A HIGH DIMENSIONAL FRAMEWORK FOR JOINT COLOR-SPATIAL SEGMENTATION

Sylvain Boltz, Éric Debreuve, Michel Barlaud

Laboratoire I3S, Université de Nice - Sophia Antipolis
2000 route des Lucioles, 06903 Sophia Antipolis, FRANCE
{boltz,debreuve,barlaud}@i3s.unice.fr

ABSTRACT

This paper deals with region-of-interest (ROI) segmentation in video sequences. The goal is to determine in successive frames the region which best matches, in terms of a similarity measure, a ROI defined in a reference frame. Color and geometry can be combined in a joint PDF. However such high-dimensional PDFs being hard to estimate, measures based on PDF distances may lead to incorrect segmentations. Here, we propose to use an estimate of the Kullback-Leibler divergence adapted to high-dimensional PDFs. It is defined from the samples using the kth-nearest neighbor (kNN) framework and it is differentiated for active contour implementation and expressed in both the continuous form and a kNN form. Results are presented on standard sequences.

***Index Terms*—** Segmentation, similarity measure, multivariate distributions, multimodal distributions, kNN, active contour

1. INTRODUCTION

The goal of region-of-interest (ROI) segmentation is to determine in successive frames the region which best matches, in terms of a similarity measure, an ROI (user-)defined in a reference frame. A similarity measure is a distance between two data sets, the reference data set and a candidate, or target, data set. In the discrete framework, each data set is composed of one data vector per pixel of the region. If the similarity measure does not necessarily make use of a one-to-one match between the pixels of the regions.

Two aspects of similarity measures between the reference region and a target region can be distinguished: radiometry, which indicates if the regions have similar color distributions, and geometry, which correlates where these colors are present in each regions. Similarity measures based solely on radiometry include distances between color histograms or probability density functions (PDF). For example, the Bhattacharya distance was used in tracking [1]. However, the absence of geometric information implies that several candidate regions can appear as good matches.

As an alternative, geometry can be added by means of a motion model used to compute the pointwise residual between reference and candidate regions. A function of the residual can serve as a similarity measure, classically, the sum of squared differences (SSD), functions used in robust estimation such as the sum of absolute differences (SAD), or statistic measures [2].

The geometric constraint can be soften by adding geometry to the PDF-based approach, *i.e.*, by defining a joint geometric/radiometric PDF [3]. This later approach leads to high-dimensional PDFs. Although there are efficient [4] methods to estimate multivariate PDFs using Parzen windowing, limitations appear when the dimension of the domain of definition of the PDFs increases. This is described in [4] as the curse of dimensionality: as the dimension of the data space increases, the space sampling gets sparser. Dilating the Parzen window is not a satisfying solution since it implies over-smoothing of the PDFs. Consequently, PDF-based similarity measures might not be estimated accurately enough for segmentation. To overcome this high-dimension problem, a PDF estimate based on a k-th nearest neighbor (kNN) search was proposed [4, 5, 2] and used to define a consistent entropy estimate [6].

Segmentation using color distributions [7] or pointwise residual [8] has already been considered in the litterature. However, these two methods have limitations, recents methods try to combine multiple features (spatial data, gradient, wavelets, motion) to perform accurate matching [3] or segmentation [9]. The PDFs are then high-dimensionals and some assumptions have to be made (*e.g.*: independence between components, gaussian assumptions).

This paper has two main contributions in segmentation. First we propose to compute the Kullback-Leibler divergence, or distance, between high-dimensional PDFs using the kNN framework. This new estimate is efficient for high-dimensional PDFs [6, 5] with very weak assumptions on the PDFs. Second, we propose to apply in segmentation, a joint radiometric/geometric called feature-spatial [3, 10], originally proposed in bounding box tracking. The equations for a complete, kNN-based active contour implementation are provided.

This paper is organized as follows. Section 2 defines the Kullback-Leibler distance on geometric/radiometric data, it presents state-of-the-art methods to estimate it, the Ahmad-Lin entropy estimation, the Parzen windowing method, and the limitations when combining both. Section 4 presents the kNN approach and the kNN-based expression of the Kullback-Leibler distance. Section 3 plugs this distance in a segmentation method through active contours. Sections 5 provide some results of segmentation performed on two standard sequences. Finally Section 6 concludes and gives some perspectives.

2. PROBLEM STATEMENT

2.1. Similarity measures with a soft geometric constraint

Let I_{ref} and I_{target} be, respectively, the reference frame in which the ROI Ω_R is (user-)defined and the candidate, or target, frame in which the region Ω_T best matches the ROI, in terms of a given similarity measure, is to be searched for. This search amounts to finding the region Ω_T which minimizes

$$J(\Omega_T) = D(I_{\mathrm{ref}}(\Omega_R), I_{\mathrm{target}}(\Omega_T)) \tag{1}$$

where D is a similarity measure, or distance, between the two sets of data. Ω_T and Ω_R are subsets of $\mathbb{R}^2$ (or subsets of $\mathbb{N}^2$ in the discrete framework).

For clarity, the reference data set $I_{\mathrm{ref}}(\Omega_R)$ will be denoted by R and the target data set $I_{\mathrm{target}}(\Omega_T)$ will be denoted by T. Thus,

$R(x)$ resp. $T(x)$, $x \in \Omega_R$ resp. $x \in \Omega_T$, represent corresponding samples from their respective region. Traditionally, $R(x)$ and $T(x)$ are a triplet of color components in a given color space, *e.g.*, RGB or YUV.

Two aspects of similarity measures can be distinguished: radiometry which indicates if the regions have similar colors and geometry which correlates where these colors are present in the regions. Measures based solely on radiometry include distances between the probability density functions (PDF) of the color information in the regions.

A widely used distance is the Kullback-Leibler divergence, or distance,

$$D_{\mathrm{KL}}(T,R) = \int_{\mathbb{R}^d} f_T(\alpha)\ \log \frac{f_T(\alpha)}{f_R(\alpha)}\ \mathrm{d}\alpha \tag{2}$$

$$= -H(f_T) + H_\times(f_T, f_R) \tag{3}$$

where f_T is the PDF of data set T, f_R is the PDF of data set R, H is the Shannon entropy and $H_\times$ is the cross entropy, also called relative entropy or likelihood.

The geometric constraint can be soften by expressing it in the PDF-based approach, *i.e.*, by adding geometry to the original radiometric data [3]. Formally, the PDF $f_R(\alpha)$ resp. $f_T(\alpha)$ is built on $\alpha = T(x) = \{I_{\mathrm{target}}(x), x\}$ for $x \in \Omega_T$ resp. on $\alpha = R(x) = \{I_{\mathrm{ref}}(x), x\}$ for $x \in \Omega_{\mathrm{ref}}$. The PDF is now built on a joint geometric/radiometric data set of the reference region and the target region: R(x) and T(x) are now 5-dimensionals (three color components and two spatial components).

2.2. Spatial features registration

Although the geometric constraint is soft, we can improve the matching by aligning the spatial features from the reference frame to the target frame. The spatial features of R are registered on the spatial features of T by transforming the spatial features of R: $R(x) = \{I_{target}(x), x + \varphi(x)\}$ where φ is the transformation of the object from the frame R to frame T. For coherence with the segmentation, the transformation is estimated with the same energy

$$\varphi = \arg\min_{\varphi} D_{\mathrm{KL}}(f_R, f_T). \tag{4}$$

R and T are now data sets $\{I_{\mathrm{ref}}(x), x + \varphi(x)\}$, $x \in \Omega_R$, and $\{I_{\mathrm{target}}(x)), x\}$, $x \in \Omega_T$, respectively, where φ is a geometric transformation representing the motion of the ROI between the reference frame and the target frame. Estimation (19) being provided in the discrete framework, it is not differentiable. Its minimization could be performed by an exhaustive search procedure in a search window. For computational considerations, it will be performed using a suboptimal search procedure, the diamond search.

3. SEGMENTATION USING ACTIVE CONTOURS

3.1. Ahmad-Lin estimate of entropy and Parzen windowing

Ω_U defining dataset $U = \{I(x), x\}$ for x in Ω_U (U being either T or R, Ω_U being either Ω_T or Ω_R) has the following Shannon entropy

$$H(U) = -\int_{\mathbb{R}^d} f_U(\alpha)\ \log f_U(\alpha)\ \mathrm{d}\alpha \tag{5}$$

It can be rewritten from the Ahmad-Lin estimate [11]

$$\hat{H}_{\mathrm{AL}}(f_U) = -\frac{1}{|U|}\int_{\Omega_U} \log f_U(U(x))\, dx \tag{6}$$

where $|U|$ is the area of Ω_U. Since the actual PDF f_U is unknown, it must be estimated. A common practice is to use the non-parametric, Parzen windowing method. The Parzen method for PDF estimation makes no assumption about the actual PDF. Consequently, the estimated PDF cannot be described in terms of a small number of parameters, as opposed to, say, a Gaussian distribution defined by its mean and variance. This method is therefore qualified as non-parametric. It approximates the density at sample s with the relative number of samples $k(s)/|U|$ falling into the open ball of volume v centered on s

$$\hat{f}_U(s) = \frac{1}{|U|}\sum_{x\in\Omega_U} K_\sigma(s - U(x)). \tag{7}$$

3.2. Shape derivative

The energy to be minimized through active contours is the kullback leibler distance (19). In addition, as the distribution of the object can be characterized by a subregion inside the object, we propose to add a maximum area constraint with a weighting λ

$$J(T) = D_{\mathrm{KL}}(T,R) - \lambda|T|. \tag{8}$$

Using (6), this energy can be rewritten as follows

$$J(T) = \frac{1}{|T|}\int_{\Omega_T} \log f_T(T(x))\, dx \tag{9}$$

$$- \frac{1}{|T|}\int_{\Omega_T} \log f_R(T(x))\, dx - \lambda|T|, \tag{10}$$

showing the dependencies with respect to the domain Ω_T. Therefore, we propose to rely on the shape derivative framework [7] to determine the derivative of (9) with respect to Ω_T in the direction V. Neglecting the shape derivative of the distribution f_T, we obtain

$$\mathrm{d}J_r(T,V) = \int_{\partial\Omega_T} \mathcal{A}(s)V(s)\cdot N(s)\mathrm{d}s \tag{11}$$

where N is the inward unit normal to $\partial\Omega_T$ and with

$$\mathcal{A}(s) = -\frac{1}{|T|}(D_{\mathrm{KL}}(T,R) - \log f_T(s) + \log f_R(s)) + \frac{1}{|T|^2}\int_{\Omega_T}\left(1 - \frac{K_\sigma(T(x)-T(s))}{f_T(T(x))}\right)\mathrm{d}x + \lambda. \tag{12}$$

The minimization of the energy is then performed using the active contour technique [8, 7] using a B-spline implementation. An initial contour is iteratively deformed according to V chosen such that derivative is negative or equal to zero at each iteration. The minimum is reached when the derivative is equal to zero. The corresponding shape of the active contour represents the segmentation. To ensure the negativity of the shape derivative, we choose

$$V(s) = -\mathcal{A}(s)\ N(s). \tag{13}$$

Let us remind that the PDF is built on a joint geometric/radiometric data set of the reference region and the target region: R(x) and T(x) are now 5-dimensional (three color components and two spatial components). Consequently, the PDF domain of definition becomes high-dimensional. The sparsity of this high-dimension data space makes the PDF estimation, and therefore the similarity measure estimation, even more problematic. Let us present a new framework for computation high dimensional PDFs and the Kullback-Leibler distance.

4. THE K-TH NEAREST NEIGHBOR (KNN) FRAMEWORK

The first difficulty in using the Parzen method (see (7)) is the critical choice of the window size σ [4]. Another difficulty is due to what is informally called the curse of dimensionality. As the dimension of the data space increases, the space sampling gets sparser. Therefore, less samples fall into the Parzen windows centered on each sample, making the PDF estimation less reliable. Dilating the Parzen window does not solve this problem since it leads to over-smoothing the PDF. In a way, the limitations of the Parzen method come from the fixed size of the window: the method cannot adapt to the local sample density.

The k-th nearest neighbor (kNN) framework provides an advantageous alternative. It allows to estimate the entropy of a PDF directly from the data set, *i.e.*, without explicitly estimating the PDF. Nevertheless, this entropy estimate derives from the kNN PDF estimation method [4], p. 181. This method shows good performance for multivariate data [5].

4.1. Locally adaptive PDF estimation

In the Parzen method, the density of U at sample s is related to the number of samples falling into a window of fixed size σ centered on the sample (see Eq. (7)). The kNN method is the *dual* approach: the density is related to the size of the window σ necessary to include its k nearest neighbors. Let us note $\sigma(s) = \rho_k(U, s)$ the distance to the k-th nearest neighbor of s among the data set U.

This variable size estimate is called locally adaptive. It can be performed it two different ways. The first way is called balloon estimate [5]

$$\hat{f}_U(s) = \frac{1}{|U|} \sum_{x \in \Omega_U} K_{\sigma(s)}(s - U(x)) \tag{14}$$

$$\stackrel{u.k.}{=} \frac{k}{v_d \, |U| \, \rho_k^d(U, s)} \tag{15}$$

where v_d is the volume of the unit ball in $\mathbb{R}^d$. The kernel size $\sigma(s) = \rho_k(U, s)$ depends on the point s where $\hat{f}$ is evaluated. Expression (14) can be reduced to the simple expression (15) when $K_{\sigma(s)}$ is a uniform kernel. The second way is called sample point estimate [5]

$$\hat{f}_U(s) = \frac{1}{|U|} \sum_{x \in \Omega_U} K_{\sigma(U(x))}(s - U(x)). \tag{16}$$

The kernel sizes $\sigma(U(x)) = \rho_k(U, U(x))$ depend on the samples $U(x)$.

We will consider the balloon estimate as it is the underlying PDF estimate in the kNN expression of entropy. However, we will discuss later how the sample point estimate can be useful. Although the distance is usually computed in the Euclidean sense, other distances can be used. Let us remind that the data are a subset of $\mathbb{R}^d$ with $d = 5$. The choice of k appears to be a much less critical than the choice of the window size in the Parzen method. Actually, when the kNN approach is used for parameter estimation, k must be greater than the number of parameters and such that $k/|U|$ tends toward zero when both k and $|U|$ tends toward infinity. A typical choice is $k = \sqrt{|U|}$.

4.2. Kullback-Leibler Distance estimation

Based on the Ahmad-Lin entropy estimate (6) and the kNN PDF estimation (14), a consistent and unbiased entropy estimate was proposed with a proof of consistency under weak conditions on the underlying PDF [6]. The kNN estimate of the Shannon entropy is equal to

$$\hat{H}_{\mathrm{kNN}}(f_T) = \log(v_d \, (|T| - 1)) - \psi(k) + d \, \mu_T(\log \rho_k(T)) \tag{17}$$

where $\mu_T(g)$ is the mean of g for all the values taken over data set T where ψ is the Polygamma function Γ'/Γ. Note that estimate (17) does not depend on the PDF $\hat{f}_T$. Informally, the main term in estimate (17) is the mean of the log-distances to the k-th nearest neighbor of each sample.

In the same framework, the cross entropy of two data sets R and T can be approximated by

$$\hat{H}_{\times,\mathrm{kNN}}(f_T, f_R) = \log(v_d \, |R|) - \psi(k) + d \, \mu_T(\log \rho_k(R)). \tag{18}$$

Note again that estimate (18) does not depend on any PDF and that its main term is the mean of the log-distances to the k-th nearest neighbor among data set R of each sample of T. Finally, since the Kullback-Leibler distance is a difference between a cross entropy and a Shannon entropy (see Eq. (2)), the kNN estimate of this distance is equal to

$$D_{\mathrm{KL}}(T, R) = \log \frac{|R|}{|T| - 1} + d[\mu_T(\log \rho_k(R)) - \mu_T(\log \rho_k(T))] \tag{19}$$

4.3. kNN-based shape derivative

Using the expression (15) for f_R and f_T and the expression (19) for D_{KL}, the shape derivative (12) is equal to

$$\begin{aligned} \mathcal{A}(s) = & -\frac{d}{|T|}[\mu_T(\log \rho_k(R)) - \mu_T(\log \rho_k(T)) \\ & - \log \rho_k(R, s) + \log \rho_k(T, s)] + \lambda \\ & + \frac{1}{|T|}\left[1 - \frac{1}{k} \sum_{x \in \nu(T,s)} \left(\frac{\rho_k(T, x)}{\rho_k(T, s)}\right)^d\right] \end{aligned} \tag{20}$$

where $\nu(T, s)$ is the support of $K_{\sigma(T(s))}$, the uniform kernel centered at $T(s)$ of half width $\rho_k(T, s)$ (see Eq. (15)). By definition, there are exactly k samples in $\nu(T, s)$.

Keeping expression (15) for f_R and f_T but replacing K_σ with $K_{\sigma(T(x))}$ (the approach of (16)), the sum in (20) turns into the cardinality of $\{x | s \in \nu(T, x)\}$ since $\rho_k(T, s)$ at the denominator becomes $\rho_k(T, x)$.

5. EXPERIMENTAL RESULTS

In this section we will compare two methods, the Kullback-Leibler distance computed through kNN but with no geometry kNN-KL (no spatial features, R and T are 3-Dimensionals) and the Kullback-Leibler distance computed through kNN with geometry kNN-KL-G (spatial features, R and T are 5-Dimensionals). k of the kNN framework is set to 3. The reference histograms for kNN-KL and kNN-KL-G are built over a region Ω_R on frame 1 for ``Erik'', Fig. 1, and frame 74 for ``Football'', Fig. 1, using a manual segmentation. The goal is to find the corresponding region Ω_T in frame 6 for ``Erik'' and frame 75 for ``Football''. We

initialize with a circle far from the solution to show the stability of the method.

First, we present results on sequence ``Erik'' composed of 288x352-pixel frames (see Fig. 1). This sequence shows a translating man over a static background. This sequence was chosen because its motion is very simple, while it is composed of many colors which will lead to complex color histograms. Some parts of the background have similar colors than Erik. Therefore kNN-KL includes it as object while kNN-KL-G detects their spatial features are not correct so it does not include it as object. These results did not use maximum area constraint, $\lambda = 0$. Results are presented on sequence ``Football'' composed of 288x352-pixel frames (see Fig. 1). This sequence shows fast and articulate motions. Some parts of the public on the upper part of the video have the same colors as the player. kNN-KL-G excludes again them as their spatial features are not correct while kNN-KL includes them in the segmentation. The maximum area constraint was tuned to segment the whole object $\lambda = 0.004$ in both cases. The Kullback-Leibler distance kNN-KL-G slightly increase when taking the legs of the player as their are articulated (error of registration in the spatial features). However, as the geometric constraint is soft, it increases less than with segmenting the public, the player is then correctly segmented with the help of maximum area constraint.

6. CONCLUSION AND FUTURE WORKS

The presented results show that the kNN framework can be applied to active contour segmentation, especially in a high-dimensional context such as joint feature-spatial segmentation using a PDF-based criterion. Future works will consider a motion model more complex than translation, possibly dealing with articulated motions. We will also apply this high-dimensional framework for segmentation using multiple features [9] (motion, texture...).

Fig. 1. Segmentation on sequence "Erik", resp. "Football", on frame 6, resp. 75: *(from left to right and top to bottom)* region of interest Ω_R manually segmented on frame 1, resp. 74; initialization of the segmentation; result Ω_T of segmentation with method kNN-KL; result Ω_T of segmentation with method kNN-KL-G.

7. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Hilton Head Island, SC, USA, 2000.

[2] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, and L. Pronzato, "A minimum-entropy procedure for robust motion estimation," in *IEEE Internatinal Conference on Image Processing (ICIP)*, Atlanta, GA, USA, 2006.

[3] A. Elgammal, R. Duraiswami, and L. S. Davis, "Probabilistic tracking in joint feature-spatial spaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, WI, USA, 2003.

[4] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, 1992.

[5] G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *The Annals of Statistics*, vol. 20, no. 3, pp. 1236–1265, 1992.

[6] M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," *Journal of Nonparametric Statistics*, vol. 17, no. 3, pp. 277–297, 2005.

[7] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients?," *SIAM Journal of Applied Mathematics*, vol. 1, no. 2, pp. 2128–2145, 2003.

[8] D. Cremers and S. Soatto, "Motion competition: A variational framework for piecewise parametric motion segmentation," *International Journal of Computer Vision*, vol. 62, no. 3, pp. 249–265, 2005.

[9] T. Brox, M. Rousson, R. Deriche, and J. Weickert, "Unsupervised segmentation incorporating colour, texture, and motion," in *Computer Analysis of Images and Patterns*, Groningen, The Netherlands, 2003, vol. 2756 of *LNCS*, pp. 353–360.

[10] S. Boltz, E. Debreuve, and M. Barlaud, "High-dimensional statistical distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN, USA, 2007.

[11] I. Ahmad and P.-E. Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 372–375, 1976.