Linearized Multidimensional Earth-Mover’s-Distance Gradient Flows

Carlos S. Mendoza*, José-Antonio Pérez-Carrasco, Aurora Sáez, Student, IEEE,
Begoña Acha, Member, IEEE and Carmen Serrano, Member, IEEE

Abstract—This article presents the first framework capable of performing active contour segmentation using Earth Mover’s Distance (EMD) to measure dissimilarity between multidimensional feature distributions. EMD is the best known and best understood cross-bin histogram distance measure, and as such it allows for meaningful comparisons between distributions, unlike bin-to-bin measures which only account for discrepancies on a bin-to-bin basis. Since EMD is obtained with linear programming techniques, its differential structure with respect to variations in bin weights as the active contour evolves is expressed by means of sensitivity analysis. Euler-Lagrange equations are then derived from the computed sensitivity at every iteration to produce gradient descent flows. We validate our approach with color image segmentation, in comparison with state-of-the-art Bhattacharyya (bin-to-bin) and one-dimensional EMD (cross-bin) active contours. Some unique advantages of cross-bin comparison are highlighted in our segmentation results: better perceptual value and increased robustness with respect to the initialization.

Index Terms—Color Segmentation, Active contours, Maximal Discrepancy, Cross-bin Metrics, Signature Representation.

I. INTRODUCTION

Recent segmentation methods perform image segmentation by finding a contour that minimizes certain energy function. Ideally, when the contour coincides with the real boundary of the object, the energy will be at its global minimum.

In the continuous setting, the energy functional is defined over continuous variables and image domains. Active curve evolution driven by gradient-descent optimization has been the most prevalent and flexible choice in the literature to minimize global functionals [1]. Alternatively, in the discrete setting an image is represented as a graph such that the minimum cut also minimizes the energy functional. Graph-cut methods utilize combinatorial min-cut/max-flow algorithms for computing minimum s-t cuts on graphs. They present the advantage of guaranteeing global optima in nearly real time [2]. However, minimizing certain energy functionals via graph cuts can pose a technically difficult problem. Each contribution proposes a graph construction to minimize a particular energy functional and, in some of these...
cases, the construction is fairly complex [3]. Known limitations of discrete domain energy optimization include metrication artifacts [4], shrinking bias [5] and suboptimality in multiple-label settings [6].

Continuous convex relaxation approaches [1] share advantages of both active curves and graph cuts, and have recently attracted significant attention in image segmentation. Recent work has shown the potential of convex optimization in solving the classical Mumford-Shah segmentation model [7] or the Pott’s model [8]. Current efforts have focused on transforming energy minimization problems into convex formulations compatible with the graph-cut framework.

In the active contours framework, the segmentation process starts with an initial contour obtained automatically or from user interaction; this contour is evolved towards object boundaries under the action of forces (gradient flows) defined on the continuous domain to achieve energy minimization via gradient descent. Energy functionals depend on the geometry of the boundary curve and on image features. Common image features comprise edges (edge-based active contours) and characteristics of the regions occupied by objects (region-based active contours) [9].

The first active contour methods were edge-based. They used edge-based functionals [10]–[15] making the derived gradient flows weak at locations with strong edges. Segmentation with edge-based active contours is affected by the limitations of edge detectors. Weak or noisy edges cause an active contour to pass over real boundaries, while spurious edges may stop the contour. Consequently, a classic drawback of edge-based active contours is their small range of capture, which requires the initial contour to be placed in close proximity to the actual objects.

Features can also be obtained from contour-traced image regions, making active contours more robust to noise and incomplete boundaries [16]–[19]. Some authors [20], [21] propose energies based on a combination of edge and region terms. Alternatively, Zhu and Yuille [22] considered only region features in their approach.

A. Region Features and Descriptors

Region features should be chosen which best discriminate objects and background when characterized over image regions [23], [24]. Obvious region features are intensity and color, but more sophisticated alternatives have been included in the energy functional: vector fields, like the optical flow field [25], [26], motion fields [27], [28], texture [28]–[30] and shape features [31], [32].

Region features usually vary within each region and the goal of a region descriptor is to formalize and measure the variation so that it can be incorporated into the energy functional. The most general region descriptor is the probability density function (PDF) of the corresponding region feature. Some methods approximate PDFs a priori: in [16], [20], [33] image intensities are modeled as Gaussian mixtures and their parameters are ascertained beforehand with an expectation maximization (EM) algorithm; in [34] the user selects image samples that make it possible to estimate the mean and variance for each region. Other methods [22], [35]–[37] estimate the parameters of the PDF at each gradient flow iteration according to the regions enclosed by the evolving contour. This approach is more robust to initialization variability.

B. Histogram Distances

PDFs, as region descriptors, can be approximated by non-parametric kernel density estimation (KDE) [38], [39]. With an appropriate choice of kernel window width,
KDE can describe the data closely. When KDE is performed using the discrete Dirac delta function as kernel and the obtained PDF is quantized, then the result is a normalized histogram: a function relating sample values of a random variable with their relative frequencies.

Two regions described non-parametrically may be compared according to different criteria [40]–[44], particularly by computing a statistical divergence, like the Kullback-Leibler divergence [40] or the Battacharya distance [41]. Both these distribution dissimilarity measures can be considered bin-to-bin measures, in that total dissimilarity is a function of the dissimilarity of each of the bins compared pairwise. An alternative approach is to use cross-bin histogram distance measures, which explicitly incorporate all-to-all comparisons between bins, and therefore consider not only intra-bin but also inter-bin dissimilarities. The disadvantages of bin-to-bin distances were explored in [45], [46]. In bin-to-bin distances only features falling in the same bin are subject to comparison. The result is that bin-to-bin measures do not properly match perceived differences between feature distributions and that they are sensitive to the bin size, since it may affect the degree of overlap between bins for two given histograms. On cross-bin distances, on the other hand, the bin size is not so critical since two bins can be compared in terms of their relative separation (ground distance) with no need for overlap. High correlation with perceived discrepancies can be achieved for cross-bin distances if good definitions of ground distances are available (i.e. if the perceptual difference between bins has been established from psycho-visual models).

One of the most trusted cross-bin distance measures available is the Wasserstein Distance, widely known in computer vision and image processing as the Earth Mover’s Distance (EMD) [45], [46]. The EMD computes the dissimilarity of two distributions in terms of a linear programming optimization problem: the Monge-Kantorovich transportation problem [47]. Indeed, the EMD quantifies distribution dissimilarity as the amount of work needed to reconfigure one distribution into the other as if they were piles of dirt spread over space. If the elemental work required to move one unit of mass from one place to another (i.e. the ground distance) is known, then an optimal way of performing the reconfiguration task can be devised. The EMD is the total work required to achieve this optimal reconfiguration.

C. Histogram-Distance Optimizing Gradient Flows

Gradient flows for energy depending on a histogram distance requires the first variation of the distance with respect to the contour, which can be derived via calculus of variations. This was done for the Bhattacharyya Distance, and compared to the Kullback-Leibler Divergence, in earlier work [41]. Although the EMD has been postulated to better correspond perception [46], its computation requires the solving of the transportation problem, and therefore its first variation cannot be formulated. The exception is for one-dimensional distributions in which ground distances equal the separation –in feature space– between the corresponding bin locations. Subject to these restrictions, an analytical solution for the EMD can be derived, and the expression for the so-called Match Distance [45] obtained. The first variation of the Match Distance is tractable, allowing for EMD gradient flows. Segmentation techniques exploiting the Match Distance can be found in [42], [43].

To the best of our knowledge, multidimensional features have never been compatible with EMD gradient flows. In this paper, we propose linearizing individual solutions to the transportation problem to facilitate the formulation of gradient flows that can maximize multi-
Multidimensional histograms are often sparse representations of the underlying PDF for a sample of random multidimensional variables. A histogram \( \{ h \} \) is a mapping from a set of \( N \)-dimensional vectors \( \{ h_u \} \) to the set of nonnegative reals. These vectors typically represent bins (or their centers) in a fixed partitioning of the relevant region of the underlying feature space, and the associated reals are a measure of the relative mass (weight) of the samples that fall into the corresponding bin, with respect to the total mass of the sample set.

Introduced in [45] and further demonstrated in [48], signatures are far more efficient in the multidimensional case. Unlike histograms (fixed-size structures), a signature \( \{ s \} = \{ s_u = (\bar{m}_u, w_u), u = 1 \ldots N \} \) represents a PDF as a set of \( N \) clusters in feature space (an adaptive partitioning). Following a clustering procedure on the samples, the underlying PDF can be described by a set of \( N \) clusters. Cluster \( s_u \) is defined by a centroid \( \bar{m}_u \) in feature space and by the fraction \( w_u \) of pixels that belong to the cluster among all the pixels in the sample (cluster weight or relative mass). Notice that histograms are a particular case of signatures for which centroids are evenly spaced in feature space. In our framework each gradient flow iteration depends on \( N \) with \( O(N^3 \log N) \), which is the known complexity of solving the transportation problem [47]. Using signatures we can adaptively partition the feature space to convey the underlying PDF, and achieve good fidelity with minimal \( N \). \( N \) can be empirically chosen, greater values will produce better fidelity in the signature representation of the underlying PDF at the cost of greater computational burden.

Distribution dissimilarity can be introduced into image segmentation energy functionals. Maximal Discrepancy Criterion (MDC) energy is inversely proportional to the EMD between feature distributions inside and outside the evolving contour [43]. Match-to-Template energy is proportional to the EMD between a model distribution and that of a given image region. This criterion often requires additional balloon forces, the relative influence of which is frequently difficult to tune [43]. Another criterion, inspired by the Mumford and Shah functional [49], aims to minimize the discrepancy between local inside/outside distributions as compared to the median distribution inside/outside the evolving contour [42].

Active contours can be evolved either explicitly or implicitly. Explicit implementations are dependent on contour parametrization, so handling points may be a daunting task, especially when more image objects are to be detected or when the topology of the contour changes [10], [22], [50], [51]. The level-set method introduced by Osher and Sethian [52] is an alternative that overcomes these difficulties and has thus become a very popular method for active contours. In the level-set approach, an evolving curve \( C(t) \) can be embedded as the zero-level set of a function \( \phi \):

\[ C(t) = \{(x, y)|\phi(x, y, t) = 0\} \text{ with } \phi(x, y, 0) = C_0 \]

where \( C_0 \) is the initial curve. \( \phi \) is very often a signed distance function taken to be positive on the inside and negative on the outside of the curve, or vice versa. The outward normal \( \bar{N} \) of \( C \) can be expressed in terms of the gradient \( \nabla \phi \) of the new function. Thus, under this representation, evolving a contour according to a force acting in the normal direction

\[ C_t = F \cdot \bar{N}, \]

is equivalent to the following partial differential equation:

\[ \phi_t = F|\nabla \phi|. \]

Furthermore, under the level-set representation, key ge-
ometric properties of the curve such as its curvature \( \kappa = \frac{d\vec{T}}{ds} \) (the magnitude of the rate of change of the contour tangent with respect to the arc length), can also be expressed in terms of \( \nabla \phi \), as will be shown in Section II-D.

D. Contributions

In this work we provide the formulation for multi-dimensional EMD gradient flows. We develop a segmentation scheme based on an MDC energy functional and establish the relations between EMD, distributions and contours via calculus of variations. The differential structure of the EMD is obtained as in [53].

II. Methodology

A. The Transportation Problem

In this work EMD will allow for comparisons between feature distributions inside/outside (I/O) the evolving contour. Considering histograms as a particular case of signatures, the EMD between distributions can be defined in terms of centroid locations and weights, as long as ground distances are known.

We consider signatures \( s^I = \{s_u\}_{u=1,...,N} \) (inside the contour) and \( s^O = \{s_v\}_{v=1,...,N} \) (outside the contour). The feature space for the whole image is clustered and then each signature is determined by the same set of \( N \) cluster centroids and different relative weights inside/outside as defined in Subsect. I-C. Ground distances (elemental work of moving one unit of mass from one cluster to another) \( \{d_{uv}\}_{u=1,...,N,v=1,...,N} \) need to be established between every pair of centroid locations, in terms of the perceived difference between the centroids in feature space, which must be available from a psycho-visual model. To compute the EMD it is necessary to solve the transportation problem, i.e. to find the optimal flows \( \{f_{uv}\}_{u=1,...,N,v=1,...,N} \) going from cluster \( u \) in \( s^I \) to cluster \( v \) in \( s^O \) for each of the \( N^2 \) combinations of centroids \( u \) and \( v \) in both signatures. Then, the optimal total work employed in the reconfiguration equals the EMD between the two signatures.

Formally, the optimization problem can be written as:

\[
\text{EMD}(s^I, s^O) = \arg\min_{f_{uv}} Z(f_{uv}, d_{uv}),
\]

where \( Z \), as defined in [45], meets the following expression:

\[
Z(f_{uv}, d_{uv}) = \sum_{u=1}^{N} \sum_{v=1}^{N} d_{uv} f_{uv},
\]

subject to the following constraints:

\[
f_{uv} \geq 0, \quad 1 \leq u \leq N, \quad 1 \leq v \leq N,
\]

\[
\sum_{u=1}^{N} f_{uv} = w^O_v, \quad 1 \leq v \leq N,
\]

\[
\sum_{v=1}^{N} f_{uv} = w^I_u, \quad 1 \leq u \leq N,
\]

\[
\sum_{u=1}^{N} \sum_{v=1}^{N} f_{uv} = 1.
\]

The meaning of these constraints is as follows. A direction of mass transfer must be chosen, for example from \( I \) to \( O \), so that flows are positively defined in that direction. All the mass flowing from/into a cluster must be equal to the initial/final available mass in that cluster. The total mass being transported equals the mass of both normalized signatures (i.e. one).

B. Simplex Method and EMD Sensitivity Analysis

The abovementioned linear programming problem can be considered in geometric terms as finding an optimum in a closed convex polytope. Here the polytope is defined by intersecting \( 2N+1 \) half spaces in an \( N^2 \)-dimensional Euclidean space. Essentially, the Simplex method [54] works by searching the vertices on the boundary of the polytope for an optimum. The Simplex algorithm begins at a starting vertex and moves along the edges of
the polytope until it reaches the vertex of the optimum solution.

Sensitivity analysis is carried out to obtain the derivatives of the optimum flows between clusters with respect to each of the variables on the right-hand side (RHS) of the constraints ($\{w_u\}, \{w_v\}$). Since the EMD depends only on these flows and the ground distances (fixed), such derivatives make it possible to obtain the sensitivity of $Z(f_{uv}, d_{uv})$ with respect to the variations in the weights in (7) and (8), through linearization of (5) in every solution. At this point, the linear programming problem is expressed in a matrix form and the sensitivity analysis from the Simplex method in [53] is retrieved.

1) Simplex Method in Matrix Form: For convenience we here reproduce some of the content of [53] regarding EMD sensitivity analysis. The problem in (4) is first represented in a matrix form. The starting matrix is then transformed to an optimal form based on the Simplex method in [53] is retrieved.

Specifically, since there are $N^2$ variables $f_{uv}$ and $N^2$ constants $d_{uv}$ in (5), we use column vectors $\vec{f}$ and $\vec{d}$, both of size $N^2$, to represent the set of flows and the ground distances as

$$\vec{f} = [f_{11} \cdots f_{1N} \cdots f_{N1} \cdots f_{NN}]^T, \quad (10)$$

$$\vec{d} = [d_{11} \cdots d_{1N} \cdots d_{N1} \cdots d_{NN}]^T. \quad (11)$$

Stacking the first three equations of the equality constraints in (7), (8) and (9), the coefficients $c_{uv}$ ($1 \leq u \leq N, 1 \leq v \leq N$), which are equal to 1, can form a 2D matrix of $2N + 1$ rows and $N^2$ columns. If

$$H = \begin{pmatrix} c_{11} & 0 & \cdots & 0 & \cdots & c_{N1} & 0 & \cdots & 0 \\ \vdots & 0 & \cdots & c_{1N} & \cdots & 0 & \cdots & 0 \\ c_{11} & \cdots & c_{1N} & \cdots & 0 & \cdots & 0 \\ \vdots \\ 0 & \cdots & 0 & \cdots & c_{N1} & \cdots & c_{NN} \\ c_{11} & \cdots & c_{1N} & \cdots & c_{N1} & \cdots & c_{NN} \end{pmatrix}, \quad (12)$$

where $c_{uv} = 1$ indicate that the flow $f_{uv}$ appears in the constraint are represented with a 1 in (12). Then if

$$\vec{b} = [w_1^O \cdots w_N^O w_1^I \cdots w_N^I]^T, \quad (13)$$

we can obtain a matrix form of (4) and (5) as

$$\text{EMD} = \arg\min_{\vec{f}} Z = \arg\min_{\vec{f}} \vec{d}^T \vec{f}, \quad (14)$$

subject to

$$H \vec{f} = \vec{b}, \quad (15)$$

$$\vec{f} \geq \vec{0}. \quad (16)$$

According to the Simplex algorithm [54], since there are $N^2$ variables and $2N + 1$ constraints in the problem, we can always formulate $2N + 1$ basic variables (i.e. variables of nonzero value), and $N^2 - (2N + 1)$ nonbasic variables (i.e. variables of zero value). Grouping all of the basic variables together and all of the nonbasic variables together, we split the flow vector $\vec{f}$ into $[\vec{f}_B^T \vec{f}_{NB}^T]^T$ where the subscript $B$ denotes basic variables, and $NB$...


\[
\begin{array}{c|c|c|c}
Z & f_B & f_{NB} & \text{RHS} \\
\hline
1 & 0 & -d_{NB}^T H_B^{-1} H_{NB} & \\
0 & 1 & d_{B}^T H_B^{-1} H_{NB} & \\
\end{array}
\]

\text{TABLE II} 
\text{REFORMULATED OPTIMAL TABLE}

denotes nonbasic variables. The ground distance vector \( \vec{d} \) and matrix \( H \), are similarly divided as \([d_{B}^T \, d_{NB}^T]^T \) and \([H_B \, H_{NB}] \), respectively. Thus the Simplex starting table is written as in Table I, where RHS denotes the right-hand side of the equations. The second row corresponds to the objective function of (14), and the third row is a vector representation of the constraints in (15) and (16). The superscript \( S \) denotes starting table.

Applying the Simplex algorithm yields an optimal table in which the sets of basic variables and nonbasic variables change, and so do all their coefficients. After matrix transformations [53], the optimal table can be reformulated as shown in Table II. This reformulated optimal table is used for sensitivity analysis.

2) Sensitivity Analysis: Based on the Reformulated Optimal Table (Table II), from [53] we retrieve the sensitivity analysis of the EMD with respect to a change in the cluster weights of the signature. Note that sensitivity analysis is performed both on the \( w^f \) and on the \( w^O \) parts, i.e., the cluster weights corresponding to the signature inside/outside of the contour. In the second row of Table II, we have \( Z = d_{B}^T H_B^{-1} \tilde{b} \). Let \( \tilde{b} \) be changed to \( \tilde{b}' \), where \( \tilde{b}' = \tilde{b} + [0 \ldots 0 \Delta w_i^O \ldots 00]^T \); i.e., one of the weights changes while the others remain the same.

Then

\[
Z' = d_{B}^T H_B^{-1} \tilde{b}' =
\begin{align*}
&= d_{B}^T H_B^{-1} \tilde{b} + d_{B}^T H_B^{1/2} [0 \ldots 0 \Delta w_i^O \ldots 00]^T \\
&= d_{B}^T H_B^{-1} \tilde{b} + k_i \Delta w_i^O ,
\end{align*}
\]

where

\[
k_i = \sum_{i=1}^{2N} (d_{B}^T)(\Delta w_i^O)_{li} ,
\]

\( i = 1, \ldots, 2N \).

Therefore (see Appendix A in [53]),

\[
\frac{\partial Z}{\partial w_i^O} = \lim_{\Delta w_i^O \to 0} \frac{\Delta Z}{\Delta w_i^O} = k_i ,
\]

\[
\frac{\partial Z}{\partial w_i^f} = \lim_{\Delta w_i^f \to 0} \frac{\Delta Z}{\Delta w_i^f} = k_{u+N} ,
\]

with \( v, u = 1, \ldots, N \).

This takes care of the change of \( Z \) due to EMD sensitivity. To incorporate the effect of subsequent normalization let us incorporate the change of weight in one cluster as the weights of the other clusters in the same signature change. Considering this constraint leads (Appendix B in [53]) to

\[
\frac{\partial Z}{\partial w_i^O} = k_v - \sum_{j \neq v} k_j \frac{w_i^O}{\sum_{j' \neq v} w_{j'}^O} ,
\]

\[
\frac{\partial Z}{\partial w_i^f} = k_{u+N} - \sum_{k \neq \overline{u}} k_{k+N} \frac{w_k^f}{\sum_{k' \neq \overline{u}} w_{k'}^f} ,
\]

\( v, u, j, i = 1, \ldots, N \).

Although Zhao et. al. [53] obtained the sensitivity in order to establish the directions of maximum EMD variation in the image domain, we adopt an analogue approach in order to obtain gradient descent evolutions for deforming contours. This contribution occupies the rest of this Section.

C. Multidimensional EMD Gradient Flows

Once we have reviewed some basic results on the Simplex algorithm, and we have retrieved work by Zhao et al. [53] on sensitivity analysis, we are ready to present our contribution, consisting on the derivation of...
gradient flows that maximize multidimensional EMD-derived energies. A common criterion for segmenting with distributions is the Maximal Discrepancy Criterion (MDC), which was in fact the basis for the methodology in [43]. For a given contour, this criterion establishes an energy \( E \) inversely proportional to the EMD between the feature signatures inside (I) and outside (O) the contour:

\[
E = -\text{EMD}(s^I, s^O) .
\]  

(23)

Minimization of this energy is equivalent to the following optimization problem

\[
\underset{\phi(\vec{x})}{\text{argmax}} \text{EMD} ,
\]  

(24)

where the contour is embedded as the zero level-set of the function \( \phi(\vec{x}) \) defined over the image domain \( \Omega \). For a given signature pair, we can model the EMD variations with respect to the signature weights by linearizing the optimal \( Z \) and using \( \frac{\partial Z}{\partial w_i} \) according to (21) and (22). The optimization problem becomes

\[
\underset{\phi(\vec{x})}{\text{argmax}} \underset{f_{uv}(\phi(\vec{x}))}{\text{argmin}} Z(f_{uv}(\phi(\vec{x})), d_{uv}) ,
\]  

(25)

and the derivative of the EMD with respect to the signature weights will be approximated for each individual signature pair following resolution of the linear programming problem:

\[
\frac{\partial \text{EMD}}{\partial w^I_u} \simeq \frac{\partial Z_{\text{min}}}{\partial w^I_u} ,
\]  

\[
\frac{\partial \text{EMD}}{\partial w^O_v} \simeq \frac{\partial Z_{\text{min}}}{\partial w^O_v} .
\]  

(26)

(27)

The chain rule can then be used to produce an expression of the first variation of the EMD with respect to \( \phi \),

\[
\frac{\delta \text{EMD}}{\delta \phi} = \sum_{u=1}^{N} \frac{\partial \text{EMD}}{\partial w^I_u} \frac{\delta w^I_u}{\delta \phi} + \sum_{v=1}^{N} \frac{\partial \text{EMD}}{\partial w^O_v} \frac{\delta w^O_v}{\delta \phi} .
\]  

(28)

If \( H(\cdot) \) is the Heaviside function, and \( \phi(\vec{x}) \) is positive/negative for \( \vec{x} \) inside/outside the contour, we can write the signature weights in terms of \( \phi(\vec{x}) \) as

\[
w^I_u(\phi(\vec{x})) = \frac{\int_{\Omega} C_u(\vec{x}) H(\phi(\vec{x})) d\vec{x}}{\int_{\Omega} H(\phi(\vec{x})) d\vec{x}} ,
\]  

\[
w^O_v(\phi(\vec{x})) = \frac{\int_{\Omega} C_v(\vec{x}) (1 - H(\phi(\vec{x}))) d\vec{x}}{\int_{\Omega} (1 - H(\phi(\vec{x})) d\vec{x}} ,
\]  

(29)

(30)

where \( \Omega \) is the whole image domain and \( C_i(\vec{x}) \) charts the membership of \( \vec{x} \) to the \( i \)-th cluster in the signature:

\[
C_i(\vec{x}) = \begin{cases} 
1 & \text{if } \vec{x} \in s_i \\
0 & \text{otherwise},
\end{cases}
\]  

(31)

Following [24] it is immediate to determine the first variation of the weights with respect to \( \phi \)

\[
\frac{\delta w^I_u}{\delta \phi} = -\frac{\delta(\phi)}{A_I} \left( w^I_u - C_u(\vec{x}) \right) ,
\]  

\[
\frac{\delta w^O_v}{\delta \phi} = \frac{\delta(\phi)}{A_O} \left( w^O_v - C_v(\vec{x}) \right) .
\]  

(32)

(33)

Note that \( \delta(\phi) \) is the continuous Dirac delta function, \( A_{I/O} \) are the areas of the regions left inside/outside of the contour, and \( u/v \) index the clusters in the signatures inside/outside the contour. Finally, the contour evolution minimizing the energy can be obtained by combining (23), (28), (32) and (33) to produce the first variation of the energy \( E \) with respect to \( \phi \), and plugging the result into a gradient descent scheme:

\[
\phi_t = -\frac{\delta E}{\delta \phi} = \frac{\delta \text{EMD}}{\delta \phi} =
\]  

\[
= \frac{\delta(\phi)}{A_O} \sum_{v=1}^{N} \left( w^O_v - C_v(\vec{x}) \right) \frac{\partial \text{EMD}}{\partial w^O_v} -
\]  

\[
- \frac{\delta(\phi)}{A_I} \sum_{u=1}^{N} \left( w^I_u - C_u(\vec{x}) \right) \frac{\partial \text{EMD}}{\partial w^I_u} .
\]  

(34)

As previously stated in [35], some regularization of \( \delta(\phi) \) needs to be adopted. We prefer using \( |\nabla \phi| \), which has the advantage of extending the gradient descent to all level-sets of \( \phi \) [55]. This time rescaling does not affect the steady state solution, but it does remove stiffness near the zero level sets of \( \phi \). Only the speed of descent, not its direction, is affected [55].
As shown in Section I-C, this gradient descent of $\phi$ implicitly affects its zero-level set according to the following contour evolution [24]:

$$\frac{\partial C}{\partial t} = FN,$$

where

$$F = \frac{1}{A_O} \sum_{v=1}^{N} \left( w^O_v - C_v(x) \right) \frac{\partial \text{EMD}}{\partial w^O_v} - \frac{1}{A_I} \sum_{u=1}^{N} \left( w^I_u - C_u(x) \right) \frac{\partial \text{EMD}}{\partial w^I_u},$$

and $\tilde{N}$ stands for the contour’s normal. This balloon force will minimize energy $E$ as long as we solve the linear programming problem for every time $t = t_0 + \Delta t$ in the discretization of the gradient descent equation, and update the values of $\frac{\partial \text{EMD}}{\partial w^I_u}$ and $\frac{\partial \text{EMD}}{\partial w^O_v}$ according to (26) and (27).

### D. Contour Regularization

The energy functional in (23) can be thought of as accounting for the fidelity of estimation of the optimal level-set function to observed features in the enclosed regions. However, this cost function does not take into consideration some plausible properties of the optimal solution, and, as a result, minimizing (23) alone may be too sensitive to measurement noise and/or errors in the data. To eliminate this drawback, we can attempt to filter out the spectral components of the solution corresponding to the noise subspace. For the case at hand, we can regularize the solution by constraining the length of the active contour [10] to favor smooth contours, in which case the energy functional is given by

$$E = -\text{EMD}(s^I, s^O) + \mu \text{Length}(C),$$

where $\mu > 0$ is a regularization constant controlling the trade-off between fidelity and stability. Translating this into the level-set framework leads to the following optimization problem:

$$\arg\min_{\phi(x)} \left( -\text{EMD}(\phi) + \mu \int_{\Omega} |\nabla H(\phi)| d\Omega \right).$$

The gradient flow associated with minimizing the cost functional in (38) can be shown [35] to be equal to

$$\phi_t = \frac{\delta \text{EMD}}{\delta \phi} - \mu \kappa \delta(\phi),$$

where $\kappa$ is the contour curvature –as defined in Section I-C– and meets

$$\kappa = \nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right).$$

This is equivalent to contour evolution in the normal direction according to the force

$$F_{\text{reg}} = F - \mu \kappa,$$

with $F$ as defined in (36).

### III. EXPERIMENTS

In this Section we provide some experimental evidence for the main advantages of our method. The main contribution of our work is the possibility of evolving active contours according to gradient flows that optimize the EMD between multidimensional distributions. EMD has proven to correlate better with perception [45], [46], and EMD-optimizing gradient flows have been proposed for one-dimensional distributions in the past [42], [43]. In [43], EMD gradient flows were compared with bin-to-bin Bhattacharyya gradient flows [41].

We compare our proposed framework (EMD N-D) with Bhattacharyya (B) gradient flows [41] and with one-dimensional EMD (EMD 1-D) gradient flows [43] in the context of color image segmentation.

#### A. Methodologies for Comparison

Bhattacharyya gradient flows as presented in [41] are readily applicable to distributions in any dimensions expressed either as histograms or as signatures.
Bhattacharyya distance does not account for any cross-bin interactions in the distributions and therefore makes comparisons on a bin-to-bin basis.

Bhattacharyya distance maximization in the active contour formulation is equivalent to the following optimization problem:

\[
\arg\min_{\phi(\vec{x})} \tilde{B}(\phi(\vec{x})) = \arg\min_{\phi(\vec{x})} \sum_{i=1}^{N} \sqrt{w_i^I w_i^O}, \quad (42)
\]
given that Bhattacharyya distance \( B = -\log \tilde{B} \). Contour evolution in the normal direction producing the gradient descent for such optimization therefore corresponds \cite{41} to the following force:

\[
F(\vec{x}) = - \frac{\tilde{B}}{2} \left( A^{-1}_I - A^{-1}_O \right) - \frac{1}{2} A^{-1} \left( \sqrt{\tilde{w}^I \cdot \tilde{C}(\vec{x})} + \sqrt{\tilde{w}^O \cdot \tilde{C}(\vec{x})} \right) + \frac{1}{2} A^{-1} \sqrt{\tilde{w}^I \cdot \tilde{C}(\vec{x})}, \quad (43)
\]

where

\[
\tilde{C}(\vec{x}) = [C_1(\vec{x}), \ldots, C_N(\vec{x})]^T, \quad (44)
\]
with \( C_i(\vec{x}) \) defined as in (31), and

\[
w_i^{I/O} = [w_1^{I/O}, \ldots, w_N^{I/O}]. \quad (45)
\]

EMD 1-D gradient flows as proposed in \cite{43} benefit from a closed-form solution to the transportation problem for one-dimensional distributions. Extra constraints for this closed solution are: (i) the unit transportation cost from one bin to another (ground distance) equals the distance between the bins, and (ii) the bins sample the distribution uniformly. As a consequence EMD 1-D can be applied to distributions of gray-level intensity values obtainable from color images. We used a widely preferred rule for the conversion: \( GL = 0.2989 R + 0.5870 G + 0.1140 B \). Constraint (ii) requires the use of histograms instead of signatures.

If the energy \( E = -\text{EMD}(P_{in}, P_{out}) \) as for EMD N-D with maximum discrepancy criterion, \cite{43} proved that such energy can be minimized in terms of the cumulative distribution functions CDF of probability density functions \( P_i \), inside and outside the evolving contour \( C \) embedded in \( \phi(\vec{x}) \):

\[
\arg\max_{\phi(\vec{x})} \text{EMD}(P_{in}, P_{out}) = \arg\max_{\phi(\vec{x})} \sum_{i=1}^{N} |CDF_{in}(z_i) - CDF_{out}(z_i)|, \quad (46)
\]
As a consequence, the resulting force field to provide a gradient descent maximization of EMD for contour \( C \) is as follows:

\[
F = \sum_{z_i} -s_i F_{z_i}, \quad (47)
\]
where

\[
s_i = \text{sign}((CDF_{in}(z_i) - CDF_{out}(z_i))) \quad (48)
\]
\[
F_{z_i} = \left( T_{z_i} \left( \frac{1}{A_{in}} + \frac{1}{A_{out}} \right) - \frac{C_{in}}{A_{in}} - \frac{C_{out}}{A_{out}} \right), \quad (49)
\]
\[
T_{z}(x,y) = \begin{cases} 1, & I(x,y) \leq z, \\ 0, & I(x,y) > z, \end{cases} \quad (50)
\]
z\(_i\) is the \( i \)-th discretized value (bin center) of the random variable \( z \), \( I(x,y) \) is the realization of the random variable at image location \((x,y)\), \( A_{in/out} \) is the area inside/outside of the contour, and \( C_{in/out} \) is the number of pixels inside/outside of the contour for which \( I(x,y) \leq z_i \).

As for EMD N-D, we can add a length regularizer:

\[
F_{\text{reg}} = F - \mu \kappa, \quad (51)
\]
for both B and EMD 1-D.

B. Color Signatures

To illustrate the properties of our novel gradient flows, we propose a segmentation scenario previously incompatible with EMD gradient flows: color images. Image color is usually represented as a three-component vector, in accordance with the tri-stimulus theory of human vision that models the frequency response of the
The 1976 CIE $L^*a^*b^*$ (CIELAB or Lab for simplicity) color space was designed according to psycho-visual experiments to allow Euclidean distances measured in the new space to be made proportional to the perceived similarity between colors [57], [58]. We therefore employ the Lab space coordinates as a three-dimensional feature in our segmentation scheme. Ground distances are made equal to the Euclidean distance in Lab, and thus the cost of mass transfer between bins matches perceptual differences.

C. Implementation

For implementing the proposed and benchmark gradient flows we used a sparse scheme following the well-validated methodology in [60], based on integer-valued level-set functions and move-in/move-out operations corresponding to the sign of the forces obtained from (36), (43) and (47) for each of the methods.

Due to the impossibility of approximating second-order derivatives of contour shape by sparse level-set functions, the length regularizer, which evolves with the curvature, is implemented via binary convolutions. The idea of using convolution operations to generate mean curvature motion was originally proposed by Merriman et al. in their work on diffusion generated mean curvature motion [61]. In [60], this idea is further pursued. Their solution is to approximate curvature diffusion by convolving the characteristic function (binarization of the regions inside/outside of the contour) with Gaussian kernels and thresholding the result. As Gaussian $\sigma$ becomes smaller, the approximation gets better.

Clustering of feature space is therefore necessary for signature construction. Many techniques for both supervised and unsupervised clustering are available in literature. Here, we opted to use supervised k-means clustering [59], i.e. Euclidean-distance k-means with an empirically fixed number of clusters. The EMD ground distances – i.e. the work needed to transfer one unit of mass from one cluster to another – are thus made equal to the Euclidean distance between cluster centers in the feature space (Lab).

An example of color image signature is provided in Fig. 1. Note how the signature representation conveys the color content of the image as a set of clusters (colored spheres) with different masses (sphere sizes).
smoothing behavior, arising from approximations performed for computational efficiency, is different from that obtained from standard level-set-based curvature motion. However, the aforementioned authors argue that this thresholded type of smoothing regularization is actually more desirable in many practical image processing applications because it only eliminates small structures with high curvature while leaving coarse scale structures intact. For details please see [60].

Using a Matlab-based implementation our methodology has a computational cost of only 0.056 ms per pixel (i.e., about 15 seconds for a $512 \times 512$ image). For a signature with $N = 30$ clusters, this is higher than that of the main benchmark approach (0.033 ms per pixel) but in the same order of magnitude. The results, however, are significantly better, as is described in the following Sections.

### D. Maximal Discrepancy Results

We will now analyze some of the advantages introduced by the proposed technique. We will benchmark results in terms of state-of-the-art bin-to-bin Bhattacharyya gradient flows [41], and EMD 1-D [43], analyzing the main benefits derived from cross-bin distance maximization for multidimensional (color) features. We chose a fixed number of clusters $N = 30$ in $Lab$ space. For the EMD N-D case, ground distances will equal Euclidean distances in feature space. For EMD 1-D we used gray-level histograms with 30 bins.

We will evaluate results for synthetic/natural images, with different initializations (assisted and generic) using the three schemes. For our proposed framework, we will study the influence of the weight of the regularizer and the number of clusters in the model. This will provide a means of justifying the techniques presented.

1) Segmentation Accuracy: For an example set of 8 images we provide a detailed comparison of the results produced by the three methods (EMD 1-D, B, EMD N-D). Manual segmentations were provided by an external operator for ground truth. In Fig. 2-9 we show the results for synthetic/natural images, initialized with a rectangle overlapping the object of interest (assisted initialization) or with a generic pattern of circular regions. For the
Fig. 3. Segmentation of ‘Blue Fish’ image. (a) Color distribution in $\text{Lab}$ space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) B final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for B final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, B and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), B (blue), EMD N-D (green).

referenced figures, $N = 30$ and the regularizer is implemented via Gaussian filtering of the characteristic function ($\sigma = 3$).

All these examples show a specific set of results:

- The signature of the image.
- The original image with the initialization contour.
- The distribution of clusters inside (red) and outside (blue) the initialization contour. Intermediate colors (shades of purple) indicate partial membership inside/outside. Blue and red are complementary colors that were chosen so that the percentages of the cluster lying inside (red) and outside (blue) the contour can be easily grasped visually.

- Final contours for EMD 1-D, B and EMD N-D.
- The distribution of clusters inside (red) and outside (blue) the EMD 1-D, B, and EMD N-D final contours.
- The evolution of the discrepancy functional at each iteration of contour evolution for the three methods. The value of the discrepancy is normalized by the value attained at convergence. The purpose of this graph is to prove that the contours actually maximize (local) discrepancy by all three methods.
- The evolution of sensitivity, specificity and Dice’s coefficient according to the ground truth for the three methods.
- Final sensitivity, specificity and Dice’s coefficient for EMD 1-D, B and EMD N-D (Table III).

In Fig. 2-6 segmentation is initialized with a rectangular selection. In Fig. 2, the three methods achieve acceptable results. EMD 1-D is color blind and favors a segmentation into bright and dark regions, which in this case produces a segmentation with good accuracy (Dice = 0.945). B distinguishes different colors but is unable to establish meaningful relations between colors in different clusters. For B all color clusters are equally different to each other. This does not prevent good segmentation (Dice = 0.965) since the initialization is close enough to the solution. The best result is achieved for EMD N-D (Dice = 0.970). In Fig. 3, on the other hand, EMD 1-D fails to provide satisfactory results (Dice = 0.489) since gray levels do not differentiate the object from the background. B again provides good results but misses the upper part of the foreground object. Note in Fig.
3h-i, that this missing part of the object corresponds to a (electric blue) cluster that is completely inside for EMD N-D (Dice = 0.956) and almost entirely outside for B (Dice = 0.933). This cluster was also left outside the initialization (Fig. 3b-c) but EMD N-D successfully captures its location in the color space and correctly assigns it to the interior of the contour. In Fig. 4 the object shape makes the rectangular initialization exhibit a cluster distribution very far removed from the solution. The worst results are for EMD 1-D (Dice = 0.672). In Fig. 4g note how EMD 1-D tends to provide solutions that separate clusters by planes normal to the \(L\) axis, since luminance is closely related to gray level. The results for B are slightly better (Dice = 0.840), but again reveal (see Fig. 4h) B’s inability to switch clusters from inside/outside sets. EMD N-D again obtains accurate results (Dice = 0.978). Note how in Fig. 4i EMD N-D successfully finds the best configuration in feature space to maximize discrepancy. In Fig. 5 the inability of B to reassign clusters results in very poor accuracy (Dice = 0.647). EMD 1-D finds a separation approximately normal to the \(L\) axis and assigns part of the object to the background (Dice = 0.887). EMD N-D finds the object with good accuracy (Dice = 0.956). Fig. 6 shows an example with a synthetic image of pure colors and added Gaussian noise in the three color planes. In this case EMD 1-D (yellow contour) selects the colors with similar luminance (Dice = 0.744). B (blue contour) produces worse results (Dice = 0.859) in this case, since the initialization is far removed from the solution. In fact B produces a distribution of inside/outside clusters that matches that of the initialization. Finally, EMD N-D produces perfect results (Dice = 1) and dissects the color space properly thanks to its awareness of cluster location in 3-D.

In Fig. 7 and 8 we performed segmentations with the three methods according to a generic initialization scheme (a grid of \(L \times L\) evenly spaced circles, with \(L = 3\) and 4, respectively). With this type of initialization B performs very poorly in both cases (Dice = 0.443, 0.730 respectively). EMD 1-D achieves better results (Dice = 0.847) for Fig. 7, but not as good as those for EMD N-D, since the gray level of the object is very close to the gray level of the background in some boundary-adjacent regions (corresponding to the clusters with highest \(b\) coordinate in Fig. 7g). EMD 1-D does provide very good accuracy (Dice = 0.988) in synthetic Fig. 8 where luminance discrimination matches

<table>
<thead>
<tr>
<th>Method</th>
<th>EMD 1-D</th>
<th>B</th>
<th>EMD N-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Tree'</td>
<td>0.900</td>
<td>0.996</td>
<td>0.945</td>
</tr>
<tr>
<td>'Flower'</td>
<td>0.999</td>
<td>0.345</td>
<td>0.672</td>
</tr>
<tr>
<td>'Blue Fish'</td>
<td>0.395</td>
<td>0.913</td>
<td>0.488</td>
</tr>
<tr>
<td>'Pelican'</td>
<td>0.815</td>
<td>0.992</td>
<td>0.887</td>
</tr>
<tr>
<td>'Stripes'</td>
<td>1.000</td>
<td>0.488</td>
<td>0.744</td>
</tr>
<tr>
<td>'Landscape'</td>
<td>0.948</td>
<td>0.721</td>
<td>0.847</td>
</tr>
<tr>
<td>'Rectangles'</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

TABLE III
Sensitivity, Specificity and Dice’s Coefficient for Test Images.
Fig. 4. Segmentation of ‘Flower’ image. (a) Color distribution in Lab space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) B final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for B final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, B and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), B (blue), EMD N-D (green).

Fig. 5. Segmentation of ‘Pelican’ image. (a) Color distribution in Lab space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) B final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for B final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, B and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), B (blue), EMD N-D (green).

the solution. In both cases EMD N-D performs well (Dice = 0.950, 0.988 respectively).

Note that EMD N-D performance is high for both manual and automatic initializations. The experiments with automatic initialization clearly reveal that EMD N-D is color-aware and knows how to dissect the color space, independently of the initialization. On the other hand, B contours have no reference of inter-cluster relations and therefore, can adjust memberships (shades of purple) but not really shuffle clusters in and out of the evolving B contour. We conducted an experiment in which the ‘Flower’ image was segmented by EMD N-D using different initializations (a grid of \( L \times L \) evenly spaced circles with \( L = 1, 2, \ldots, 5 \)) and no regularizer (so that only the EMD maximization effect was revealed). The standard deviation of the final EMD value for varying initializations with our method was 0.0076 and
Fig. 6. Segmentation of ‘Stripes’ image. (a) Color distribution in \( L\alpha b \) space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) \( B \) final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for \( B \) final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, \( B \) and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), \( B \) (blue), EMD N-D (green).

the average overlap between solutions was 99.97%. This suggests high robustness against initialization variability, so our method can be considered automatic. This property is also present in EMD 1-D, and has now been made available for N-dimensional feature spaces. We know that in cases when a global optimizer is intractable (as in active contour evolution), local optimizers (e.g. gradient descent) are heavily influenced by the profusion of local extrema, and it is desirable to have an energy functional with as few local extrema as possible. We propose that this functional should take into account estimated probability density functions (e.g. histograms or signatures) in a cross-bin fashion. The opposite occurs with \( B \), for which the description of the color contents in the inside/outside regions is too poor (bin-to-bin) to provide good behavior of extrema in the discrepancy functional. Alternatively, EMD contours seem always to find their way in feature space, the reason being that they incorporate information on the location of bins/clusters.

Even in cases where the maximal discrepancy criterion is not adequate (there are multiple semantic regions of equal importance), like in Fig. 9, this consistency in the space of features can be appreciated for EMD contours. Random initialization produces the inside/outside distribution illustrated in Fig. 9c. The EMD N-D contour finds a maximal discrepancy result that separates blue and yellow hues together and apart from the rest of the image (Fig. 9i). This is also the case for EMD 1-D which successfully separates bright and dark regions of the image (Fig. 9d). Note that this image has not been included in Table III. The reason is that no ground truth is possible for this image since it does not include a single semantic region on a definite background.

2) Influence of the Regularizer: To characterize the effects of the regularizer on contour length and segmentation results, Fig. 10 shows an example segmentation with varying degrees of regularization. Note how in Fig. 10b and Fig. 10d-g contour length is reduced as the variance of the Gaussian filter is increased, and the maximum achieved EMD is reduced as a consequence of the increasing weight of the regularization term. As the regularizer gains relative importance in the optimization problem, the EMD part of the optimization suffers, producing shorter contours. In this case the best segmentation occurs for \( \sigma = 2 \) (see Fig. 10c).
3) **Influence of the Number of Clusters:** The number of cluster is of critical importance in the computational performance of the algorithms. In each evolution iteration EMD sensitivity is computed with complexity $O(N^3 \log N)$ where $N$ is the number of clusters in the signature. On the other hand, higher numbers for $N$ result in closer representations of the underlying color distribution.

![Fig. 7. Segmentation of ‘Landscape’ image. (a) Color distribution in Lab space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) B final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for B final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, B and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), B (blue), EMD N-D (green).](image)

![Fig. 8. Segmentation of ‘Rectangles’ image. (a) Color distribution in Lab space. (b) Original image with initialization contour. (c) Inside (red) and outside (blue) distribution for initial contour. Intermediate colors indicate partial membership. (d) EMD 1-D final contour. (e) B final contour. (f) EMD N-D final contour. (g) Distribution inside/outside EMD 1-D final contour. (h) Distribution for B final contour. (i) Distribution for EMD N-D final contour. (j) Evolution of discrepancies/distances EMD 1-D, B and EMD N-D (normalized by maximum value attained). (k) Evolution of sensitivity. (l) Evolution of specificity. (m) Evolution of Dice’s coefficient. Colors: EMD 1-D (yellow), B (blue), EMD N-D (green).](image)

Fig. 11 shows segmentations for the ‘Crayons’ image with varying number of signature clusters. We use the same regularizer ($\sigma = 3$) as for Fig. 2-9. Notice how, for a precise initialization over the orange part of the image, as the number of clusters increases the contour is more selective for the desired hues. When the number of clusters is too low, the underlying distribution is not faithfully conveyed, and therefore the contour is not capable of telling the difference between color shades. Bigger signatures allow precise color discrimination but
IV. CONCLUSIONS

In this paper we propose a novel active contour framework for segmentation based on the Earth Mover’s Distance between multidimensional feature distributions. The lack of closed-form expressions for multidimensional EMD (or one-dimensional EMD with arbitrary ground distances) required us to linearize the solution to the transportation problem in order to approximate the differential structure of EMD. The approximate derivatives were then employed to build a force field that optimizes the EMD discrepancy between the distributions kept inside/outside of the evolving contour.

We have shown results for some test images and illustrated some of the nice properties of EMD as a means of measuring dissimilarity between distributions (initialization independence and improved accuracy [45], [46]). Although the computational cost of multidimensional EMD rises with signature size $N$ as $N^3 \log N$, the descriptive ability of EMD when comparing distributions is very high even for low $N$. The cost is also compensated by the approximate implementation of level-set gradient flows.

We have presented results involving three-dimensional (color) features in a maximum discrepancy scheme. This context was chosen for its immediacy and interest in the field, but the applications of the methodology proposed herein go far beyond color image segmentation.
Fig. 11. Influence of $N$ on ‘Crayons’ image segmentation. (First column) Images showing initialization and final EMD contour. (Second column) Color distribution in Lab. (Third column) Distribution inside (red) and outside (blue) EMD final contour. Intermediate colors indicate partial membership. (First row) $N = 5$. (Second row) $N = 10$. (Third row) $N = 20$. (Forth row) $N = 40$.

The same framework can be used to address segmentation of any multidimensional feature that is piecewise smooth in the image domain (e.g. texture). All energies based on distribution discrepancy can be derived analogously, as we did for the maximum discrepancy energy. This includes discrepancies with respect to a pre-trained distribution model. Also multi-region active contour schemes can be devised using coupled contours [62] with minor modifications. Ultimately, all image processing procedures that rely on gradient flows can now benefit from multidimensional EMD-based energy terms, including (but not limited to) image registration, enhancement and optical flow estimation.

Finally, although we have proposed an active contour optimization for our EMD-based energy functional, an alternative scheme can be devised in which the differential structure of EMD can be exploited by means of an iterative graph-cut approach (e.g. GrabCut [63]). Segmenting with distributions is not exclusive to the active contour framework, and a graph-cut implementation would possibly entail alternative/additional benefits.

That, however, is beyond the scope of this paper and we leave it for future exploration.

ACKNOWLEDGMENTS

This research was partially funded by TEC2010-21619-C04-02 (CICYT, Spain), and also by a Ph.D. fellowship grant from University of Seville, Spain.

REFERENCES


Carlos S. Mendoza received his Ph.D. in Electronics and Signal Processing from the University of Sevilla, Spain, in 2011. His Ph.D. dissertation was titled “Image Processing in Medicine – Advances for Phenotype Characterization, Computer-Assisted Diagnosis and Surgical Planning.” During 2012, Dr. Mendoza enjoyed a double affiliation shared between University of Sevilla and the Sheikh Zayed Institute for Pediatric Surgical Innovation – Childrens National Medical Center of Washington DC, USA. He is presently a Madrid-MIT M+Vision Fellow at the Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT) in Boston, USA. His current research interests include image segmentation, registration, texture and shape analysis and machine learning for computer-assisted diagnosis, image-guided interventions, surgical planning and imaging biomarkers.

José-Antonio Pérez-Carrasco received his Ph.D. in Electronicas and Signal Processing from the University of Sevilla, Spain, in 2011. His Ph.D. dissertation was titled “Simulation Tool for Assembling and Analyzing Hierarchical AER Systems for Visual Processing”. His current research interests include image processing and its medical applications, visual perception and real-time processing.

Aurora Sáez received her degree in Technical Industrial Engineering from the University of Cordoba, Spain, in 2003 and Telecommunication Engineering from the University of Seville, Spain, in 2008. Since 2009, she is Ph.D. student in the Signal Processing and Communications Department of the University of Sevilla, Spain. Her research interests include color image processing with biomedical applications, focusing on skin images and cellular imaging.

Begoña Acha received her Ph.D. in Communication Engineering in 2002, from the University of Sevilla, Spain. She has been teaching and conducting research at the University of Sevilla since 1996. Her present position is Tenured Professor at Signal Processing and Communications Department, University of Sevilla. Her current research activities include works in the field of color image processing and its medical applications. Skin images, retinal images, computed tomography images are examples of applications of her research. She has published numerous papers in journals and conferences and she has co-authored the book “Color Image Processing with Biomedical Applications”, published by SPIE Press, Bellingham, Washington, USA, 2011.
Carmen Serrano received her Ph.D. degree in Communication Engineering in 2002, from the University of Sevilla, Spain. She has been teaching and conducting research at the University of Sevilla since 1996. At present, she is a Tenured Professor in the Signal Processing and Communications Department of the University of Sevilla. Her research interest is mainly focused on image processing. In particular, she conducts research on the detection of pathological signs in medical images and on the segmentation of anatomical structure for surgical planning. One of her main research fields is color image processing. She has developed algorithms for computer-assisted diagnosis of burn images, pigmented lesions of the skin and retinal images. Most of her research has been applied to medical images, including skin images as well as radiological images from modalities such as computed tomography and magnetic resonance imaging. She has co-authored the book “Color Image Processing with Biomedical Applications”, published by SPIE Press, Bellingham, Washington, USA, 2011, and she has published numerous papers in journals and prestigious conferences.