Coarse-to-Fine Segmentation With Shape-Tailored Continuum Scale Spaces

Naeemullah Khan[1], Byung-Woo Hong[2], Anthony Yezzi[3], and Ganesh Sundaramoorthi[1]
[1]KAUST, Saudi Arabia [2] Chung-Ang University, Korea [3] Georgia Tech, USA
{naeemullah.khan,ganesh.sundaramoorthi}@kaust.edu.sa, hong@cau.ac.kr, ayezzi@ece.gatech.edu

Abstract

We formulate an energy for segmentation that is designed to have preference for segmenting the coarse over fine structure of the image, without smoothing across boundaries of regions. The energy is formulated by integrating a continuum *of scales from a scale space computed from the heat equation within regions. We show that the energy can be optimized without computing a continuum of scales, but instead from a* single *scale. This makes the method computationally efficient in comparison to energies using a discrete set of scales. We apply our method to texture and motion segmentation. Experiments on benchmark datasets show that a continuum of scales leads to better segmentation accuracy over discrete scales and other competing methods.*

1. Introduction

Segmentation of images using low-level cues plays a key role in computer vision. An image consists of many different structures at different *scales*, and thus the notion of *scale space* [24], which consists of blurs of the image at all degrees, has been central to computer vision. The need for incorporating scale space in segmentation is well-recognized [40]. Further, there is evidence from human visual studies (e.g., [18, 35]) that the coarse scale, i.e., from high levels of blurring, is predominantly processed before the fine scale. This *coarse-to-fine* principle has led to many efficient algorithms that are able to capture the coarse structure of the solution, which is often most important in computer vision. Therefore, it is natural for segmentation algorithms to use scale space and operate in a coarse-to-fine fashion.

Existing methods for segmentation that incorporate scale have either one of the following limitations. First, most segmentation methods (e.g., [6, 25, 2]) based on scale spaces consider *global* scale spaces that are computed on the whole image, which does not capture the fact that there exist multiple *regions* of the segmentation at different scales, and this could lead to the removal and/or displacement of important structures in the image, for instance, when large structures are blurred across small ones, leading to an inaccurate segmentation. Second, algorithms that use a coarse-to-fine principle (e.g., [5, 33]) do so *sequentially* (see Figure 1) so that the algorithm operates at the coarser scale and then uses the result to initialize computation at a finer scale. While this warm start may influence the finer scale result, there is no guarantee that the coarse structure of the segmentation is preserved in the final solution.

In this paper, we develop an algorithm that simultaneously addresses these two issues. Specifically, we formulate a novel multi-region energy for segmentation, which integrates a *continuum* of scales from *Shape-Tailored Scale Spaces*. These scale spaces are defined within regions of the segmentation, and thus they prevent removal or displacement of important structures. By integrating over a *continuum* of scales of the scale space determined by the heat equation, we show that this energy has preference to coarse structure of the data without ignoring the fine structure. We show that it operates in a *parallel* coarse-to-fine fashion (see Figure 1). That is, it is initially dominated by the coarse structure of the data, then segments finer structure of the data, while preserving the structure from the coarse-scale of the data. We provide analytic solutions for the optimization of the energy, which leads to a computationally more efficient method than similar energies integrating discrete scales. We apply our algorithm to the problem of texture segmentation, and show our method outperforms discrete scale spaces and existing state of the art. We also apply our method to motion segmentation, show the advantage of the shape-tailored continuum scale space, and show outperformance against existing state of the art.

1.1. Related Work

Scale space theory [24, 53, 15, 27] has a long and rich history as a theory for analyzing images, and we only provide brief highlights. The idea is that an image consists of structures at various different scales (e.g., a leaf of a tree exists at a different scale than a forest), and thus to analyze an image without a-priori knowledge, it is necessary to consider the image at *all* scales. This is accomplished by blurring the image at a continuum of kernel sizes. The

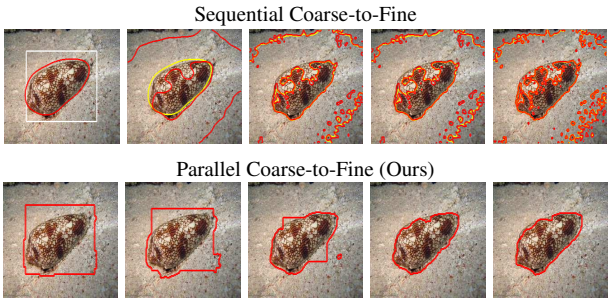

Figure 1. [Top]: Sequential coarse-to-fine methods use the result of segmentation (red) from the coarse scale to initialize (yellow) the finer scales, and may lose coarse structure of the coarse segmentation solution without additional heuristics. Note that the result of segmentation of the coarse scale is the left image in red (the blurred image is not shown), and towards the right segmentation is done at finer scales. [Bottom]: Our parallel coarse-to-fine approach considers a continuum of scales all at once and has a coarse-to-fine property. The evolution is shown from left to right.

most common kernel is a Gaussian, which is known to be the only scale space satisfying certain axioms such as not introducing any new features as the image is blurred [29]. Scale space has been used to analyze structures in images (e.g., [13, 50, 29, 44]). This has had wide ranging applications in stereo and optical flow [31], reconstruction [20, 49], key-point detection in wide-baseline matching [30], design of descriptors for matching [17], shape matching [7], and curve evolution [43], among others.

Gaussian scale spaces have also been used in image segmentation, most notably in texture segmentation [14, 39, 6, 25, 42], which occur frequently in natural images [2]. While these methods capture important scale information, they use a *global* scale space defined on the entire image, which does not capture the characteristic scale of features within regions and blurs across segmentation boundaries. Anisotropic scale spaces [40, 4] have been applied to reduce blurring across boundaries, but this could blur across regions where edges are not salient. Recently, [23] have addressed this issue by computing scales locally within the evolving regions of the segmentation. However, only a discrete number of scales are used and thus the method does not exhibit coarse-to-fine behavior. Such methods for segmentation have been numerically implemented with various optimization methods, including level sets [38], convex methods [41, 26], and others [47]. The energy we consider is not convex, and thus we rely on gradient descent on curves. The energy we consider involves optimization with partial differential equation (PDE) constraints, and thus we build on optimization methods from [3, 11].

Coarse-to-fine methods, where coarse representations of the image or objective function are processed and then finer aspects of the data are successively revealed, have a long history in computer vision [5]. In these methods, data or the objective function is smoothed, and the smoothed problem is solved. The result is used to initialize the problem with less smoothing, where finer details of the data are revealed. The hope is that this finer result retains aspects of coarse solution, while gradually finding finer detail. However, without additional heuristics such as restricting the finer solution to be around the solution of the coarse problem, there is no guarantee that coarse structure is preserved when solving the finer problem. Recently, [33] provided analysis and derived closed form solutions for the smoothing of the objective in problems of point cloud matching. Our method uses a single energy integrating over a continuum of scales in *parallel*, rather than a sequential approach where multiple energies from coarse to fine are solved. This guarantees that the coarse and fine scale aspects of the desired solution are obtained.

Since we also apply our method to the problem of segmenting moving objects in video based on motion, we highlight some aspects of that literature most relevant to this work. Methods for motion segmentation are based on optical flow (e.g., [45]). Piecewise parametric models for motion of regions in segmentation are used in e.g., [52, 10]. Non-parametric warps are used for motion models (e.g.,[37, 46, 54]). Our goal here is *not* to estimate motion, but rather we use existing techniques for motion estimation, and improve the segmentation of regions by merely replacing a single scale formulation with our novel continuum scale space approach.

2. Continuum Shape-Tailored Energy

In this section, we construct a coarse-scale preferential energy without blurring across segments. To achieve this, we introduce a Shape-Tailored *Continuum* Scale Space. A Shape-Tailored Scale Space avoids blurring across regions, and a continuum of scales obtains a coarse-to-fine property.

2.1. Shape-Tailored Heat Scale Space

The Gaussian Scale Space, constructed by smoothing the image with a Gaussian at a continuum of scales (variances), can be generalized to be defined within regions (subsets of the image) of arbitrary shape by using the heat equation (see Figure 2). The solution to the heat equation defaults to Gaussian smoothing when the domain is $\mathbb{R}^2$. The heat equation, defined in a region R, is:

$$\begin{cases} \partial_t u(t,x) = \Delta u(t,x) & x \in R,\, t > 0 \\ \nabla u(t,x) \cdot N = 0 & x \in \partial R,\, t > 0 \\ u(0,x) = I(x) & x \in R \end{cases} \tag{1}$$

where $u : [0, +\infty) \times R \to \mathbb{R}^k$ denotes the scale space, $R \subset \Omega \subset \mathbb{R}^2$ is the domain (or subset) of the image Ω, $I : \Omega \to \mathbb{R}^k$ ($k \geq 1$ is the number of channels) is the image, ∂R denotes the boundary of R, N is the unit outward normal

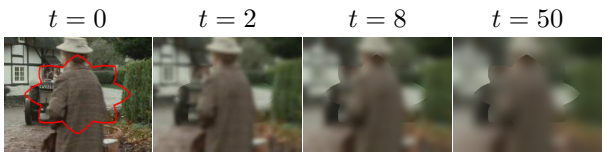

Figure 2. Shape-tailored scale space (solution of heat equation within regions with boundary in red) for various times (scales). Notice the quick diffusion of fine scale structures, and the persistence of coarse structure. The persistence of coarse structure is important to our coarse-to-fine segmentation scheme.

vector to R, ∇ denotes the vector of partials, Δ denotes the Laplacian, ∂_t denotes the partial derivative with respect to t, and t is the scale parameter parameterizing the scale space. Increasing t indicates increasing amount of smoothing.

The construction of scale space using the heat equation is useful for segmentation as it allows us to conveniently compute coarse scales of the data *within* regions of a segmentation. If the regions are chosen to be the correct segmentation, this avoids blurring data across segmentation boundaries. However, one does not know the segmentation a-priori, and thus the regions are simultaneously optimized with the scale spaces in the optimization problem defined next.

2.2. Coarse-Scale Preferential Energy

The Gaussian scale space is relevant in defining our coarse-scale preferential energy as the heat equation removes the fine structure of the image in short time, and spends more time removing coarse structure (see Figure 2) [9]. Therefore, a data term integrating the scale space over the scale parameter of the heat equation gives preference to segmentations separating the coarse over the fine structure. We thus propose the following energy for segmentation integrating over a continuum of scales:

$$E = \sum_{i=1}^{N} \int_{R_i} \int_0^T |u_i(t,x) - a_i|^2 w(t)\,\mathrm{d}t\,\mathrm{d}x + \mathrm{Reg}(\partial R_i), \tag{2}$$

where $T > 0$ is the final time, $\{R_i\}_{i=1}^N$ are a collection of regions forming the segmentation, $a_i \in \mathbb{R}^k$ is the average of $u_i(t,\cdot)$, and $w : \mathbb{R}^+ \to \mathbb{R}$ is a function that weights each scale. It can be shown that a_i is independent of t. This energy is the mean-squared error of the image within the region across all scales. It generalizes common single scale segmentation models, including piecewise constant Mumford-Shah (Chan-Vese [51, 34]). Reg denotes usual curve regularization that will be discussed in the implementation section, Section 3.3.

To further demonstrate the coarse preference of our energy, we write the data term of the energy in Fourier domain. For simplicity, we choose $w(t) = 1$; other weights lead to a similar conclusion. Choosing the whole domain as a region, the data term can be written in Fourier domain as:

Lemma 1 *Suppose* $I : \mathbb{R}^2 \to \mathbb{R}$ *and* $a = \int_{\mathbb{R}^2} I(x)\,\mathrm{d}x = \int_{\mathbb{R}^2} u(t,x)\,\mathrm{d}x$. *Then*

$$\int_0^\infty \int_{\mathbb{R}^2} |u(t,x) - a|^2\,\mathrm{d}x\,\mathrm{d}t = \int_{\mathbb{R}^2} |H(\omega)\hat{I}(\omega)|^2\,\mathrm{d}\omega, \tag{3}$$

where $H(\omega) = \frac{1}{\sqrt{2|\omega|}}$, $\hat{I}$ *denotes the Fourier transform, and* ω *denotes frequency.*

The proof can be found in supplementary materials. The function H decays the high frequency components of I at a linear rate, thus the energy gives preference to the coarse image structure. Without integrating over the scale space, the energy in Fourier domain would result in $H = 1$, which has equal preference to coarse and fine structure.

3. Optimization and Scale Weighting

We now derive the optimization scheme for the energy (2), and propose and analyze weight choices.

3.1. Constrained Optimization Problem

The energy (2) is optimized with respect to the regions. Since the integrand of the energy depends on the regions nonlinearly, as the heat equation has a non-linear dependence on the region, the energy is not convex, and thus we apply gradient descent. In order to compute the gradient, we formulate the energy minimization as a constrained optimization problem. That is, we treat the minimization of the energy (2) as defined on both the regions R_i *and* u_i with the constraint that u_i satisfies the heat equation (1). This formulation allows us to apply the technique of Lagrange multipliers, which makes computations simpler since the nonlinear dependence of u_i on R_i is decoupled.

Since all data terms of the energy in (2) have the same form, we focus on computing the gradient for any one term. For convenience in notation, we avoid the subscript i denoting the index of the region. Using Lagrange multipliers, we formulate the energy as a function of region R, u, and the Lagrange multiplier $\lambda : [0,T] \times R \to \mathbb{R}^k$ with the constraint that u satisfies the heat equation:

$$E(R,u,\lambda) = \int_0^T \int_R f(u)\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \int_R (\nabla\lambda \cdot \nabla u + \lambda \partial_t u)\,\mathrm{d}x\,\mathrm{d}t, \tag{4}$$

where $f(t,u) = (u-a)^2 w(t)$. We have excluded the dependencies on x, t for convenience of notation. We have also provided a more general form of the squared error with a general function f of u. The second term comes from the weak form of the heat equation. Integrating by parts to

move the gradient from λ to ∇u gives the classical form of the heat equation in (1). Therefore, the second term in (4) is indeed obtained by Lagrange multipliers.

We may now compute the gradient for E (4) by deriving the optimizing conditions in u and λ. Details are found in supplementary materials. Optimizing in λ simply results in the original heat equation constraint, so we compute the optimizing condition for u by computing the derivative (variation) of E with respect to u. This results in a solution for λ as given below:

Lemma 2 (PDE for Lagrange Multiplier λ) *The Lagrange multiplier λ satisfies the following heat equation with forcing term, evolving backwards in time:*

$$\begin{cases} \partial_t \lambda(t,x) + \Delta\lambda(t,x) = f_u(t,u(t,x)) & x \in R \times [0,T] \\ \nabla\lambda(t,x) \cdot N = 0 & x \in \partial R \times [0,T] \\ \lambda(T,x) = 0 & x \in R \end{cases}, \tag{5}$$

where f_u denotes the partial with respect to the second argument.

Duhamel's Principle [12] leads to the following solution:

Lemma 3 (Lagrange Multiplier λ) *The solution of (5) can be written as*

$$\lambda(t,x) = -\int_t^T F(s-t,x;s)\,\mathrm{d}s. \tag{6}$$

where $F(\cdot,\cdot;s) : [0,T] \times R \to \mathbb{R}$ is the solution of the forward heat equation (1) with zero forcing and initial condition $f_u(u)$ evaluated at time s, i.e.,

$$\begin{cases} \partial_t F(t,x;s) - \Delta F(t,x;s) = 0 & x \in R \times [0,T] \\ \nabla F(t,x;s) \cdot N = 0 & x \in \partial R \times [0,T] \\ F(0,x;s) = f_u(s,u(s,x)) & x \in R \end{cases}. \tag{7}$$

In the case that $f(t,u) = (u-a)^2 w(t)$, λ can be expressed as

$$\lambda(t,x) = -2\int_t^T (u(2s-t,x) - a) w(s)\,\mathrm{d}s. \tag{8}$$

The formula for λ in (8) is convenient for particular choices of the weight w as taking the limit as T gets large leads to the energy gradient being computable without explicitly computing the scale space u, as shown in the next section.

With the optimizing conditions for u and λ of E, we can now compute the gradient of the energy E with respect to R in terms of λ and u:

Proposition 1 *The gradient of E with respect to the boundary ∂R can be expressed as*

$$\nabla_{\partial R} E = \int_0^T \left[f(u) + \nabla\lambda \cdot \nabla u + \lambda \partial_t u \right] \mathrm{d}t \cdot N, \tag{9}$$

where N is the normal vector to ∂R.

3.2. Weighting Functions

We now explore possible choices of weights, w. Some choices of weights may have convenient solutions for the gradient that does not require computation of the scale-space u, which makes the computational cost much less expensive than the generic formula (9). As observed in the experiments, all have a coarse-to-fine behavior, but each differs in the extent of this property. Calculations are provided in supplementary materials.

Exponential With Positive Exponent (ExpPos): We consider the weight $w(t) = e^{1/\alpha[(t/T)^2-1]}\mathbf{1}_{[0,T]}(t)$, where $\alpha > 0$ and $\mathbf{1}$ denotes the indicator function. Here, the weight increases with scale so that the largest scales between 0 and T are weighted the most. We truncate at a finite T. This is because for large scales, the image is blurred too much to be used in segmentation, and very large scales should have either low or zero weight. This weighting exhibits the most coarse-to-fine behavior of any weightings we consider. Although this is the ideal weighting, to the best of our knowledge, the gradient (9) cannot be written in a form that does not require computation of the scale space. Thus, it is computationally more costly than other weightings we consider. However, typically T is chosen small (e.g., $T = 10$ for a 256×256 image) in comparison to other weightings, which offers cost savings.

Truncated Uniform Weight (Uniform): We consider the weight function $w(t) = \mathbf{1}_{[0,T]}(t)$. This uses a uniform weight on all scales between 0 and T. Since we want to avoid very large scales ($T \to \infty$), we choose a finite T. The gradient when T is large (but still finite) is approximated as

$$\nabla_{\partial R} E \cdot N \approx (u_0 + a)(aT - U_T) + \frac{1}{2}|\nabla U_T|^2, \tag{10}$$

where u_0 is initial condition to the heat equation (original data), and

$$\begin{cases} U_T(x) - T\Delta U_T(x) = T u_0(x) & x \in R \\ \nabla U_T \cdot N = 0 & x \in \partial R \end{cases}. \tag{11}$$

U_T is the integral of the scale space from 0 to T and this can be approximated as the solution of (11) (see supplementary). The advantage of (10) is that it does not require explicit computation of the scale space, and (11) can be solved efficiently iteratively. Indeed, in gradient descent of R, the solution for the previous iteration can be used as a warm start for the next iteration. Analysis of the approximation is in supplementary.

Exponential With Negative Exponent (ExpNeg): We consider the weight $w(t) = e^{-(1/\alpha)t}$ for all $t \in [0,\infty)$, where $\alpha > 0$. A small value of α implies that only the small scales are relevant. A large value of α includes larger scales, which is desired. The intuition for using this weighting is that it includes moderately large scales with non-negligible

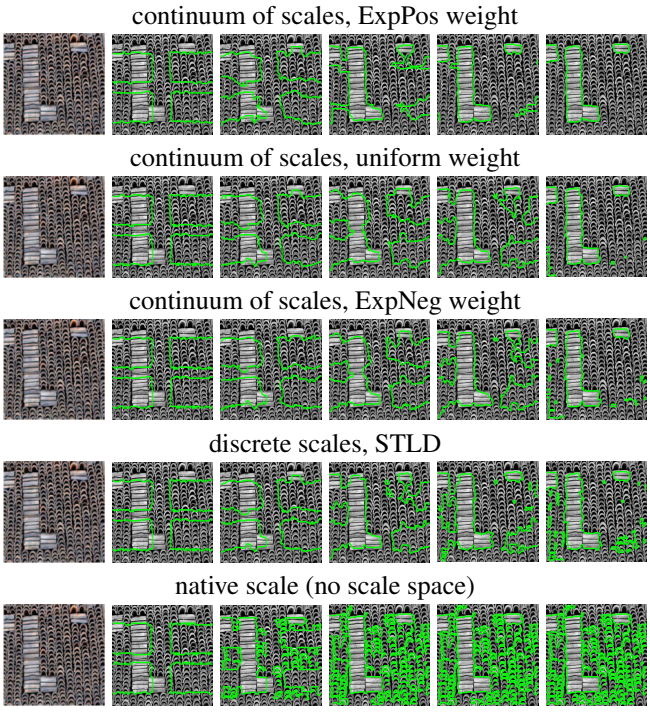

Figure 3. **Visualization of Energy Optimization for Various Scale Weightings**. *We compare usual segmentation of the native image scale, a discrete shape-tailored scale space (STLD), ExpPos, Uniform, and ExpNeg weightings for the continuum scale space. No coarse-to-fine behavior is exhibited for the native image scale and STLD. The continuum scale spaces give coarse-to-fine behavior, with ExpPos more so than other weightings.*

weight as desired, it disregards very large scales as desired by having exponentially decaying weight, and it has an exact solution for the gradient that does not require the computation of scale space. One can show that the gradient is

$$\nabla_{\partial R} E \cdot N = a\alpha(a+2u_0) - u_0 U_{2\alpha} + \frac{1}{4\alpha} U_{2\alpha}^2 - \frac{1}{2}|\nabla U_{2\alpha}|^2, \tag{12}$$

where $U_{2\alpha}$ solves (11) with T replaced by 2α. Like the uniform weighting, the gradient yields a form that does not require the computation of the scale space. An advantage over the uniform case is that the solution is exact.

3.3. Multi-Region Segmentation

We now present the numerical implementation of the gradient descent for energy (2), when there are multiple regions. The term involving regularization is discussed later. Let $G_i N_i$ be the gradient of the i^{th} summand of E in (2), where N_i is the outward normal to R_i. For instance, $G_i N_i$ can be any one of the expressions (9), (10), (12). As shown in [56], the gradient of the full energy evaluated at a point x is just the sum of $G_i N_i$ for all i such that $x \in \partial R_i$. For a point $x \in \partial R_i \cap \partial R_j$, this yields that the gradient is $(G_i - G_j)N_i$.

To achieve sub-pixel accuracy, we use relaxed indicator functions $\phi_i : \Omega \to [0,1]$ for $i = 1, \ldots, N$ to represent the regions, similar to level set methods [38]. R_i is where ϕ_i is larger than $\phi_j, j \neq i$. By abuse of notation, denote by G_i the quantity multiplying the normal vector of region R_i in either of (9), (10), (12), which is defined in the entire region R_i. We extend it from R_i to $D(R_i)$, a small dilation of R_i, by solving for G_i in $D(R_i)$. The extension beyond the region is done so that the evolution of ϕ_i can be defined around the curve, as in level set methods. Following [38] to convert a curve to a level set evolution, the update scheme for ϕ_i inducing the regions gradient descent is Algorithm 1.

Algorithm 1 Multi-Region Gradient Descent

1: Input: An initialization of ϕ_i
2: **repeat**
3: Set regions: $R_i = \{x \in \Omega : i = \text{argmax}_j \phi_j(x)\}$
4: Compute dilations, $D(R_i)$, of R_i
5: Compute band pixels $B_i = D(R_i) \cap D(\Omega \backslash R_i)$
6: Compute G_i in B_i from (9), (10), or (12)
7: Update pixels $x \in D(R_i) \cap D(R_j)$ as follows:

$$\phi_i^{\tau+\Delta\tau}(x) = \phi_i^{\tau}(x) - \Delta\tau(G_i(x) - G_j(x))|\nabla\phi_i^{\tau}(x)| + \Delta\tau \cdot \varepsilon\Delta\phi_i^{\tau}(x).$$

8: Update all other pixels as

$$\phi_i^{\tau+\Delta\tau}(x) = \phi_i^{\tau}(x) + \Delta\tau \cdot \varepsilon\Delta\phi_i^{\tau}(x).$$

9: Clip between 0 and 1: $\phi_i = \max\{0, \min\{1, \phi_i\}\}$.
10: **until** regions have converged

The update of the ϕ_i in Line 7 of Algorithm 1 involves the term $\Delta\phi_i^{\tau}$, which provides smoothness of the curve. More sophisticated regularizers (such as length regularization) may be used, but we have found this simple regularization sufficient. We choose $\varepsilon = 0.005$ in experiments, and this does not need to be tuned, as it is mainly for inducing regularity for computation of derivatives of ϕ. Further, considering the scale space naturally induces regularity.

4. Application to Motion Segmentation

In this section, we show how the results of the previous section can be applied to motion segmentation. Motion segmentation is the problem of segmenting objects and/or regions with similar motions computed using multiple images of the object(s). One of the challenges of motion segmentation is that motion is inferred through a sparse set of measurements (e.g., along image edges or corners), and thus the motion signal is typically only reliable for segmentation in sparse locations. By using a scale space formulation of an energy for motion segmentation, coarse representations

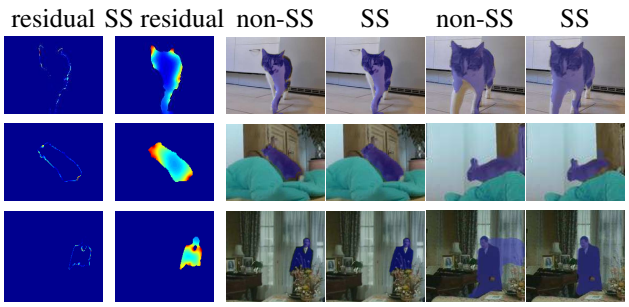

Figure 4. Motion residuals at a single scale are sparse (left column), leading to difficulties in using these cues in segmentation (non-SS). Motion cues at a continuum of scales (SS) provide a richer signal (2nd column), which improves segmentation. Segmentations (in purple) are shown for a frame (middle two) and a few frames ahead (right two). Although errors in the non-SS approach are subtle between frames, they quickly propagate across frames, compared to our approach.

of the motion signal are integrated and more significantly impact the segmentation. This property increases the reliability of motion segmentation (Figure 4), and the coarse-to-fine approach captures the coarse-structure without being impacted by fine-scale distractions at the outset.

With this motivation, we reformulate the motion segmentation problem with scale space. Let $I_0, I_1 : \Omega \to \mathbb{R}^k$ be two images of a sequence where Ω is the domain of the image. For a given region R_i, we define a mapping $w_i : R_i \to \Omega \subset \mathbb{R}^2$, which we call a warp or deformation that back warps I_1 to I_0. We assume that I_0 and I_1 are related through w_i by the Brightness Constancy Assumption, except for occlusions, as in typical works in the optical flow [45]. Define the energy

$$E_{mseg} = \sum_{i=1}^{N} \int_{R_i} \int_0^T [1 - m(x)]|u_i(t,x)|^2 w(t)\,\mathrm{d}t\,\mathrm{d}x - \int_{R_i} m(x) \log p_{R_i}(I_0(x))\,\mathrm{d}x + \mathrm{Reg}(\partial R_i), \quad (13)$$

where u_i is the scale space of the difference of I_0 and the back-warping of I_1 in the un-occluded region $R_i \backslash O_i$:

$$u_{0,i} = \begin{cases} I_1(w_i(x)) - I_0(x) & x \in R_i \backslash O_i \\ 0 & x \in O_i \end{cases}, \quad (14)$$

and $m : \Omega \to [0,1]$ is the motion ambiguity function. Note that the energy in the case $m = 0$ is equivalent to integrating over all scales the difference of the scale spaces of I_0 and of $\hat{I}_1$ (defined as $I_1 \circ w_i$ inside $R_i \backslash O_i$ and I_0 in O_i). Note that $\hat{I}_1$ is used rather than $I_1 \circ w_i$ as the latter does not correspond to I_0 in the occlusion. This energy requires that the regions are chosen so that *all* scales of the images between 0 and T match. The motion ambiguity function m indicates whether the motion at a pixel is reliable for segmentation (1 in a textureless or occluded region and 0 otherwise). In case the motion is ambiguous, local color histograms p_{R_i} within regions are used for grouping. As is typical in optical flow [45], we set the occlusion to be a threshold of the residual: $O_i = \{x \in R_i : |I_1(w_i(x)) - I_0(x)|^2 > \beta\}$.

The optimization involves iterative alternating updates of the warps and the regions. To update warps, we use the method of warp estimation in [55]. To update the regions, we use the results of the previous section and use the exponential weight with negative exponent, for computational efficiency. This yields the gradient of the i^{th} data terms in (13) approximately as

$$\left[(1-m)(\frac{\alpha}{4}U^2 - u_0 U - \frac{1}{2}|\nabla U|^2) - m \log p_{R_i}(I_0)\right] N_i, \quad (15)$$

where U is the solution of (11) using $T = 2\alpha$ and right hand side $u_{0,i}$. The gradient descent of E_{mseg} is then given by Algorithm 1, choosing G_i to be the component of (15) multiplying N_i. We apply our method frame-by-frame. Then we propagate the result to the next frame via the computed warp to warm-start the segmentation in the next frame.

5. Experiments

5.1. Texture Segmentation

Datasets and Methods Compared: We first test our method on texture segmentation, a task where multiscale information is important. We test on two datasets used in [23]. The Brodatz Synthetic Dataset has 198 images generated from textures in Brodatz and random shapes from MPEG dataset. The second is the Real-World Texture Dataset, which consists of 256 textured images obtained from photographs of real-world scenes. We use RGB color channels and binned oriented gradients at four angles, as the features for segmentation. Since the contribution in this paper is the use of shape-tailored scale spaces at a continuum of scales, we compare to [23] (STLD), which uses scale space but only considers a discrete number of scales. For reference, we include other segmentation methods. We use the abbreviations ExpPos, Uniform, and ExpNeg for the positive exponent exponential, uniform, and negative exponent exponential weights in our method. The methods are all initialized with a standard box tessellation.

Results on Brodatz: First, we compare on Brodatz with different weighting schemes introduced in Section 3.2 for continuum scale spaces against STLD. To compare weightings and not the quality of various approximations, we use (9) to compute the gradient. Images are 128×128 and we choose $\alpha = T = 10$ (corresponding to the max scale used in STLD) for all weightings. Results are displayed in Table 1. All weightings give similar results, and all are significantly more accurate than STLD. This indicates that using

Brodatz Synthetic Dataset

	Contour		Region metrics					
	F-meas.		GT-cov.		Rand. Index		Var. Info.	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
ExpPos (ours)	**0.41**	**0.41**	**0.80**	**0.80**	**0.79**	**0.79**	**0.68**	**0.68**
ExpNeg (ours)	0.39	0.39	0.78	0.78	0.77	0.77	0.68	0.68
Uniform (ours)	0.40	0.40	0.79	0.79	0.78	0.78	0.68	0.68
STLD	0.33	0.33	0.71	0.71	0.70	0.70	0.74	0.74

Real-World Texture Dataset

	Contour		Region metrics					
	F-meas.		GT-cov.		Rand. Index		Var. Info.	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
ExpNeg (ours)	**0.60**	**0.60**	**0.91**	**0.91**	**0.91**	**0.91**	**0.45**	**0.45**
STLD	0.58	0.58	0.87	0.87	0.87	0.87	0.59	0.59
non-STLD	0.17	0.17	0.81	0.81	0.82	0.82	0.77	0.77
mcg [2]	0.51	0.54	0.74	0.82	0.77	0.85	0.80	0.66
gPb [1]	0.50	0.54	0.74	c0.84	0.78	0.86	0.80	0.65
CB [21]	0.48	0.52	0.64	0.70	0.66	0.75	0.89	0.78
SIFT	0.10	0.10	0.55	0.55	0.59	0.59	1.44	1.44
Entropy [19]	0.08	0.08	0.74	0.74	0.75	0.75	0.95	0.95
Hist-5 [36]	0.14	0.14	0.66	0.66	0.70	0.70	1.18	1.18
Hist-10 [36]	0.13	0.13	0.66	0.66	0.70	0.70	1.19	1.19
Chan-Vese [8]	0.14	0.14	0.71	0.71	0.73	0.73	1.04	1.04
LAC [28]	0.09	0.09	0.55	0.55	0.58	0.58	1.41	1.41
Global Hist [32]	0.12	0.12	0.65	0.65	0.67	0.67	1.12	1.12

Table 1. **Results on Texture Segmentation Datasets**. Algorithms are evaluated using contour and region metrics. Higher F-measure for the contour metric, ground truth covering (GT-cov), and rand index indicate better fit to the ground truth, and lower variation of information (Var. Info) indicates a better fit to ground truth.

continuum scale space leads to increased performance.

Results on Real-World Texture Images: Since all results for different weightings are similar, we now use ExpNeg for comparison on the Real-World Texture Dataset because of its speed. Results, in Table 1, for $\alpha = 20$, show that the accuracy of the continuum scale space is greater than discrete scales (STLD). Sample representative visual results are shown in Figure 5.

Next, we test our approach with different choices of α using the ExpNeg weighting. We also compare against STLD in terms of speed and accuracy. Results are shown in Table 2. Results of STLD show that more than one scale is necessary, and faster speed by using fewer scales leads to degradation of the segmentation. Second, results of ExpNeg show that the results are stable across different parameter choices for α. Finally, a speed comparison is performed between ExpNeg and STLD. Note that each scale that is used in STLD requires the solution of a PDE, whereas our approach of ExpNeg requires only a single PDE. This makes our continuum scale space approach computationally less expensive, as confirmed in Table 2. Our approach also requires only a single parameter in contrast to STLD that requires choosing a list of scales.

5.2. Motion Segmentation

Datasets: We test our method on the Freiburg-Berkeley Motion Segmentation (FBMS-59) [37] dataset. FBMS-59 consists of two sets - training, 29 sequences, and test, 30 sequences. Videos range between 19 and 800 frames, and

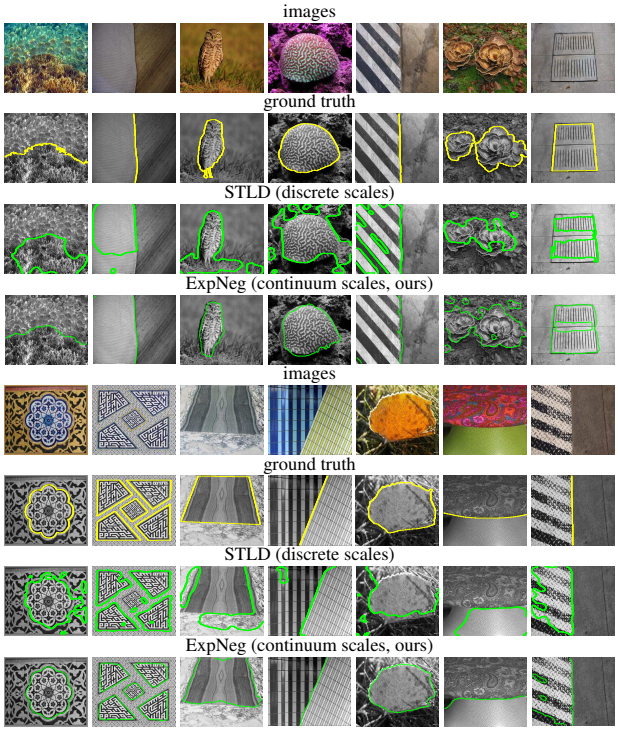

Figure 5. **Sample representative results on Real-World Texture Dataset**. *We compare the best two methods (ours) and STLD (using discrete scale spaces).*

STLD Scale Comparison

	Contour		Region metrics					
	F-meas.		GT-cov.		Rand. Index		Var. Info.	
STLD scales	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
4	0.56	0.56	0.85	0.85	0.85	0.85	0.63	0.63
20	0.55	0.55	0.84	0.84	0.84	0.84	0.64	0.64
4,8,12,16,20	**0.58**	**0.58**	**0.87**	**0.87**	**0.87**	**0.87**	**0.59**	**0.59**

ExpNeg Parameter α Comparison

	Contour		Region metrics					
	F-meas.		GT-cov.		Rand. Index		Var. Info.	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
$\alpha = 20$	0.60	0.60	0.91	0.91	0.91	0.91	0.45	0.45
$\alpha = 30$	0.60	0.60	0.90	0.90	0.90	0.90	0.46	0.46
$\alpha = 50$	0.60	0.60	0.90	0.90	0.90	0.90	0.46	0.46

Speed Comparison

method	average iterations	average time
ExpNeg ($\alpha = 20$)	12.9 ± 4.4	10.3 sec
STLD (scale 4,8,12,16,20)	16 ± 4.1	83.7 sec

Table 2. **Analysis of Scale Parameters and Speed**. [Top]: Comparison of different scale choices for discrete scale spaces (STLD). [Middle]: Results for different α in continuum scale space with ExpNeg weight. [Bottom]: Speed comparison on a single processor for ExpNeg continuum scale space and STLD.

have multiple objects.

Comparison: To demonstrate the advantage of our continuum space energy over a corresponding single scale energy, we compare to [55]. Our approach replaces the single scale motion term there with the energy (13). Further, additional regularization used in [55] is not used, as the

	Training set (29 sequences)				Test set (30 sequences)			
	P	R	F	$N/65$	P	R	F	$N/69$
[16]	79.17	47.55	59.42	4	77.11	42.99	55.20	5
[37]	81.50	63.23	71.21	16	74.91	60.14	66.72	20
[48]	83.00	70.10	76.01	23	77.94	59.14	67.25	15
[22]	86.91	71.33	78.35	25	87.57	**70.19**	**77.92**	25
[55]	89.53	70.74	79.03	26	91.47	64.75	75.82	27
ExpNeg (ours)	**93.04**	**72.68**	**81.61**	**29**	**95.94**	65.54	77.87	**28**

Table 3. FBMS-59 results. Average precision (P), recall (R), F-measure (F), and number of objects detected (N) over all sequences in training and test datasets. Higher values indicate superior performance. All methods are fully automatic.

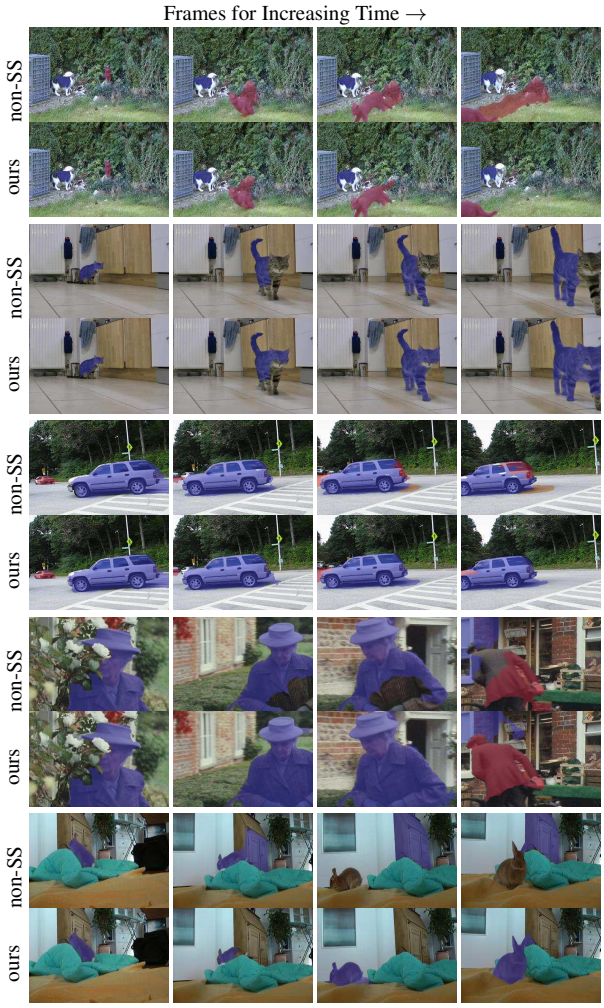

Figure 6. Sample visual results on representative sequences for the FBMS-59 dataset (segmented objects in purple and red). The change of energy to integrate over all scales (our approach) is generally less sensitive to clutter than using an energy that contains only one scale (non-SS).

scale-space provides inherent regularization. Since we test on benchmarks, we also compare to other state-of-the-art approaches, although our main purpose is to show the improvements that occur by merely using our continuum scale space energy.

Initialization: We initialize each with a segmentation of optical flow from [45] between frame 1 and 20.

Parameters: Our method with ExpNeg weighting requires one parameter α in (12). We choose it to be $\alpha = 20$ by selecting it based on a few sequences from the training set. Other parameters e.g., histogram sizes are chosen based on [55].

Results on FBMS-59: Figure 6 shows some representative visual results of our method and the single scale approach. Table 3 shows quantitative results of the two approaches, as well as other state-of-the-art methods. Visual results show our approach generally avoids distracting clutter and thus prevents leakages in comparison to the single scale approach. In many cases, it also captures more of the object. Quantitative results show that we improve the F-measure of [55] by about 2% on both training and test sets, and that we increase the number of objects detected. We also have highest F-measure of all competing methods.

Computational cost: The additional processing cost required for our scale space is small compared with the overall cost of [55]. Our approach adds about 5 secs per frame (one core) to the total time on average of about 30 secs per frame by [55] on a 12-core processor.

6. Conclusion

We have presented a general energy that reformulates conventional data terms in segmentation problems. This novel energy incorporates a shape-tailored *continuum* scale space. It exhibits two important properties: scales spaces are defined within regions, so that structures in different segments are not blurred across boundaries nor displaced, and a coarse-to-fine property. The latter favors that the coarse structure of the desired segmentation is obtained while finer structure becomes successively obtained, without having to rely on heuristics. Our shape-tailored continuum scale spaces have two main advantages over shape-tailored discrete scale spaces: they have a coarse-to-fine property, ignoring distracting fine-scale structure leading to more accurate solutions, and they have a speed advantage. We have shown application to both texture and motion segmentation. Experiments on two benchmark datasets in texture segmentation have shown the importance of shape-tailored continuum scale spaces with respect to existing state-of-the-art. Experiments on a motion segmentation benchmark have shown the importance of multiscale information in motion segmentation: a mere integration of the common motion residual over scale improves results, leading to a state-of-the-art method.

Acknowledgements

Partially funded by KAUST OCRF-2014-CRG3-62140401, NRF-2014R1A2A1A11051941, and NSF CCF-1526848.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011. 7

[2] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 328–335, 2014. 1, 2, 7

[3] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson. Image segmentation using active contours: Calculus of variations or shape gradients? *SIAM Journal on Applied Mathematics*, 63(6):2128–2154, 2003. 2

[4] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decompositionmodeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006. 2

[5] A. Blake and A. Zisserman. *Visual reconstruction*, volume 2. MIT press Cambridge, 1987. 1, 2

[6] X. Bresson, P. Vandergheynst, and J.-P. Thiran. Multiscale active contours. *International Journal of Computer Vision*, 70(3):197–211, 2006. 1, 2

[7] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1704–1711. IEEE, 2010. 2

[8] T. F. Chan and L. A. Vese. Active contours without edges. *Image processing, IEEE transactions on*, 10(2):266–277, 2001. 7

[9] C. Chen and H. Edelsbrunner. Diffusion runs low on persistence fast. In *2011 International Conference on Computer Vision*, pages 423–430. IEEE, 2011. 3

[10] D. Cremers and S. Soatto. Motion competition: A variational approach to piecewise parametric motion segmentation. *International Journal of Computer Vision*, 62(3):249–265, 2005. 2

[11] M. C. Delfour and J.-P. Zolésio. *Shapes and geometries: metrics, analysis, differential calculus, and optimization*, volume 22. Siam, 2011. 2

[12] L. C. Evans. Partial differential equations. 2010. 4

[13] L. Florack and A. Kuijper. The topological structure of scale-space images. *Journal of Mathematical Imaging and Vision*, 12(1):65–79, 2000. 2

[14] M. Galun, E. Sharon, R. Basri, and A. Brandt. Texture segmentation by multiscale aggregation of filter responses and shape elements. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 716–723. IEEE, 2003. 2

[15] J.-M. Geusebroek, R. Van Den Boomgaard, A. W. Smeulders, and A. Dev. Color and scale: The spatial structure of color images. In *Computer Vision-ECCV 2000*, pages 331–341. Springer, 2000. 1

[16] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2141–2148. IEEE, 2010. 8

[17] T. Hassner, V. Mayzels, and L. Zelnik-Manor. On sifts and their scales. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1522–1528. IEEE, 2012. 2

[18] J. Hegdé. Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, 84(4):405–439, 2008. 1

[19] B.-W. Hong, S. Soatto, K. Ni, and T. Chan. The scale of a texture and its application to segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 7

[20] R. Hummel and R. Moniot. Reconstructions from zero crossings in scale space. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12):2111–2130, 1989. 2

[21] P. Isola, D. Zoran, D. Krishnan, and E. H. Adelson. Crisp boundary detection using pointwise mutual information. In *Computer Vision–ECCV 2014*, pages 799–814. Springer, 2014. 7

[22] M. Keuper, B. Andres, and T. Brox. Motion trajectory segmentation via minimum cost multicuts. *IEEE International Conference on Computer Vision (ICCV)*, pages 3271–3279. 8

[23] N. Khan, M. Algarni, A. Yezzi, and G. Sundaramoorthi. Shape-tailored local descriptors and their application to segmentation and tracking. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 3890–3899. IEEE, 2015. 2, 6

[24] J. J. Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984. 1

[25] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):142–157, 2009. 1, 2

[26] N. Komodakis, M. P. Kumar, and N. Paragios. (hyper)-graphs inference through convex relaxations and move making algorithms: Contributions and applications in artificial vision. *Foundations and Trends® in Computer Graphics and Vision*, 10(1):1–102, 2016. 2

[27] G. Koutaki and K. Uchimura. Scale-space processing using polynomial representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2744–2751, 2014. 1

[28] S. Lankton and A. Tannenbaum. Localizing region-based active contours. *Image Processing, IEEE Transactions on*, 17(11):2029–2039, 2008. 7

[29] T. Lindeberg. Scale-space for discrete signals. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(3):234–254, 1990. 2

[30] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 2

[31] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981. 2

[32] O. Michailovich, Y. Rathi, and A. Tannenbaum. Image segmentation using active contours driven by the bhattacharyya

gradient flow. *Image Processing, IEEE Transactions on*, 16(11):2787–2801, 2007. 7

[33] H. Mobahi and J. W. Fisher III. Coarse-to-fine minimization of some common nonconvexities. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 71–84, 2015. 1, 2

[34] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989. 3

[35] P. Neri. Coarse to fine dynamics of monocular and binocular processing in human pattern vision. *Proceedings of the National Academy of Sciences*, 108(26):10726–10731, 2011. 1

[36] K. Ni, X. Bresson, T. Chan, and S. Esedoglu. Local histogram based segmentation using the wasserstein distance. *International Journal of Computer Vision*, 84(1):97–111, 2009. 7

[37] P. Ochs, J. Malik, and T. Brox. Segmentation of moving objects by long term video analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(6):1187–1200, 2014. 2, 7, 8

[38] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988. 2, 5

[39] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002. 2

[40] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990. 1, 2

[41] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1133–1140. IEEE, 2009. 2

[42] J. C. Rubio, J. Serrat, A. López, and N. Paragios. Unsupervised co-segmentation through region matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 749–756. IEEE, 2012. 2

[43] G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International journal of computer vision*, 11(1):25–44, 1993. 2

[44] A. Sironi, V. Lepetit, and P. Fua. Multiscale centerline detection by learning a scale-space distance transform. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2697–2704. IEEE, 2014. 2

[45] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2432–2439. IEEE, 2010. 2, 6, 8

[46] D. Sun, J. Wulff, E. Sudderth, H. Pfister, and M. Black. A fully-connected layered model of foreground and background flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2451–2458, 2013. 2

[47] G. Sundaramoorthi and B.-W. Hong. Fast label: Easy and efficient solution of joint multi-label and estimation problems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3126–3133, 2014. 2

[48] B. Taylor, V. Karasev, and S. Soatto. Causal video object segmentation from persistence of occlusions. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 4268–4276. IEEE, 2015. 8

[49] B. Ummenhofer and T. Brox. Global, dense multiscale reconstruction for a billion points. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1341–1349, 2015. 2

[50] R. Van Den Boomgaard and A. Smeulders. The morphological structure of images: The differential equations of morphological scale-space. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(11):1101–1113, 1994. 2

[51] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision*, 50(3):271–293, 2002. 3

[52] J. Y. Wang and E. H. Adelson. Representing moving images with layers. *Image Processing, IEEE Transactions on*, 3(5):625–638, 1994. 2

[53] A. P. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, volume 9, pages 150–153. IEEE, 1984. 1

[54] Y. Yang and G. Sundaramoorthi. Shape tracking with occlusions via coarse-to-fine region-based sobolev descent. *IEEE transactions on pattern analysis and machine intelligence*, 37(5):1053–1066, 2015. 2

[55] Y. Yang, G. Sundaramoorthi, and S. Soatto. Self-occlusions and disocclusions in causal video object segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4408–4416, 2015. 6, 7, 8

[56] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(9):884–900, 1996. 5