A VERTEX-BASED REPRESENTATION OF OBJECTS IN AN IMAGE

Gozde Unal, Hamid Krim

NC State University
ECE Department
Raleigh, NC 27695

Anthony Yezzi

Georgia Institute of Technology
School of Electrical Eng.
Atlanta, GA 30332

ABSTRACT

Novel polygon evolution models are introduced in this paper for capturing polygonal object boundaries in images which have one or more objects that have statistically different distributions on the intensity values. The key idea in our approach is to design evolution equations for vertices of a polygon that integrate both local and global image characteristics. Our method naturally provides an efficient representation of an object through a few number of vertices, which also leads to a significant amount of compression of image content. This methodology can effectively be used in the context of MPEG-7. We also propose usage of the Jensen-Shannon criterion as an information measure between the densities of regions of an image to capture more general statistical characteristics of the data.

1. INTRODUCTION

Emerging multimedia applications such as searching for multimedia material of one's interests over large databases quickly reveal a need for efficient access and representation of visual objects in a scene. An efficient description of an object is through its contour representation. This representation may later be compactly stored and provide a significant amount of compression of the multimedia content. Several techniques such as in [1, 2] for shape coding entailed first extracting object boundaries and later finding a polygonal approximation of the extracted contour. The locations of the resulting polygon vertices are also encoded for further data compression.

We propose in this paper, a technique that directly captures a polygonal representation of objects in images. Our approach, hence, produces vertex locations of objects in the scene in one shot rather than performing this search for meaningful locations in several steps. Our method, which provides a polygon as the resulting object description, can also be efficiently used for visual image retrieval or indexing based on an object's shape captured by a few vertices. This relates our technique to MPEG-7 framework which focuses on multimedia content [3] .

Objects are delineated (and approximated) by moving vertices of a polygon according to some energy criterion. Specifically, we develop a fast numerical approximation of an information-theoretic measure which drives the flow of polygon vertices on the basis of the prevailing intensity distributions of regions enclosed by the polygons in an image. This is, as we later show, achieved by a system of coupled Ordinary Differential Equations (ODE's). In contrast to the region-based active contours framework [4–9], where continuous curves are utilized, our approach offers significant advantages over them by providing a very compact description particularly for simply connected target shapes. Similarly, our approach differs from the early snakes [10] in that the number of vertices that are involved in our flows are remarkably lower than the number of contour points used by the snake-based methods.

An energy functional over regions delineated by a number of polygons are defined in terms of an information measure among the distributions of the related regions. The so-called generalized Jensen-Shannon divergence measure [11, 12] not only provides a more general description of the properties of the regions than only mean, and variance as the representative statistics, but also gives a natural extension to a distance criterion among a finite number of distributions over the regions. We are then able to utilize multiple polygons moving over the image to capture several meaningful regions.

In the remainder of the paper, we first derive the polygon evolution equations, then introduce the information-theoretic criterion for evolution of a single contour or multiple contours. Some simulation results are presented followed by the conclusions.

2. POLYGON EVOLUTION MODELS

In this section, we present gradient flows designed to move polygon vertices with the goal of parsing an image domain into meaningfully different regions. A general form of an energy functional over a region R which is to capture various objects can be written as a contour integral

$$E(\boldsymbol{C}) = \iint_R f(x,y)\,dx\,dy = \oint_{\boldsymbol{C}=\partial R} \langle \boldsymbol{F}, \boldsymbol{N} \rangle\, ds,$$

where $\boldsymbol{N}$ denotes the outward unit normal to $\boldsymbol{C}$ (the boundary of R), ds the Euclidean arclength element, and where $\boldsymbol{F} = (F^1, F^2)$ is chosen so that $\nabla \cdot \boldsymbol{F} = f$. It can be shown [7] that a gradient flow for $\boldsymbol{C}$ with respect to E may be written as

$$\frac{\partial \boldsymbol{C}}{\partial t} = f\boldsymbol{N}.$$

The key to our approach is to consider a closed polygon $\boldsymbol{V}$ as the contour $\boldsymbol{C}$, and with a fixed number of vertices, say $n \in \mathbb{N}$, $\{\boldsymbol{V}_1, .., \boldsymbol{V}_n\} = \{(x_i, y_i), i = 1, \dots n\}$. We may parameterize $\boldsymbol{C}$ by $p \in [0, n]$ as

$$\boldsymbol{C}(p, \boldsymbol{V}) = L(p - \lfloor p \rfloor, \boldsymbol{V}_{\lfloor p \rfloor}, \boldsymbol{V}_{\lfloor p \rfloor + 1})$$

where $\lfloor p \rfloor$ denotes the largest integer which is not greater than p, and where $L(t, \boldsymbol{A}, \boldsymbol{B}) = (1-t)\boldsymbol{A} + t\boldsymbol{B}$ parameterizes between 0 to 1 the line from $\boldsymbol{A}$ to $\boldsymbol{B}$ with constant speed. Note that the indices of $\boldsymbol{V}$ should be interpreted as modulo n so that $\boldsymbol{V}_0$ and

$\boldsymbol{V}_n$ denote the same vertex (recall $\boldsymbol{C}$ is a closed curve). Finally, note that $\boldsymbol{C}_p$ is defined almost everywhere (where $p \neq \lfloor p \rfloor$) by

$$\boldsymbol{C}_p(p, \boldsymbol{V}) = \boldsymbol{V}_{\lfloor p \rfloor + 1} - \boldsymbol{V}_{\lfloor p \rfloor}.$$

Using the above parameterization $\boldsymbol{C}_p(p, \boldsymbol{V})$, we obtain the first variation of the energy functional E as

$$E_v = \int_0^n f\big(L(p - \lfloor p \rfloor, \boldsymbol{V}_{\lfloor p \rfloor}, \boldsymbol{V}_{\lfloor p \rfloor + 1})\big) \times \langle \boldsymbol{C}_v, J(\boldsymbol{V}_{\lfloor p \rfloor + 1} - \boldsymbol{V}_{\lfloor p \rfloor}) \rangle \, dp,$$

yielding a minimization of E by way of a gradient descent flow given by a set of differential equations for each vertex V_k as

$$\frac{\partial \boldsymbol{V}_k}{\partial t} = \int_0^1 p f\big(L(p, \boldsymbol{V}_{k-1}, \boldsymbol{V}_k)\big) \, dp \, \boldsymbol{N}_{k,k-1} + \int_0^1 (1-p) f\big(L(p, \boldsymbol{V}_k, \boldsymbol{V}_{k+1})\big) \, dp \, \boldsymbol{N}_{k+1,k}, \quad (1)$$

where $\boldsymbol{N}_{k,k-1}$ (resp. $\boldsymbol{N}_{k+1,k}$) denotes the outward unit normal of edge $(\boldsymbol{V}_{k-1} - \boldsymbol{V}_k)$ (resp. $(\boldsymbol{V}_k - \boldsymbol{V}_{k+1})$), and which in the interest of space, we defer the proof of to [13]. These ODE's are solved simultaneously for each vertex of a polygon. Note how the information through the functional f is being integrated along adjacent edges for a vertex. In addition to a small number of vertices, their well-separated locations clearly distinguish our proposed approach from the snake-based methods.

Upon obtaining a polygonal evolution model for a generic energy functional, we next choose a "good" criterion that captures the underlying statistical properties of each region delineated by a contour (particularly a polygon) and defines a metric among them. This may be achieved by an information divergence measure that defines a distance among probability densities. On account of its properties [12], we pick as our energy functional the Jensen-Shannon criterion which, in its general form may be written for N probability densities (with prior probabilities, $\boldsymbol{a} = (a_1, \ldots, a_N)$ such that $\sum_{i=1}^N a_i = 1$) as

$$JS_a = H\left(\sum_{i=1}^N a_i p_i(\xi)\right) - \sum_{i=1}^N a_i H(p_i(\xi)) \quad (2)$$

where $p_i(\xi)$ denotes the probability density of pixel values in the i^{th} region, and H is the Shannon entropy. One of the many features of the Jensen-Shannon divergence is that different weights may be assigned to the distributions involved and according to their importance. Estimation of a probability density and its entropy may be carried out in a variety of ways, we hence adopt a first order approximation of a density which achieves a maximum entropy solution, which is in turn used in approximating the entropy expression as proposed in [14].

We assume a scalar random variable (r.v.) $\boldsymbol{\xi} : \Xi \subset \mathbb{R} \to \mathbb{R}$ on the intensity values of a region delineated by an active polygon, and the information available on the density of the r.v. ξ is given by

$$\int_{\Xi} p(\xi) G_j(\xi) d\xi = u_j, \quad \text{for } j = 1, .., m, \quad (3)$$

where the estimates u_j are the expectations $E\{G_j(\boldsymbol{\xi})\}$ of m different independent functions $\{G_i(\cdot)\}$ of $\boldsymbol{\xi}$. Note that there is no model assumption for the random variable $\boldsymbol{\xi}$, however, the distribution which has the maximum entropy and which is also compatible with the measurements in Eq. (3) is sought [15, 16]. A simple approximation of the entropy functional is found in [14] upon substituting a first-order approximate density in

$$H(\hat{p}(\xi)) = -\int_{\Xi} \hat{p}(\xi) \log \hat{p}(\xi) d\xi \approx H(\nu) - \frac{1}{2}\sum_{j=1}^m u_j^2, \quad (4)$$

where $H(\nu) = \frac{1}{2}(1 + \log(2\pi))$ is the entropy of a standardized Gaussian variable.

3. EVOLUTION OF A SINGLE CONTOUR

With a single contour, the image domain Ω is split into two regions, namely a region inside the contour, call it R_u, and a region outside the contour, $\Omega \backslash R_u$. Measurements u_j (resp. v_j) for region R_u (resp. R_v), are given by

$$u_j = \frac{\int_{R_u} G_j(I(\boldsymbol{x})) d\boldsymbol{x}}{|R_u|}, \quad v_j = \frac{\int_{\Omega \backslash R_u} G_j(I(\boldsymbol{x})) d\boldsymbol{x}}{|R_v|}, \quad (5)$$

with $|R_u| = \int_{R_u} d\boldsymbol{x}$, ($|R_v| = \int_{\Omega \backslash R_u} d\boldsymbol{x}$), $\boldsymbol{x} = (x, y)$.

Exploiting the approximation in Eq. (4), we define an energy functional whose optimization yields the evolution of our active contour based on a mutual information measure defined in Eq. (2). The new energy functional for two regions, is

$$JS_{a,2} = H(a_1 \hat{p}_1(\xi) + a_2 \hat{p}_2(\xi)) - a_1 H(\hat{p}_1(\xi)) - a_2 H(\hat{p}_2(\xi)). \quad (6)$$

Using (4) in Eq.(6), we have

$$\begin{aligned} \widehat{JS}_{a,2} &= H(\nu) - \frac{1}{2}\sum_{j=1}^m (a_1 u_j + a_2 v_j)^2 \\ &\quad - a_1\left(H(\nu) - \frac{1}{2}\sum_j u_j^2\right) - a_2\left(H(\nu) - \frac{1}{2}\sum_j v_j^2\right) \\ &= \frac{1}{2} a_1 a_2 \sum_{j=1}^m (u_j - v_j)^2 \end{aligned} \quad (7)$$

by noting that $a_1 + a_2 = 1$.

The image domain will be partitioned into "distinct" regions whose characteristics are well-separated as long as the measurements in a region successfully capture the statistical characteristics of the data. For instance, when images can be approximated by constants, the mean intensity over each region is an adequate measurement, i.e., the first moment of the r.v. is utilized. ($G(x) = x$). For Gaussian densities, a second order moment, the variance, is utilized. In order to capture statistical characteristics for non-Gaussian densities, we use the functions per [14]: $G_1(\xi) = \xi e^{-\xi^2/2}$ as an odd function to measure asymmetry (analogous to the third moment as a measure of skewness), and $G_2(\xi) = |\xi|$, or $e^{-\xi^2/2}$ as choices of even functions (analogous to the fourth moment as a measure of sparsity, bimodality, relative to a Gaussian density).

The choice of prior probabilities for each density, i.e. the coefficients a_1, a_2 in Eq.(7) leads to different gradient descent flows for a contour. If equal priors are assigned ($a_1 = a_2 = 0.5$), the gradient flow for a continuous contour $\boldsymbol{C}$ is found as

$$\frac{\partial \boldsymbol{C}}{\partial t} = \underbrace{\sum_{j=1}^m (u_j - v_j)\left(\frac{G_j(I) - u_j}{|R_u|} + \frac{G_j(I) - v_j}{|R_v|}\right)}_{f} \boldsymbol{N}_u, \quad (8)$$

where f is the speed of the contour. We note that this is a generalized form of the data term of the flow in [9].

In an attempt to normalize the effective contribution to the divergence measure from a region with respect to the ratio of the number of pixels in its population to the overall number of pixels, we assign priors that are proportional to the ratio of area of each region to the total area of the image domain. In this case, we arrive at the following gradient descent flow

$$\frac{\partial C}{\partial t} = \underbrace{\sum_{j=1}^{m}(u_j - v_j)\,((G_j(I) - u_j) + (G_j(I) - v_j))}_{f}N_u \quad (9)$$

whose energy functional is indeed a generalized form of the energy functional in [8]. For case of evolving polygon vertices, each vertex flow is given by substituting the total image force f into Eq.(1) and solving the ODE's simultaneously.

Note that to avoid degeneracies, such as edges intersecting each other, we developed a regularizer which balances the tendency of edges getting too close in order to better satisfy the optimization problem. The issue of the regularization, the choice of initial number of vertices, and adaptive removal and addition of vertices are detailed in [13].

Two sample simulation results that demonstrate the evolution of a single polygon over an image domain are given in Fig.1 and Fig.2. (The progression of the flow is left-right, top-bottom.) In Fig.1, the polygon vertex locations marked on the region enclosed by them are also shown at the same snapshot moments. Note the significant compression of the image content in both cases to merely a few vertices.

4. EVOLUTION OF MULTIPLE CONTOURS

The generalization of the above development is also readily achieved as we show next. Suppose there are two active contours whose inner regions are denoted by R_u, and R_v, and their common exterior by R_w. The measurements, i.e. statistics, in these respective regions are denoted by $u_j, v_j, w_j,\ j = 1, \ldots, m$ for m different measurements, with respective prior probabilities $a_1, a_2, a_3 \mid a_1 + a_2 + a_3 = 1$. The energy functional for three densities can be written as

$$JS_{a,3} = H\left(\sum_{i=1}^{3} a_i p_i(\xi)\right) - \sum_{i=1}^{3} a_i H(p_i(\xi)) \quad \approx \frac{1}{2}\sum_{j=1}^{m}[\; a_1\, a_2(u_j - v_j)^2 + a_1\, a_3(u_j - w_j)^2 + a_2\, a_3(v_j - w_j)^2] \quad (10)$$

This form of energy functional may easily be extrapolated to N active contours, thus partitioning the image domain into $N + 1$ regions. Observe that the first-order approximations to both the densities and the corresponding entropies of the regions, lead to a measure that computes a weighted sum of the mutual information measure (i.e. distances between their statistics) between all pairwise combinations of the regions.

The ternary case entails the derivation of a gradient descent flow for each of the two active contours C_u, and C_v. Taking the first variation of the energy functional in Eq.(10) w.r.t. contour C_u and C_v, yields a gradient descent flow for each contour (here we assumed constant priors), and the image forces are given by

$$\begin{aligned} f_u = &\sum_{j=1}^{m}[a_1\, a_2(u_j - v_j) + a_1\, a_3(u_j - w_j)]\frac{G_j(I) - u_j}{|R_u|} \\ &+[a_1\, a_3(u_j - w_j) + a_2\, a_3(v_j - w_j)](1-\chi_v)\frac{G_j(I) - w_j}{|R_w|}, \\ f_v = &\sum_{j=1}^{m}[-a_1\, a_2(u_j - v_j) + a_2\, a_3(v_j - w_j)]\frac{G_j(I) - v_j}{|R_v|} \\ &+[a_1\, a_3(u_j - w_j) + a_2\, a_3(v_j - w_j)](1-\chi_u)\frac{G_j(I) - w_j}{|R_w|}. \end{aligned} \quad (11)$$

where χ_u, and χ_v are the characteristic functions over the regions R_u, and R_v.

We illustrate the use of ternary evolutions in Fig.3, and Fig.4. (The progression of the flow is left-right, top-bottom.) Two contours move in such a way to maximize the approximate Jensen-Shannon divergence between densities of the three regions, namely R_u: inside the contour C_u; R_v: inside the contour C_v; R_w: the complement of $R_u \bigcup R_v$. The resulting two polygons show that the gain is again two-fold: segmentation of the target objects, and their description in terms of a small number of vertices.

5. CONCLUSIONS

We proposed novel flows which capture polygonal objects in images. We proposed the application of a Jensen-Shannon criterion as a distance among the regions delineated by polygons, and utilized an approximate implementation of the criterion. The result of our methodology provides a significant compression of image content for a variety of purposes such as image retrieval by shape templates, shape coding, content-based description of the multimedia data which are the core functionalities of MPEG-7.

6. REFERENCES

[1] K.J. O'Connell, "Object-adaptive vertex-based shape coding method," *IEEE Trans. Circuits and Systems on Video Technology*, vol. 7, pp. 251–255, 1997.

[2] B.-J. Yun, S.-W. Lee, and S.-D. Kim, "Vertex adjustment method using geometric constraint for polygon-based shape coding," *Electronics Letters*, vol. 37, pp. 754–755, 2001.

[3] "Special issue on MPEG-7," *IEEE Trans. Circuits and Systems on Video Technology*, June 2001.

[4] R. Ronfard, "Region-based strategies for active contour models," in *IJCV*, 1994, vol. 13, pp. 229–251.

[5] R. Malladi, J.A. Sethian, and B.C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE PAMI*, vol. 17, no. 2, pp. 158–173, 1995.

[6] N. Paragios and R. Deriche, "Geodesic active regions for texture segmentation," Tech. Rep., INRIA, France, 1998.

[7] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE PAMI*, vol. 18, pp. 884–900, 1996.

[8] T.F. Chan and L.A. Vese, "An active contour model without edges," in *ScaleSpace99*, 1999, pp. 141–151.

[9] A. Yezzi, Jr., A. Tsai, and A. Willsky, "A statistical approach to snakes for bimodal and trimodal imagery," in *ICCV*, 1999.

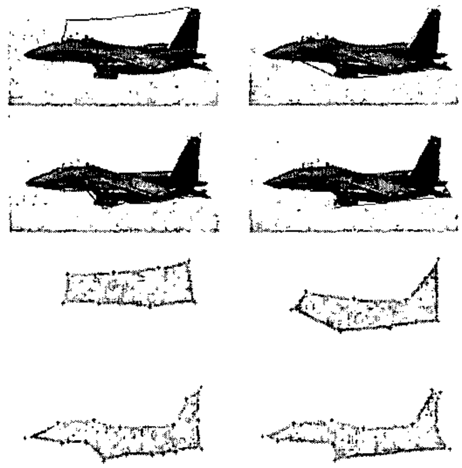

Fig. 1. Plane figure features are captured finally by 12 vertices of the polygon (bottom right picture) by flow (1) with image force f in (9) which uses only mean, i.e. $G(x) = x$ $(m = 1)$.

Fig. 2. Monarch's natural texture captured by 5 vertices of the polygon using flow (1) with image force f in (9), where $G_1(x) = xe^{-x^2/2}$, and $G_2(x) = e^{-x^2/2}$ $(m = 2$ measurements). (Polygon overlaid on the image can be seen in color on the pdf file.)

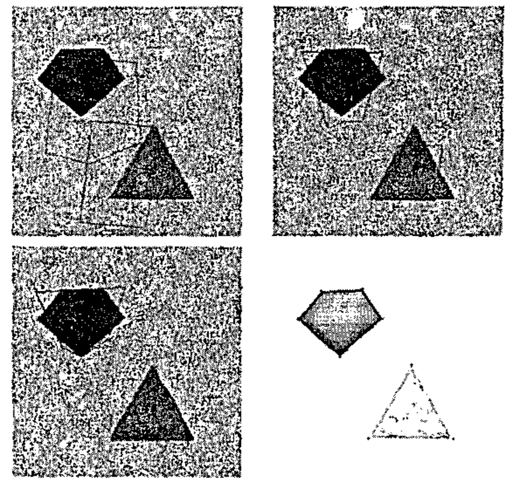

Fig. 3. Ternary flows using image forces (11) with $G(x) = x$, are used to segment this three channel (RGB) image. Result of the algorithm shown on the bottom right provides only 5 vertex locations for the first object, and 3 vertex locations for the second object.

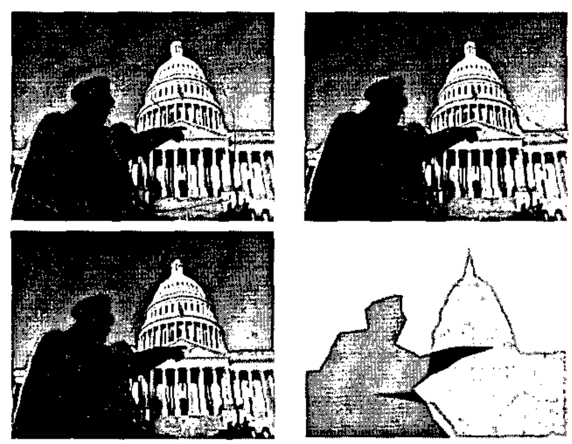

Fig. 4. Three distinct regions: sky, the capitol building, and the policeman are captured by ternary flows using (11) where $G_1(x) = xe^{-x^2/2}$, and $G_2(x) = e^{-x^2/2}$. (Polygons overlaid on the image can be seen in color on the pdf file.)

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. 1, pp. 321–331, 1988.

[11] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Trans. Information Theory*, vol. 37, pp. 145–151, 1991.

[12] A.B. Hamza, Y. He, and H. Krim, "An information divergence measure for ISAR image registration," in *IEEE Workshop on Statistical Signal Processing*, 2001, pp. 130–133.

[13] Gozde Bozkurt Unal, *Curve and Polygon Evolution Techniques for Image Processing*, Ph.D. thesis, NCSU, 2002.

[14] A. Hyvarinen, "New approximations of differential entropy for independent component analysis and projection pursuit," Tech. Rep., Helsinki Univ., 1997.

[15] E.T.Jaynes, *Information Theory and Statistical Mechanics*, p. 181, K.Ford(ed.), New York, 1963.

[16] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, 1991.