 45].

The paper has been organized as follows. Section 2 builds up the soft Mumford-Shah (SMS) model under the Bayesian rationale and the MAP estimator [21, 32], which are the formal stochastic foundations of the present model. In Section 3, the prior energy on the ownerships  $p_i$ ’s is developed based on the celebrated work of Modica and Mortola [30] on phase-field modeling and  $\Gamma$ -convergence approximation in material sciences and phase transitions. In Section 4, we analyze the main mathematical properties of the proposed SMS model, including the admissible space, hidden symmetry and symmetry breaking via weak supervision, and the existence theorems. In Section 5, we then derive the system of Euler-Lagrange equations of the SMS model for which the role of the probability simplex constraint is discussed in detail. Section 5 also introduces the alternating-minimization algorithm to compute the Euler-Lagrange equations. Finally, the numerical performance of the SMS model is demonstrated in Section 6 via both synthetic and natural test images that are sufficiently representative and generic.

Throughout the manuscript, the notation  $E[X, Y | Z]$  in the deterministic setting always denotes a quantity (often a functional)  $E$  that depends on  $X, Y$ , and  $Z$  but with  $Z$  given or fixed. Similarly,  $E[X | Y, Z]$  still denotes  $E[X, Y | Z]$  modulo some additive quantity  $g[Y, Z]$  that is often unimportant as far as the optimization on  $X$  (given  $Y$  and  $Z$ ) is concerned. These notations therefore have been closely inspired by *conditional probabilities* in the stochastic setting.

## 2 Bayesian Rationale to the New Model and Gaussian Mixture

### 2.1 Bayesian Rationale

Segmentation can be done in some feature spaces such as gradient-like highpass features or Gabor features (e.g., [35, 44, 45]). The Mumford-Shah model easily extends to such general features (e.g., [35]), even though it was originally formulated only for intensity fields. For maximal clarity in exposing the core ideas of the current work, we shall also focus only on the latter, while leaving as canonical exercises to adapt the new model for any given feature distribution.

Let  $K$  be the total number of intended patterns. As in [43, 45],  $K$  could also be treated as an unknown to be optimally estimated, which however does not add much to the most significant contribution (i.e., the modeling and computation of the “softening” procedure) of the present work.

Given an image input  $I = I(x)$  on a bounded, regular, and open domain  $\Omega$ , the primary goal of soft segmentation is to compute the ownerships:

$$p_1(x), p_2(x), \dots, p_K(x).$$

Define  $\mathbf{P}(x) = (p_1(x), p_2(x), \dots, p_K(x))$ , and

$$\Delta_{K-1} = \text{convex hull of } \vec{e}_1, \dots, \vec{e}_K,$$

where the  $(\vec{e}_i | i = 1 : K)$  denotes the canonical Cartesian basis of  $\mathbb{R}^K$ .  $\Delta_{K-1}$  is often called the *canonical*  $(K - 1)$ -simplex, or the *probability*-simplex in  $\mathbb{R}^K$ . Then

$$\mathbf{P} : \Omega \rightarrow \Delta_{K-1}, \quad x \rightarrow \mathbf{P}(x),$$

meaning that the total ownerships always add up to 100% at any pixel  $x \in \Omega$ .

Associated with each pattern label  $\omega = i$  is a smooth function  $u_i(x) \in H^1(\Omega)$ , similar to the original Mumford-Shah model. Here the Sobolev space  $H^1(\Omega)$  is defined by [39]

$$H^1(\Omega) = \{u \in L^2(\Omega) \mid \nabla u \in L^2(\Omega, \mathbb{R}^2)\}.$$

Define  $\mathbf{U}(x) = (u_1(x), u_2(x), \dots, u_K(x))$ . Then the goal of soft segmentation is to estimate the optimal vectorial pair of ownerships and patterns given an image  $I$ :

$$(\mathbf{P}_*, \mathbf{U}_*) = \operatorname{argmax}_{(\mathbf{P}, \mathbf{U})} \operatorname{Prob}(\mathbf{P}, \mathbf{U} \mid I).$$

By the Bayesian formula [21, 32], the posterior given  $I$  is expressible via

$$\operatorname{Prob}(\mathbf{P}, \mathbf{U} \mid I) = \operatorname{Prob}(I \mid \mathbf{P}, \mathbf{U}) \operatorname{Prob}(\mathbf{P}) \operatorname{Prob}(\mathbf{U}) / \operatorname{Prob}(I),$$

assuming that the mixture patterns  $\mathbf{U}$  and the mixture rules  $\mathbf{P}$  are *independent*. We shall call the first term a “mixture generation” model, since it reveals how the image data should look like given the information of the patterns and their ownerships.

By taking the logarithmic likelihood  $E[\cdot] = -\log \text{Prob}(\cdot)$ , or formally Gibbs' energy in statistical mechanics [9, 16], one attains the soft segmentation model in its “energy” form:

$$\min_{(\mathbf{P}, \mathbf{U})} E[\mathbf{P}, \mathbf{U} | I] = E[I | \mathbf{P}, \mathbf{U}] + E[\mathbf{P}] + E[\mathbf{U}], \quad (5)$$

modulo an insignificant additive constant.

Assuming that all the pattern channels are independent from each other, one has

$$E[\mathbf{U}] = E[u_1, \dots, u_K] = \sum_{i=1}^K E[u_i | i].$$

For Sobolev-regular patterns, one may impose the homogeneous Sobolev energies:

$$E[u_i | i] = E[u_i] = \alpha \int_{\Omega} |\nabla u_i|^2, \quad i = 1 : K, \quad (6)$$

for some scalar weight  $\alpha$  that models the visual sensitivity to intensity roughness. Unlike the original Mumford-Shah model, the energy for each channel has been defined on the entire image domain  $\Omega$  instead of on each “hard-cut” patch  $\Omega_i$ . Thus the energy form (6) must carry out *extrapolation* for practical applications. Long-range extrapolations are, however, often unimportant after being weighed down by their negligible ownerships  $p_i$ 's.

## 2.2 Gaussian Mixture with Smooth Mean Fields

In this subsection we discuss the mixture generation model  $\text{Prob}(I | \mathbf{P}, \mathbf{U})$  or  $E[I | \mathbf{P}, \mathbf{U}]$ .

Assume that the patterns are all Gaussian's with mean fields  $u_1, u_2, \dots, u_K$ . For simplicity also assume that they share the same variance  $\sigma^2$  (which readily generalizes to the more general case with variations). Then at any given pixel  $x \in \Omega$ ,

$$(I | \omega(x) = i) \sim N(u_i(x), \sigma^2), \quad i = 1 : K.$$

Define the Gaussian probability density function (p.d.f)

$$g(I | m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I - m)^2}{2\sigma^2}\right).$$

The the p.d.f of the mixture image  $I$  at any pixel  $x$  is given by

$$\text{Prob}(I(x) | \mathbf{P}(x), \mathbf{U}(x)) = \sum_{i=1}^K \text{Prob}(I | \omega(x) = i) \text{Prob}(\omega(x) = i) = \sum_{i=1}^K g(I | u_i(x), \sigma) p_i(x).$$

Thus *ideally* the “energy” for the mixture generation model should be given by

$$E[I | \mathbf{P}, \mathbf{U}] = -\mu \int_{\Omega} \log\left(\sum_{i=1}^K g(I | u_i(x), \sigma) p_i(x)\right), \quad \text{for some scalar } \mu > 0, \quad (7)$$

provided that, given two fields  $\mathbf{P}$  and  $\mathbf{U}$  on  $\Omega$ , for any two distinct pixels  $x$  and  $y$ ,

$$(I(x) | \mathbf{P}, \mathbf{U}) \text{ is independent of } (I(y) | \mathbf{P}, \mathbf{U}).$$

In the current work, we shall adopt a reduced form of the complex formula (7), which is simpler and easier to manage both in theory and for computation. Assume that each soft ownership  $p_i(x)$  is closer to a hard one  $p_i(x) \simeq 1_{\Omega_i}(x)$  for  $i = 1 : K$ . Then

$$\begin{aligned}
-\log \left( \sum_{i=1}^K g(I | u_i(x), \sigma) p_i(x) \right) &\simeq -\log \left( \sum_{i=1}^K g(I | u_i(x), \sigma) 1_{\Omega_i}(x) \right) \\
&= -\sum_{i=1}^K \log g(I | u_i(x), \sigma) 1_{\Omega_i}(x) \quad (a. e.) \\
&\simeq -\sum_{i=1}^K \log g(I | u_i(x), \sigma) p_i(x) \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^K (I - u_i(x))^2 p_i(x) + const.,
\end{aligned}$$

where the additive constant only depends on  $\sigma$  and  $K$ . This suggests the following convenient energy form for the mixture generation model:

$$E[I | \mathbf{P}, \mathbf{U}] = \lambda \int_{\Omega} \left( \sum_{i=1}^K (I - u_i(x))^2 p_i(x) \right), \quad (8)$$

which amounts to a weighted least-square energy [42]. The weight  $\lambda$  reflects visual sensitivity to synthesis errors.

In combination of (5), (6), and (8), the new soft segmentation model takes the form of minimizing

$$E[\mathbf{P}, \mathbf{U} | I] = \lambda \sum_{i=1}^K \int_{\Omega} (I - u_i(x))^2 p_i(x) + \alpha \sum_{i=1}^K \int_{\Omega} |\nabla u_i|^2 + E[\mathbf{P}]. \quad (9)$$

Notice that here the ownership distribution  $\mathbf{P}$  “softens” the “hard” segmentation boundary  $\Gamma$  in the original Mumford-Shah model (4). To complete the modeling process, it suffices to properly define the prior or regularity energy  $E[\mathbf{P}]$ , which is the main task of the next section.

### 3 Modica-Mortola’s Phase-Field Model for Ownership Energy

To generalize but not to deviate too far from classical hard segmentation, it is natural to impose the following two constraints:

- (a) each pattern ownership  $p_i(x)$  has almost only two phases: on (corresponding to  $p_i = 1$ ) and off (to  $p_i = 0$ ), and the transition band in between is narrow; and
- (b) the soft boundaries, or equivalently the transition bands, are *regular*, instead of being zigzag.

In combination, one imposes the following Modica-Mortola type of energy with a double-well potential [30]:  $p_i \in H^1(\Omega)$ ,

$$E_{\varepsilon}[p_i] = \int_{\Omega} \left( 9\varepsilon |\nabla p_i|^2 + \frac{(p_i(1 - p_i))^2}{\varepsilon} \right), \quad i = 1 : K. \quad (10)$$

Here  $\varepsilon \ll 1$  controls the transition bandwidth. Since  $\varepsilon \ll 1$ , the second term necessarily demands  $p_i \simeq 0$  or 1 to lower the energy, which well resonates with the expectation in (a). The first term,

weighted by the small parameter  $\varepsilon$ , amounts to a regularity condition on each  $p_i$ , which meets the requirement in (b).

Energies in the form of (10) are very common in material sciences, including the theories of liquid crystals and phase transitions [11, 17]. Mathematically they have been well studied in the framework of  $\Gamma$ -convergence [29], which we now give a brief introduction in the present context. We also refer the reader to the works of Ambrosio and Tortorelli [1, 2] on the  $\Gamma$ -convergence approximation to the classical Mumford-Shah segmentation model.

Recall that for any  $q(x) \in L^1(\Omega)$ , its *total variation* as a Radon measure is defined by [7, 18]

$$\text{TV}[q] = \int_{\Omega} |Dq| = \sup_{\mathbf{g} \in C_0^1(\Omega, B^2)} \langle q, \nabla \cdot \mathbf{g} \rangle,$$

where  $B^2$  stands for the unit disk centered at the origin in  $\mathbb{R}^2$ . (The TV measure was first introduced into image processing by Rudin, Osher, and Fatemi [34].) Define for any  $q \in L^1(\Omega)$ ,

$$E_0[q] = \begin{cases} \text{TV}[q], & \text{if } q = 0 \text{ or } 1, \text{ a.e. on } \Omega, \\ \infty, & \text{otherwise.} \end{cases}$$

As a result, a finite energy  $E_0[q]$  necessarily implies that  $q$  has two phases only, and  $E_0[q] = \text{TV}[q] = \text{Per}(q^{-1}(1))$  is the perimeter of the support region  $V = q^{-1}(1)$ .

Further define

$$L_{[0,1]}^1(\Omega) = \{q \in L^1(\Omega) \mid q(x) \in [0, 1], \forall x \in \Omega\}$$

to be a subspace of  $L^1(\Omega)$  (as a metric space). Then Modica and Mortola's well known results in [30] readily leads to the following theorem.

**Theorem 1 ( $\Gamma$ -Convergence Approximation of a Two-Phase TV)** *For any  $q \in L_{[0,1]}^1(\Omega) \setminus H^1(\Omega)$ , we extend the definition of  $E_{\varepsilon}[\cdot]$  in (10) by defining  $E_{\varepsilon}[q] = +\infty$ . Then*

$$E_{\varepsilon} \rightarrow E_0 \quad \text{in the sense of } \Gamma\text{-convergence in the metric space } L_{[0,1]}^1(\Omega).$$

That is

(i) for any  $q_{\varepsilon} \rightarrow q$  in  $L_{[0,1]}^1(\Omega)$  as  $\varepsilon \rightarrow 0$ ,

$$\liminf_{\varepsilon \rightarrow 0} E_{\varepsilon}[q_{\varepsilon}] \geq E_0[q]; \quad \text{and}$$

(ii) for any  $q \in L_{[0,1]}^1(\Omega)$ , there exists some sequence  $(q_{\varepsilon}^* \mid \varepsilon)$ , such that  $q_{\varepsilon}^* \rightarrow q$  as  $\varepsilon \rightarrow 0$ , and

$$\lim_{\varepsilon \rightarrow 0} E_{\varepsilon}[q_{\varepsilon}^*] = E_0[q].$$

We refer the reader to Modica and Mortola [30] for a proof (with some necessary modification). Here we only point out that the ‘‘tight’’ sequence  $(q_{\varepsilon}^* \mid \varepsilon)$  in (ii) can be constructed using a smooth sigmoid transition across the hard boundary of a given two-phase function  $q$ . Recall as in the theory of neural networks [3] that a *sigmoid* transition between 0 and 1 is achieved by

$$\sigma(t) = \frac{1}{1 + e^{-t}}, \quad -\infty < t < \infty.$$

The scaling parameter  $\varepsilon$  participates in the transition by the form of  $\sigma(t/(3\varepsilon))$ . In particular,  $\varepsilon$  indeed corresponds to the width of the transition band when  $t$  is a distance function.

This theorem reveals the close connection of the particular choice of  $E_{\varepsilon}[p_i]$  in (10) to the original Mumford-Shah model.

**Proposition 1** Suppose  $p_\varepsilon$ 's “optimally” (i.e., by the above sigmoidal transition) converge to a given 2-phase pattern  $1_V(x)$  with a regular hard boundary  $\Gamma = \partial V$ . Then,

$$E_\varepsilon[p_\varepsilon] \rightarrow \text{length}(\Gamma) = \int_\Omega |D1_V(x)|.$$

Similar results have appeared in the earlier influential works of Ambrosio and Tortorelli [1, 2] on the  $\Gamma$ -convergence approximation to the Mumford-Shah model. The technique has also been extensively applied in image computation and modeling [12, 27, 28, 37, 38, 40] to overcome the difficulty in representing and computing the free boundary  $\Gamma$ .

To summarize this section, we propose the following energy model for the ownership distribution  $\mathbf{P}(x) = (p_1(x), p_2(x), \dots, p_K(x))$ :

$$E_\varepsilon[\mathbf{P}] = \sum_{i=1}^K E_\varepsilon[p_i] = \sum_{i=1}^K \int_\Omega \left( 9\varepsilon |\nabla p_i|^2 + \frac{(p_i(1-p_i))^2}{\varepsilon} \right). \quad (11)$$

One, however, must realize that different ownerships are *not* decoupled by this energy though it has appeared so. The energy  $E_\varepsilon[\mathbf{P}]$  must be coupled with the constraint of the probability-simplex:

$$\mathbf{P} : \Omega \rightarrow \Delta_{K-1}, \quad \text{or} \quad \sum_{i=1}^K p_i(x) \equiv 1, \quad p_i \geq 0, \quad \forall x \in \Omega.$$

In particular, for small  $\varepsilon$ , although (11) implies that each ownership  $p_i$  tends to polarize to 0 or 1 independently, they have to cooperate with each other under the above simplex constraint to optimally share the ownerships.

## 4 Soft Mumford-Shah (SMS) Segmentation

### 4.1 The Model and Admission Space

Combining the preceding two sections, we have developed the complete formula for soft Mumford-Shah segmentation with  $K$ -patterns:

$$\min_{\mathbf{P}, \mathbf{U}} E[\mathbf{P}, \mathbf{U} | I] = \lambda \sum_{i=1}^K \int_\Omega (u_i - I)^2 p_i + \alpha \sum_{i=1}^K \int_\Omega |\nabla u_i|^2 + \sum_{i=1}^K \int_\Omega \left( 9\varepsilon |\nabla p_i|^2 + \frac{(p_i(1-p_i))^2}{\varepsilon} \right), \quad (12)$$

with the constraint that

$$\mathbf{P} : \Omega \rightarrow \Delta_{K-1}, \quad \text{the probability } (K-1)\text{-simplex,}$$

i.e.,  $p_i \geq 0, i = 1 : K$ , and  $\sum_{i=1}^K p_i = 1$ . As discussed previously, it is this simplex constraint that induces coupling among different channels into the seemingly decoupled model (12).

Besides the simplex constraint, the last term in the energy (12) requires  $p_i \in H^1(\Omega)$  for  $i = 1 : K$ . Similarly, the second term requires each pattern  $u_i \in H^1(\Omega)$ . Then with the assumption that

“the given image  $I \in L^2(\Omega)$ ,”

$E[\mathbf{P}, \mathbf{U} | I]$  is well defined and finite for any admissible patterns  $\mathbf{U}$  and pattern ownership distribution  $\mathbf{P}$ :

$$\text{adm}_K = \{(\mathbf{P}, \mathbf{U}) \mid p_i, u_i \in H^1(\Omega), i = 1 : K; \quad \mathbf{P} : \Omega \rightarrow \Delta_{K-1}\}. \quad (13)$$



## 4.2 Breaking the Hidden Symmetry via Weak Supervision

Let  $S_K$  denote the permutation group of  $\{1, \dots, K\}$ . Each permutation  $\sigma \in S_K$  is a 1-to-1 map:

$$\sigma : \{1, \dots, K\} \rightarrow \{1, \dots, K\},$$

so that  $(\sigma(1), \dots, \sigma(K))$  is a re-arrangement of  $\{1, \dots, K\}$ . For any  $K$ -tuple  $\vec{F} = (f_1, \dots, f_K)$ , one defines

$$\vec{F}_\sigma = (f_{\sigma(1)}, f_{\sigma(2)}, \dots, f_{\sigma(K)}).$$

**Theorem 2 (Hidden Symmetry of SMS)** *For any  $\sigma \in S_K$ ,*

$$E[\mathbf{P}_\sigma, \mathbf{U}_\sigma | I] = E[\mathbf{P}, \mathbf{U} | I].$$

*In particular, suppose*

$$(\mathbf{P}^*, \mathbf{U}^*) = \operatorname{argmin}_{(\mathbf{P}, \mathbf{U}) \in \operatorname{adm}_K} E[\mathbf{P}, \mathbf{U} | I]$$

*is an optimal pair. Then for any  $\sigma \in S_K$ ,  $(\mathbf{P}_\sigma^*, \mathbf{U}_\sigma^*)$  is a minimizer as well.*

The proof is straightforward and thus omitted. Such symmetry not only worsens the non-uniqueness of the non-convex energy functional in (12), but also potentially jitters intermediate solutions in iterative computational schemes (i.e., hysterical transitions in the admissible space).

To break the permutation symmetry, we turn to a weak supervision scheme in which a user specifies  $K$  distinct domain patches:

$$Q_1, Q_2, \dots, Q_K,$$

and imposes the symmetry-breaking conditions:

$$p_i|_{Q_j} = \delta_{ij}, \quad i, j = 1 : K, \tag{14}$$

where  $\delta_{ij}$  denotes Kronecker’s delta. That is, a user requires each given patch  $Q_i$  to be a “pure” pattern exclusively labelled by  $i$ . Computationally this weak supervision process can be automated based on multiscale patch statistics as in the contemporary works on scene recognition [23, 24], or more generally, the learning theory [10, 41].



Figure 2: Examples of a 3-phase supervision and a 4-phase supervision to break the symmetry in the model. Such weak supervision can also be automated based on multiscale patch statistics [24, 23].

### 4.3 Existence Theorems for Non-Supervision and Supervision

In this subsection, we establish the existence theorems for the soft Mumford-Shah segmentation model (12) without or with the supervision (14).

**Theorem 3 (Existence Theorem for Unsupervised SMS)** *Suppose  $I \in L^2(\Omega)$ . Then for any positive modeling parameters  $(\lambda, \alpha, \varepsilon)$ , a minimizer to the unsupervised soft Mumford-Shah model (12) must exist.*

We will need the following lemma for the proof.

**Lemma 1** *Let  $(f_n | n)$  be a sequence of functions in  $L^2(\Omega)$ , and  $(p^n | n)$  a sequence of nonnegative measurable functions on  $\Omega$  and valued in  $[0, 1]$ . Suppose*

- (i)  $p^n \rightarrow p^*$ , a.e. on  $\Omega$ , and  $\int_{\Omega} p^* > 0$ ; and
- (ii)  $\int_{\Omega} f_n^2 p^n \leq A$  for some  $A > 0$  and  $n = 1 : \infty$ .

*Then there exists some function  $\rho \in L^2(\Omega)$ , such that*

- (a)  $\rho \geq 0$  and  $\int_{\Omega} \rho = 1$ , and
- (b) for some fixed  $B > 0$ ,  $|\int_{\Omega} f_n \rho| \leq B$  for  $n = 1 : \infty$ .

*Proof.* Denote the Lebesgue measure of a measurable set  $W$  by  $|W|$ . Since  $p^* \geq 0$  and  $\int_{\Omega} p^* > 0$ , there must exist some  $c > 0$ , such that

$$V = \{x \in \Omega \mid p^* > 2c\} \quad \text{has a finite but positive measure.}$$

On the other hand, by Egorov's theorem [14] on a.e. convergence, there must exist a subset  $W \subset V$ , such that

- (a')  $|V - W| \leq \frac{|V|}{2}$ , and hence  $|W| > 0$ , and
- (b')  $p^n \rightarrow p^*$  uniformly on  $W$ .

In particular, there exists some  $N$ , such that for any  $n > N$ ,  $p^n > c$  on  $W$ . Define

$$\rho(x) = \frac{1_W(x)}{|W|} \in L^2(\Omega).$$

Then  $\int_{\Omega} \rho = 1$ , and for any  $n > N$ ,

$$\int_{\Omega} f_n^2 \rho = \frac{1}{c|W|} \int_W f_n^2 c \leq \frac{1}{c|W|} \int_{\Omega} f_n^2 p^n \leq \frac{A}{c|W|}.$$

Thus by the Schwarz inequality (or  $E[X]^2 \leq E[X^2]$  in probability theory),

$$|\int_{\Omega} f_n \rho| \leq \left( \int_{\Omega} f_n^2 \rho \right)^{1/2} \leq \left( \frac{A}{c|W|} \right)^{1/2}, \quad n > N$$

The lemma holds if one defines  $B = \max \left( \left( \frac{A}{c|W|} \right)^{1/2}, |\int_{\Omega} f_1 \rho|, \dots, |\int_{\Omega} f_N \rho| \right)$ . □

We are ready to prove Theorem 3.

*Proof.* Take the special pattern distribution:

$$u_i \equiv 0, \quad i = 1 : K; \quad p_1 \equiv 1, \quad p_j \equiv 0, \quad j = 2 : K.$$

Then

$$E[\mathbf{P}, \mathbf{U} \mid I] = \lambda \int_{\Omega} I^2 < \infty.$$

Thus the infimum of the energy must be finite. Let  $(\mathbf{P}^n, \mathbf{U}^n \mid n) \subseteq \text{adm}_K$  (see (13)) be a minimizing sequence for the soft Mumford-Shah energy (12).

Due to the third term in the energy and the simplex constraint, for each channel  $i$ ,  $(p_i^n \mid n)$  must be bounded in  $H^1(\Omega)$ . By the  $L^2$ -weak compactness, there must exist some  $\mathbf{P}^* \in L^2(\Omega, \mathbb{R}^K)$ , and a subsequence of  $(\mathbf{P}^n \mid n)$ , which after relabelling shall still be denoted by  $(\mathbf{P}^n \mid n)$  for convenience, such that

$$\mathbf{P}^n \rightarrow \mathbf{P}^* \quad \text{in } L^2(\Omega, \mathbb{R}^K), \quad n \rightarrow \infty.$$

Then by the  $L^2$  lower semi-continuity of Sobolev measures,

$$9\varepsilon \int_{\Omega} |\nabla p_i^*|^2 \leq \liminf_{n \rightarrow \infty} 9\varepsilon \int_{\Omega} |\nabla p_i^n|^2, \quad i = 1 : K. \quad (15)$$

Furthermore, with possibly another round of subsequence refinement, one can assume

$$\mathbf{P}^n(x) \rightarrow \mathbf{P}^*(x), \quad a.e. \quad x \in \Omega, \quad n \rightarrow \infty.$$

Since the probability simplex  $\Delta_{k-1}$  is closed and  $\mathbf{P}^n(x) \in \Delta_{K-1}$ , one concludes that

$$\mathbf{P}^*(x) \in \Delta_{K-1}, \quad a.e. \quad x \in \Omega.$$

And by Fatou's Lemma [14, 25], one has

$$\int_{\Omega} \frac{(p_i^*(1 - p_i^*))^2}{\varepsilon} \leq \liminf_{n \rightarrow \infty} \int_{\Omega} \frac{(p_i^n(1 - p_i^n))^2}{\varepsilon}, \quad i = 1 : K. \quad (16)$$

(In fact, the equality holds by *Lebesgue's Dominated Convergence* [25].)

After the above subsequence selection on  $\mathbf{P}^n$ 's, one naturally has an associated subsequence of  $(\mathbf{U}^n \mid n)$ , which for convenience is still denoted by  $(\mathbf{U}^n \mid n)$  after relabelling. For each *specific* channel  $i$ , we then consider two scenarios separately.

Suppose  $p_i^*(x) \equiv 0, a.e. \quad x \in \Omega$ . We then define for that channel

$$u_i^*(x) \equiv 0, \quad x \in \Omega. \quad (17)$$

Such a channel is called a “dumb” channel.

Otherwise, one must have  $\int_{\Omega} p_i^* > 0$ , and from the first term in (12),

$$\int_{\Omega} (u_i^n - I)^2 p_i^n \leq \text{const.}, \quad n = 1 : \infty.$$

Since  $\int_{\Omega} I^2 p_i^n \leq \int_{\Omega} I^2$ , by the triangle inequality,

$$\int_{\Omega} (u_i^n)^2 p_i^n \leq \text{const.}, \quad n = 1 : \infty,$$

where the constant only depends on  $I$  and the model parameters. Then by Lemma 1, there exists some  $\rho_i(x) \geq 0$ , with  $\int_{\Omega} \rho_i = 1$ , some constant  $B_i > 0$  such that

$$\left| \int_{\Omega} u_i^n \rho_i \right| \leq B_i, \quad n = 1 : \infty.$$

On the other hand, by the second term in the energy (12),

$$\int_{\Omega} |\nabla u_i^n|^2 \leq C_i = C_i(I, \lambda, \alpha, \varepsilon), \quad n = 1 : \infty,$$

for some constant  $C_i$  independent of  $n$ . Then by the generalized Pointcaré inequality [13, 25] on  $\Omega$ ,

$$\|w - \langle w, \rho_i \rangle\|_{L^2} \leq A_i \|\nabla w\|_{L^2},$$

where  $A_i = A_i(\rho_i, \Omega)$  is independent of  $w \in H^1(\Omega)$ , one concludes that

$$\|u_i^n\|_{L^2} \leq D_i = D_i(A_i, B_i, C_i), \quad n = 1 : \infty,$$

for some constant  $D_i$ . As a result,  $(u_i^n | n)$  must be bounded in  $H^1(\Omega)$ . By the  $L^2$ -weak compactness of bounded  $H^1$ -sequences, there is a subsequence of  $(u_i^n | n)$ , for convenience still denoted by  $(u_i^n | n)$  after relabelling, such that

$$u_i^n \rightharpoonup u_i^* \in L^2(\Omega), \quad n \rightarrow \infty,$$

converging in the sense of both  $L^2$  and almost everywhere. Then by the lower semi-continuity,

$$\int_{\Omega} |\nabla u_i^*|^2 \leq \liminf_{n \rightarrow \infty} \int_{\Omega} |\nabla u_i^n|^2.$$

Finally, since  $u_i^n(x) \rightarrow u_i^*(x)$  and  $p_i^n(x) \rightarrow p_i^*(x)$ , a.e.  $x \in \Omega$ , Fatou's Lemma gives

$$\int_{\Omega} (u_i^* - I)^2 p_i^* \leq \liminf_{n \rightarrow \infty} \int_{\Omega} (u_i^n - I)^2 p_i^n.$$

Combining both cases just analyzed above, we have established that

$$\lambda \sum_{i=1}^K \int_{\Omega} (u_i^* - I)^2 p_i^* + \alpha \sum_{i=1}^K \int_{\Omega} |\nabla u_i^*|^2 \leq \liminf_{n \rightarrow \infty} \lambda \sum_{i=1}^K \int_{\Omega} (u_i^n - I)^2 p_i^n + \alpha \sum_{i=1}^K \int_{\Omega} |\nabla u_i^n|^2. \quad (18)$$

Together with (15) and (16), this implies

$$E[\mathbf{P}^*, \mathbf{U}^* | I] \leq \liminf_{n \rightarrow \infty} E[\mathbf{P}^n, \mathbf{U}^n | I] = \inf_{(\mathbf{P}, \mathbf{U})} E[\mathbf{P}, \mathbf{U} | I],$$

and hence  $(\mathbf{P}^*, \mathbf{U}^*)$  must be a minimizer.  $\square$

The proof reveals an important behavior of the model (12). If certain channel  $i$  becomes dumb (i.e.,  $p_i^* \equiv 0$ ), it has often been introduced unnecessarily in the first place, and the associated optimal pattern  $u_i^*$  could be any featureless constant image.

**Theorem 4** *Suppose  $I \in L^2(\Omega)$ . Then an optimal pattern-ownership pair must exist to the soft Mumford-Shah segmentation model (12) with supervision (14), assuming that each patch  $Q_i$  has a positive Lebesgue measure  $|Q_i| > 0$ .*

*Proof.* The proof is almost identical to the unsupervised case above, and simplifies substantially by noticing that no channel could become dumb due to supervision. Furthermore, the functions  $\rho_i$ 's in the above proof can be directly set to be

$$\rho_i = \frac{1}{|Q_i|} 1_{Q_i}(x), \quad i = 1 : K,$$

without the necessity of turning to Lemma 1.  $\square$

#### 4.4 Mixture of Homogeneous Gaussians

When each pattern  $i$  is a homogeneous Gaussian  $N(m_i, \sigma)$  with a distinct mean value  $m_i$ , one has

$$u_i(x) \equiv m_i, \quad x \in \Omega, \quad i = 1 : K.$$

Define  $\mathbf{m} = (m_1, \dots, m_K)$ . As a result, the soft Mumford-Shah model (12) simplifies to

$$\min_{(\mathbf{P}, \mathbf{m})} E[\mathbf{P}, \mathbf{m} \mid I] = \lambda \sum_{i=1}^K \int_{\Omega} (I - m_i)^2 p_i + \sum_{i=1}^K \int_{\Omega} \left( 9\varepsilon |\nabla p_i|^2 + \frac{(p_i(1-p_i))^2}{\varepsilon} \right). \quad (19)$$

**Theorem 5** *Suppose  $I \in L^2(\Omega)$ . Then a minimizer pair  $(\mathbf{P}^*, \mathbf{m}^*)$  to  $E[\mathbf{P}, \mathbf{m} \mid I]$  exists for both the unsupervised and supervised cases.*

The proof can be derived readily from the previous general cases and is hence left out. When  $K = 2$ , a similar model was proposed earlier by Shen [37] under the symmetrization transform:

$$p_1(x) = \frac{1 - z(x)}{2}, \quad p_2(x) = \frac{1 + z(x)}{2}, \quad z \in [-1, 1].$$

The model (19) could be considered as the soft version of Chan and Vese's model [8] from the point of view of region-based active contours. Chan and Vese have demonstrated that such a piecewise constant Mumford-Shah model (or the CV model as popularly referred to in the present literature) is already powerful enough for a number of applications including medical imaging.

## 5 Euler-Lagrange Equations and Computation on $(K - 1)$ -Simplex

### 5.1 Euler-Lagrange Equations on $(K - 1)$ -Simplex

To minimize the energy for the soft Mumford-Shah segmentation

$$E[\mathbf{P}, \mathbf{U} \mid I] = \lambda \sum_{i=1}^K \int_{\Omega} (u_i - I)^2 p_i + \alpha \sum_{i=1}^K \int_{\Omega} |\nabla u_i|^2 + \sum_{i=1}^K \int_{\Omega} \left( 9\varepsilon |\nabla p_i|^2 + \frac{(p_i(1-p_i))^2}{\varepsilon} \right), \quad (20)$$

one resorts to its gradient-descent flow or Euler-Lagrange equations. In this section, we discuss these equations and their practical computational schemes.

The first-order partial variation on  $\mathbf{U}$  given  $\mathbf{P}$  leads to, for  $i = 1 : K$ ,

$$\alpha \Delta u_i + \lambda(I - u_i)p_i = 0, \quad \text{on } \Omega; \quad \frac{\partial u_i}{\partial \mathbf{n}} = 0, \quad \text{along } \partial\Omega,$$

where  $\mathbf{n}$  stands for the outer normal vector field along  $\partial\Omega$ . Thus the Euler-Lagrange equations on the patterns are all in the form of linear Poisson equations with variable coefficient fields:

$$-\alpha \Delta u_i + (\lambda p_i)u_i = f_i, \quad i = 1 : K,$$

with Neumann adiabatic boundary conditions, where the source terms are  $f_i(x) = \lambda p_i(x)I(x)$ .

The first-order variation on the ownerships  $\mathbf{P}$  is carried out on the probability  $(K - 1)$ -simplex  $\Delta_{K-1}$ , which is a compact manifold (with border) of codimension 1 embedded in  $\mathbb{R}^K$ . Chan and Shen [6] developed a general framework for modelling and computing image features that “live” on general manifolds, and especially those that are embedded in  $\mathbb{R}^K$ . We shall follow the approach there.

Without the simplex constraint on the ownerships, for any given  $\mathbf{U}$ , the first order variation of the soft energy  $E$  under  $\mathbf{P} \rightarrow \mathbf{P} + \delta\mathbf{P}$  is given by

$$\delta E = \int_{\Omega} \sum_{i=1}^K V_i \delta p_i dx + \int_{\partial\Omega} \sum_{i=1}^K v_i \delta p_i d\mathcal{H}^1,$$

where  $\mathcal{H}^1$  is the 1-D Hausdorff measure along  $\partial\Omega$ , and

$$V_i = \lambda(u_i - I)^2 - 18\varepsilon\Delta p_i + 2\varepsilon^{-1}p_i(1 - p_i)(1 - 2p_i), \quad (21)$$

$$v_i = 18\varepsilon \frac{\partial p_i}{\partial \mathbf{n}}, \quad \text{along } \partial\Omega. \quad (22)$$

Define  $\mathbf{V} = (V_1, \dots, V_K)$  and  $\mathbf{v} = (v_1, \dots, v_K)$ . Then

$$\delta E = \int_{\Omega} \mathbf{V} \cdot \delta\mathbf{P} dx + \int_{\partial\Omega} \mathbf{v} \cdot \delta\mathbf{P} d\mathcal{H}^1,$$

which holds for any *free* variation of  $\mathbf{P}$  in  $\mathbb{R}^K$ , or one writes in the free-gradient form

$$\frac{\partial E}{\partial_f \mathbf{P}} = \mathbf{V}|_{\Omega} + \mathbf{v}|_{\partial\Omega}.$$

In reality,  $\mathbf{P} \in \Delta_{K-1}$ . Let  $T_{\mathbf{P}}\Delta_{K-1}$  denote the tangent space of  $\Delta_{K-1}$  at any single point  $\mathbf{P} \in \Delta_K$ , and

$$\pi : T_{\mathbf{P}}\mathbb{R}^K \rightarrow T_{\mathbf{P}}\Delta_{K-1}$$

the orthogonal projection onto the tangent space in  $\mathbb{R}^K$ . Since the normal direction of the tangent plane is given by  $\mathbf{1}_K/\sqrt{K} = (1, \dots, 1)/\sqrt{K}$ , the projection operator is explicitly given by, for any  $\mathbf{w} \in T_{\mathbf{P}}\mathbb{R}^K$ ,

$$\pi(\mathbf{w}) = \mathbf{w} - \mathbf{1}_K \langle \mathbf{w}, \mathbf{1}_K \rangle / K = \mathbf{w} - \langle \mathbf{w} \rangle \mathbf{1}_K, \quad \text{with } \langle \mathbf{w} \rangle = \frac{1}{K} \sum_{i=1}^K w_i.$$

The constrained gradient of  $E$  on  $\Delta_{K-1}$  is therefore given by

$$\frac{\partial E}{\partial \mathbf{P}} = \pi \left( \frac{\partial E}{\partial_f \mathbf{P}} \right) = (\mathbf{V} - \langle \mathbf{V} \rangle \mathbf{1}_K)|_{\Omega} + (\mathbf{v} - \langle \mathbf{v} \rangle \mathbf{1}_K)|_{\partial\Omega}.$$

In particular, the system of Euler-Lagrange equations on  $\mathbf{P}$  given  $\mathbf{U}$  is given by

$$\begin{cases} V_i(x) - \langle \mathbf{V} \rangle(x) = 0, & x \in \Omega, \\ v_i(z) - \langle \mathbf{v} \rangle(z) = 0, & z \in \partial\Omega, \end{cases} \quad (23)$$

for  $i = 1 : K$ . The coupling among different channels is evident from these two formulae.

**Lemma 2** *Suppose  $\mathbf{P} : \Omega \rightarrow \Delta_{K-1}$ . Then for any  $z \in \partial\Omega$ ,  $\langle \mathbf{v} \rangle(z) = 0$ , where the boundary “flux”  $\mathbf{v}$  is defined in (22).*

*Proof.* This is obtained by direct computation: at any  $z \in \partial\Omega$ ,

$$\begin{aligned}\langle \mathbf{v} \rangle &= \frac{1}{K} \sum_{i=1}^K v_i = \frac{18\varepsilon}{K} \sum_{i=1}^K \frac{\partial p_i}{\partial \mathbf{n}} \\ &= \frac{18\varepsilon}{K} \frac{\partial}{\partial \mathbf{n}} \left( \sum_{i=1}^K p_i \right) = \frac{18\varepsilon}{K} \frac{\partial 1}{\partial \mathbf{n}} = 0.\end{aligned}$$

□

As a result, the boundary conditions in (23) simplify to the ordinary Neumann conditions  $\partial p_i / \partial \mathbf{n} = 0, i = 1 : K$ . Combining all the above derivations, we have established the following theorem.

**Theorem 6 (Euler-Lagrange Equations)** *The system of Euler-Lagrange equations of  $E[\mathbf{P}, \mathbf{U} \mid I]$  are given by*

$$\begin{aligned}-\alpha \Delta u_i + (\lambda p_i) u_i &= (\lambda p_i) I, \\ -18\varepsilon \Delta p_i + 2\varepsilon^{-1} p_i (1 - p_i) (1 - 2p_i) &= \langle \mathbf{V} \rangle - \lambda (u_i - I)^2, \quad i = 1 : K,\end{aligned}\tag{24}$$

on  $\Omega$ , all with Neumann boundary conditions along  $\partial\Omega$ . Here  $\mathbf{V} = \mathbf{V}(\mathbf{P}, \mathbf{U})$  is defined as in (21). Furthermore, under supervision (14), the ownerships must satisfy the interpolation conditions:

$$p_i|_{Q_j} = \delta_{i,j}, \quad i, j = 1 : K,$$

Or equivalently, the equations on  $p_i$ 's in (24) hold on  $\Omega \setminus (\cup_{i=1}^K Q_i)$  with

$$\text{Neumann conditions along } \partial\Omega, \text{ and Dirichlet conditions along } \cup_{i=1}^K \partial Q_i : p_i|_{\partial Q_j} = \delta_{i,j}.$$

Similarly, one has the following result for the piecewise constant SMS model (19), which carries much lower complexity compared with the full SMS model.

**Proposition 2 (Euler-Lagrange Equations for Piecewise Constant SMS)** *The Euler-Lagrange equations for  $E[\mathbf{P}, \mathbf{m} \mid I]$  in (19) are given by*

$$\begin{aligned}m_i = \langle I \rangle_{p_i} &:= \frac{\int_{\Omega} I p_i}{\int_{\Omega} p_i}, \\ -18\varepsilon \Delta p_i + 2\varepsilon^{-1} p_i (1 - p_i) (1 - 2p_i) &= \langle \mathbf{V} \rangle - \lambda (m_i - I)^2, \quad i = 1 : K,\end{aligned}\tag{25}$$

with Neumann conditions for all the ownerships  $p_i$ 's along  $\partial\Omega$ .

## 5.2 Computation of the Euler-Lagrange Equations

Computationally, as well practiced in multivariate optimization problems, (24) and (25) can be solved via the algorithm of *alternating minimization* (AM) [12, 36]. The AM algorithm is closely connected to the celebrated EM (expectation-maximization) algorithm in statistical estimation problems with hidden variables [20, 26]. In the current context, the ownership distributions  $p_i$ 's could be treated as the hidden variables.

Like EM, the AM algorithm is progressive. Given the current ( $t = n$ ) best estimation of the patterns  $\mathbf{U}^n = (u_i^n \mid i = 1 : K)$ , by solving

$$\mathbf{P}^n = \operatorname{argmin}_{\mathbf{P}} E[\mathbf{P} \mid \mathbf{U}^n, I],$$

or equivalently,

$$-18\varepsilon\Delta p_i + 2\varepsilon^{-1}p_i(1-p_i)(1-2p_i) = \langle \mathbf{V}^n \rangle - \lambda(u_i^n - I)^2, \quad i = 1 : K, \quad (26)$$

with Neumann boundary conditions, one obtains the current best estimation of the ownerships  $\mathbf{P}^n = (p_i^n \mid i = 1 : K)$ . Subsequently, based on  $\mathbf{P}^n$ , by solving

$$\mathbf{U}^{n+1} = \operatorname{argmin} E[\mathbf{U} \mid \mathbf{P}^n, I],$$

or equivalently,

$$-\alpha\Delta u_i + (\lambda p_i^n)u_i = (\lambda p_i^n)I, \quad i = 1 : K, \quad (27)$$

with Neumann boundary conditions, one completes a single round of pattern updating  $\mathbf{U}^n \rightarrow \mathbf{U}^{n+1}$ . The same procedure applies to the piecewise constant soft Mumford-Shah equations in (25).

Since the system (27) is linear and decoupled, the main computational complexity resides in the integration of (26), which is coupled and nonlinear due to the simplex constraint and the double-well potential in the energy. Define  $e_i(x) = (u_i(x) - I(x))^2$  and  $\mathbf{e} = (e_i \mid i = 1 : K)$ . In order to solve

$$-18\varepsilon\Delta p_i + 2\varepsilon^{-1}p_i(1-p_i)(1-2p_i) = \langle \mathbf{V} \rangle - \lambda e_i \quad (28)$$

given  $\mathbf{e}$  and  $\mathbf{V} = \mathbf{V}(\mathbf{P}, \mathbf{U}) = \mathbf{V}(\mathbf{P}, \mathbf{e})$  (see (21)), first notice that

$$\begin{aligned} \langle \mathbf{V}(\mathbf{P}, \mathbf{e}) \rangle &= \frac{1}{K} \sum_{i=1}^K (-18\varepsilon\Delta p_i + e_i + 2\varepsilon^{-1}p_i(1-p_i)(1-2p_i)) \\ &= \frac{1}{K} \sum_{i=1}^K e_i + \frac{2\varepsilon^{-1}}{K} \sum_{i=1}^K (2p_i^3 - 3p_i^2) + \frac{2\varepsilon^{-1}}{K}, \end{aligned}$$

since  $\sum_{i=1}^K p_i = 1$  and  $\Delta \left( \sum_{i=1}^K p_i \right) = 0$ . We also split the double-potential force in (28) by

$$p_i(1-p_i)(1-2p_i) = p_i(1-p_i)^2 - p_i^2(1-p_i).$$

In combination, the nonlinear equation (28) can then be solved iteratively:

$$\dots \rightarrow \mathbf{P}^{(j)} \rightarrow \mathbf{P}^{(j+1)} \rightarrow \dots$$

by the following linearization procedure:

$$\begin{aligned} -18\varepsilon\Delta p_i^{(j+1)} + 2\varepsilon^{-1}p_i^{(j+1)} \left(1 - p_i^{(j)}\right)^2 &= f_i^{(j)}, \\ f_i^{(j)} &= -e_i + \langle V(\mathbf{P}^{(j)}, \mathbf{e}) \rangle + 2\varepsilon^{-1} \left(p_i^{(j)}\right)^2 \left(1 - p_i^{(j)}\right), \end{aligned}$$

with Neumann adiabatic boundary conditions for all the channels  $i = 1 : K$ . This system of linear Poisson equations can be conveniently integrated using any elliptic solvers. The detailed *numerical analysis* on the convergence rates of the entire algorithm above, however, is still an open problem and well deserves some systematic investigation.



## 6 Computational Examples

In this section, we present the computational results of the proposed soft Mumford-Shah model. Notice that the extension of the above SMS models to color images is straightforward by having the gray values  $u_i$ 's replaced by RGB vectors. (We, however, must remind the reader that *perceptually* RGB may not be the most ideal representation of colors compared with other nonlinear approaches, e.g., brightness-chromaticity [6] and HSV [5].)

Figures 3 and 4 illustrate the performance of the SMS model on two synthetic images with multiple phases. Figure 3 shows a typical T-junction and Figure 4 a 3-phase image with a narrow bottleneck. Plotted in the figures are the hard segments obtained from the SMS model via the hardening formulae (2) and (3).

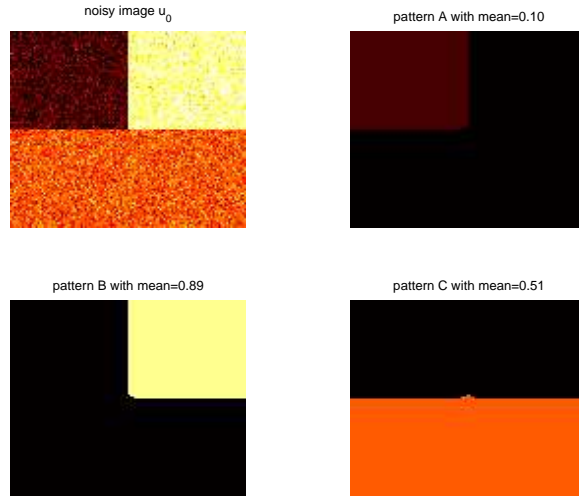


Figure 3: Synthetic image of a T-junction: hard segmentation from the SMS model via “hardening” formulae (2) and (3). The 120-degree regularization behavior at the junction point is also well known in the classical Mumford-Shah model [33].

Plotted in Figure 5 are the hardened segments obtained from the soft Mumford-Shah segmentation model via formulae (2) and (3). For this application, a user specifies three small patches (three rectangles in this example)  $Q_1, Q_2$ , and  $Q_3$ , and the SMS model proceeds with the extra interpolation conditions in (14) for the ownerships.

In Figure 6, another example of a natural image is segmented via the SMS model and the “hardening” formulae (2) and (3). A user supervises with three patches  $Q_1, Q_2$  and  $Q_3$ , and designates the two on the body to a pattern ownership  $p_{\text{body}}$  and the third (from the ocean) to  $p_{\text{ocean}}$ . If the three are treated as distinct patterns, the SMS model still works, but one needs an extra step of high-level vision processing (e.g., based on Grenander’s graph models [19]) to group the skin-tone and the purple-shirt patterns in order to capture the entire body faithfully.

Finally, plotted in Figures (7) and (8) are the ownerships from the SMS model based on the 3-phase and 4-phase supervision separately in Figure 2. The stochastic nature of the outcomes (i.e., the softly transiting ownerships  $p_i$ 's instead of hard segmentation) is closer to the way how a human subject may perceive such a natural scene. In particular, the SMS model seems to be consistent with the most recent theory that *hard* pattern segments may not be absolutely necessary for natural scene recognition [24, 23].

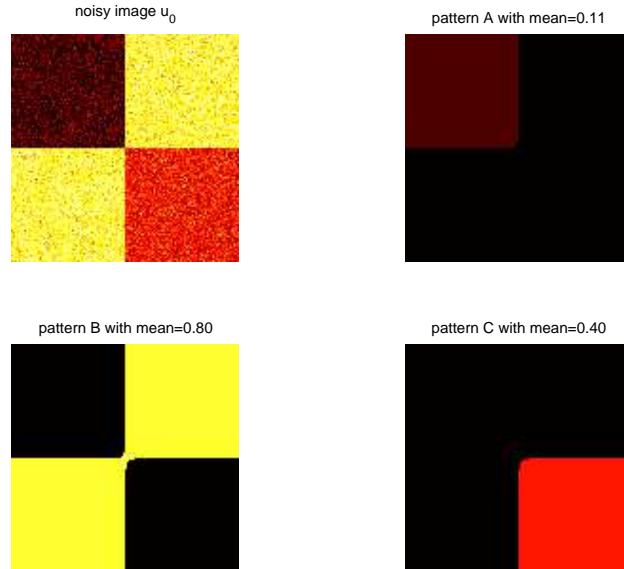


Figure 4: Synthetic image of a narrow bottleneck: hard segmentation from the SMS model via “hardening” formulae (2) and (3). The thickening regularization at the bottleneck junction can be explained similarly by the classical Mumford-Shah model for which minimum-surface or “soap-foam” behavior arises due to the surface tension energy. Also see the recent work by Kohn and Slastikov [22] for the singularity analysis of a similar problem arising from micromagnetism.

## Acknowledgments

The author is very grateful to Prof. Alan Yuille for an enlightening discussion after the current work was first presented. For their generous teaching and continual inspiration, the author is always profoundly indebted to Professors Gil Strang, Tony Chan, Stan Osher, David Mumford, Jean-Michel Morel, and Stu Geman. The author must thank his wonderful former teacher, Prof. Dan Kerstan at the Psychology Department of the University of Minnesota, for his first introduction on mixture image models and stochastic visual processing several year ago. The author also thanks the Institute of Mathematics and its Applications (**IMA**) and the Institute of Pure and Applied Mathematics (**IPAM**) for their persistent role in supporting this new emerging field. Finally, the author would like to dedicate this paper to his dear friends Yingnian Wu and Song-Chun Zhu for the unique friendship cultivated by the intellectually rich soil of vision and cognitive sciences.

## References

- [1] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via  $\Gamma$ -convergence. *Comm. Pure Appl. Math.*, 43:999–1036, 1990.
- [2] L. Ambrosio and V. M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital.*, 6-B:105–123, 1992.
- [3] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [4] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag New York, Inc., 1998.

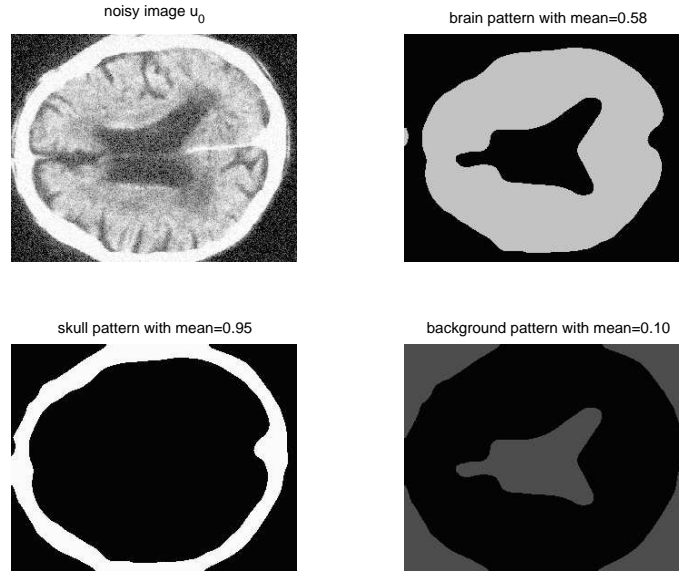


Figure 5: A real noisy brain image: hard segmentation from the SMS model via “hardening” formulae (2) and (3).

- [5] T. F. Chan, S.-H. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color models. *J. Visual Comm. Image Rep.*, 12(4):422–435, 2001.
- [6] T. F. Chan and J. Shen. Variational restoration of non-flat image features: models and algorithms. *SIAM J. Appl. Math.*, 61(4):1338–1361, 2000.
- [7] T. F. Chan and J. Shen. *Image Processing and Analysis: variational, PDE, wavelet, and stochastic methods*. SIAM Publisher, Philadelphia, 2005.
- [8] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, 2001.
- [9] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York and Oxford, 1987.
- [10] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. AMS*, 39(1):1–49, 2001.
- [11] J. L. Ericksen. *Advances in Liquid Crystals*, chapter Equilibrium theory of liquid crystals, pages 233–299. Academic Press, New York, 1976.
- [12] S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.
- [13] L. C. Evans. *Partial Differential Equations*. Amer. Math. Soc., 1998.
- [14] G. B. Folland. *Real Analysis - Modern Techniques and Their Applications*. John Wiley & Sons, Inc., second edition, 1999.
- [15] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.

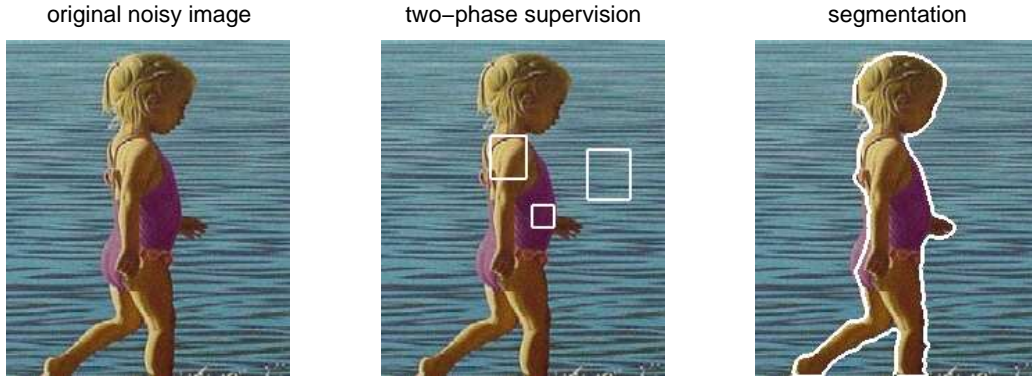


Figure 6: Hard segmentation from the SMS model via “hardening” formulae (2) and (3), based on a 2-phase supervision. Denote the two rectangles on the body by  $Q_1$  and  $Q_2$ , and the third by  $Q_3$ . Supervision provides the ownership interpolation condition:  $p_{\text{body}} = 1$  on  $Q_1 \cup Q_2$  and 0 on  $Q_3$ , while  $p_{\text{ocean}} = 1$  on  $Q_3$  and 0 on  $Q_1 \cup Q_2$ . Patch selection can also be automated based on multiscale patch statistics (e.g., Li and Perona [23]).

- [16] W. Gibbs. *Elementary Principles of Statistical Mechanics*. Yale University Press, 1902.
- [17] V. L. Ginzburg and L. D. Landau. On the theory of superconductivity. *Soviet Phys. JETP*, 20:1064–1082, 1950.
- [18] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston, 1984.
- [19] U. Grenander. *Lectures in Pattern Theory. I. II. and III.* Springer, 1976-1981.
- [20] A. D. Jepson and M. J. Black. Mixture models for image representation. *PRECARN ARK Project Technical Report ARK96-PUB-54*, March 1996.
- [21] D. C. Knill and W. Richards. *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996.
- [22] R. V. Kohn and V. V. Slastikov. Geometrically constrained walls. *Preprint*, 2004.
- [23] F.-F. Li and P. Perona. A Bayesian hierarchical model for learning natural scene categories. *IEEE CVPR*, 2(2):524–531, 2005.
- [24] F.-F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *Proc. Nat’l Acad. Sci. (USA)*, 99(14):9596–9601, 2002.
- [25] E. H. Lieb and M. Loss. *Analysis*. Amer. Math. Soc., second edition, 2001.
- [26] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley-Interscience, 2002.
- [27] R. March. Visual reconstruction with discontinuities using variational methods. *Image Vision Comput.*, 10:30–38, 1992.
- [28] R. March and M. Dozio. A variational method for the recovery of smooth boundaries. *Image Vision Comput.*, 15:705–712, 1997.
- [29] G. Dal Maso. *An Introduction to  $\Gamma$ -Convergence*. Birkhauser, Boston, 1992.

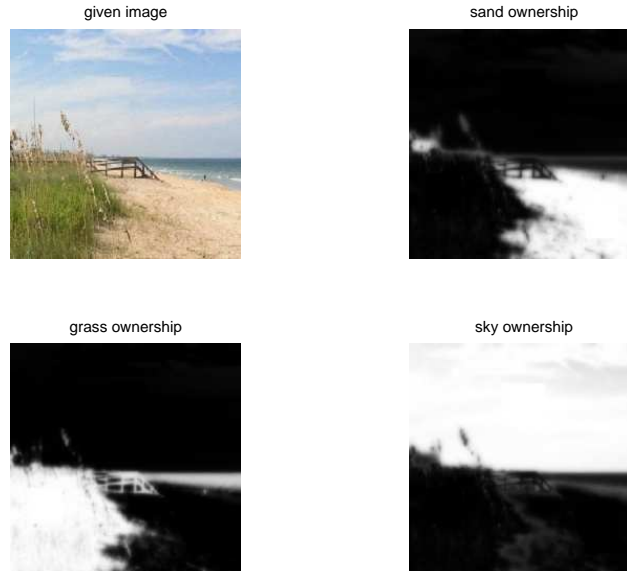


Figure 7: Soft Mumford-Shah segmentation with three phases corresponding to the supervision on the *left* panel of Figure 2. Plotted here are the three ownership distributions  $p_1(x)$ ,  $p_2(x)$ , and  $p_3(x)$ . Due to “under”-supervision, namely the number  $K$  of specified patterns is less than that of the visually meaningful ones, the grass pattern has “absorbed” the ocean pattern due to the greenish color they happen to share.

- [30] L. Modica and S. Mortola. Un esempio di Gamma-convergenza. *Boll. Un. Mat. Ital. B*, 5-14(1):285–299, 1977.
- [31] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*, volume 14 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Boston, 1995.
- [32] D. Mumford. *Geometry Driven Diffusion in Computer Vision*, chapter “The Bayesian rationale for energy functionals”, pages 141–153. Kluwer Academic, 1994.
- [33] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.
- [34] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [35] B. Sandberg, T. F. Chan, and L. Vese. A level-set and Gabor-based active contour algorithm for segmenting textured images. UCLA Department of Mathematics CAM report 02-39, 2002.
- [36] J. Shen. Bayesian video dejittering by BV image model. *SIAM J. Appl. Math.*, 64(5):1691–1708, 2004.
- [37] J. Shen.  $\Gamma$ -convergence approximation to piecewise constant Mumford-Shah segmentation. *Lec. Notes Comp. Sci.*, 3708:499–506, 2005.
- [38] J. Shen. On the foundations of vision modeling II. Mining of mirror symmetry of 2-D shapes. *J. Visual Comm. Image Rep.*, 16:250–270, 2005.

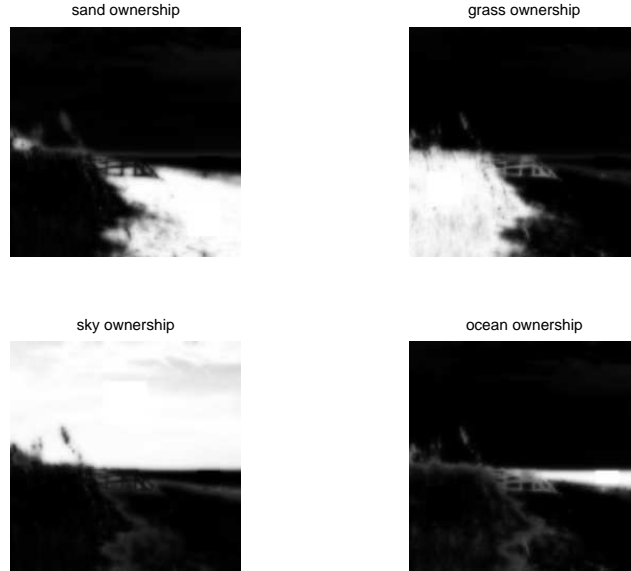


Figure 8: Soft Mumford-Shah segmentation with four phases corresponding to the supervision on the *right* panel of Figure 2. Plotted here are the four ownership distributions  $p_1(x)$ ,  $p_2(x)$ ,  $p_3(x)$ , and  $p_4(x)$ . Unlike Figure 7, the narrow ocean pattern is now softly segmented due to the extra fourth patch  $Q_4$ .

- [39] J. Shen. Piecewise  $H^{-1} + H^0 + H^1$  images and the Mumford-Shah-Sobolev model for segmented image decomposition. *Appl. Math. Res. Exp.*, in press, 2005.
- [40] J. Shen and Y.-M. Jung. Weberized Mumford-Shah model with Bose-Einstein photon noise. *Appl. Math. Optim.*, to appear, 2005.
- [41] S. Smale and D.-X. Zhou. Shannon sampling and function reconstruction from point values. *Bull. Amer. Math. Soc.*, 41:279–305, 2004.
- [42] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, MA, 1993.
- [43] Z. Tu and S. C. Zhu. Image segmentation by data-driven Markov chain Monte Carlo. *IEEE Trans. PAMI*, 24(5):657–673, 2002.
- [44] S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principle and its applications to texture modeling. *Neural Computation*, 9:1627–1660, 1997.
- [45] S. C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(9):884–900, 1996.