# High-dimensional statistical distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry

Sylvain Boltz Eric Debreuve Michel Barlaud
Laboratoire I3S, University of Nice-Sophia Antipolis, CNRS, France
{boltz,debreuve,barlaud}@i3s.unice.fr

## Abstract

*This paper deals with region-of-interest (ROI) tracking in video sequences. The goal is to determine in successive frames the region which best matches, in terms of a similarity measure, an ROI defined in a reference frame. Two aspects of a similarity measure between a reference region and a candidate region can be distinguished: radiometry which checks if the regions have similar colors and geometry which checks if these colors appear at the same location in the regions. Measures based solely on radiometry include distances between probability density functions (PDF) of color. The absence of geometric constraint increases the number of potential matches. A soft geometric constraint can be added to a PDF-based measure by enriching the color information with location, thus increasing the dimension of the domain of definition of the PDFs. However, high-dimensional PDF estimation is not trivial. Instead, we propose to compute the Kullback-Leibler distance between high-dimensional PDFs without explicitly estimating the PDFs. The distance is expressed directly from the samples using the $k$-th nearest neighbor framework. Tracking experiments were performed on several standard sequences.*

## 1. Introduction

The goal of region-of-interest (ROI) tracking 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.

Two aspects of a similarity measure between a reference region and a target region can be distinguished: radiometry which checks if the regions have similar colors and geometry which checks if these colors appear at the same location in the regions. Similarity measures based solely on radiometry include distances between color histograms or probability density functions (PDF). For example, the Bhattacharya distance was used for tracking [6, 13] and mutual information [17] was used for registration. The Kullback-Leibler divergence (or, informally, distance) and the Hellinger distance have also been considered in various applications [7]. However, the absence of geometric constraint 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 a 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) or functions used in robust estimation [10] such as the sum of absolute differences (SAD). The geometric constraint being strictly defined by the motion model, these measures might be less efficient if the model is not coherent with the actual motion. Indeed, it might generate too many outliers in the residual, including in the framework of robust estimation. Moreover, even if the model is globally coherent with the actual motion, the choice of the function of the residual is implicitly linked to an assumption on the PDF of the residual, *e.g.*, Gaussian for SSD or Laplacian for SAD. This might not be valid in case of occlusion for example.

The geometric constraint can be soften, *e.g.*, by cascading a strict geometry approach and a radiometric approach [16] or by adding geometry to a PDF-based approach, *i.e.*, by defining a joint radiometric/geometric PDF [7]. This later approach leads to high-dimensional PDFs. Although there are efficient [14] and fast [19] 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 [14] 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 tracking. To overcome this difficulty, a PDF estimator based on a $k$-th nearest neighbor (kNN) search was proposed [8]. This approach was then used to define a consistent entropy estimator [3, 9, 11].

1-4244-1180-7/07/$25.00 ©2007 IEEE

In this paper, we propose to compute the Kullback-Leibler distance between high-dimensional PDFs using the kNN framework. In this context, explicit estimation of the PDFs will not be necessary since the distance is expressed directly from the samples. This estimator being well-adapted to high-dimension, it can be applied to extended radiometric/geometric data. Section 2 provides some notations and general comments. Then, Section 3 reminds 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 proposed for ROI tracking. Finally, Sections 5 and 6 provide some results and comments of tracking performed on several standard sequences.

## 2. Similarity measures for tracking

### 2.1. Generalities

Let $I_{\mathrm{ref}}$ and $I_{\mathrm{tgt}}$ be, respectively, the reference frame in which the ROI is (user-)defined and the target frame in which the region which best matches the ROI, in terms of a given similarity measure, is to be searched for. This search amounts to finding the geometric transformation $\hat{\varphi}$ such that

$$\hat{\varphi} = \arg\min_{\varphi} D(I_{\mathrm{ref}}(\Omega), I_{\mathrm{tgt}}(\varphi(\Omega))) \tag{1}$$

where $D$ is a similarity measure, or distance, between two sets of data and $\Omega$ is the domain of the ROI. Domain $\Omega$ is a subset of $\mathbb{R}^2$ (or a subset of $\mathbb{N}^2$ in the discrete framework).

For clarity, the reference data set $I_{\mathrm{ref}}(\Omega)$ will be denoted by $R$ and the target data set $I_{\mathrm{tgt}}(\varphi(\Omega))$ will be denoted by $T$, the geometric transformation being made implicit. Thus, $R(x)$ and $T(x)$, $x \in \Omega$, represent corresponding samples from their respective region. Traditionally, $R(x)$ is a triplet of color components in a given color space, *e.g.*, RGB or YUV.

As noted earlier, two aspects of similarity measures can be distinguished: radiometry which checks if the regions have similar colors and geometry which checks if these colors appear at the same location in the regions. Sections 2.2, 2.3, and 2.4 focus on how geometry is involved.

### 2.2. Geometry-free similarity measures

The similarity measure can be based solely on radiometry. Classically, $D$ can be a distance between color histograms or, similarly, PDFs. The knowledge of where a given color was present within the region is lost. For example, the Bhattacharya distance was used for tracking [6, 13]

$$D_{\mathrm{BHA}}(T, R) = \int_{\mathbb{R}^d} \sqrt{f_R(s)\ f_T(s)}\ \mathrm{d}s \tag{2}$$

where $f_G$ is the PDF of $G(x)$, $x \in \Omega$, and $d$ is the number of components of $R(x)$, *i.e.*, three if all color components are used.

Another widely used similarity measure is the Kullback-Leibler distance

$$D_{\mathrm{KL}}(T, R) = \int_{\mathbb{R}^d} f_T(s)\ \log \frac{f_T(s)}{f_R(s)}\ \mathrm{d}s. \tag{3}$$

Note that this "distance" is not symmetric, *i.e.*, $D_{\mathrm{KL}}(T, R)$ is not equal to $D_{\mathrm{KL}}(T, R)$. Distance (3) can be decomposed as follows

$$\begin{aligned} D_{\mathrm{KL}}(T, R) &= \int_{\mathbb{R}^d} f_T(s)\ \log f_T(s)\ \mathrm{d}s \\ &\quad - \int_{\mathbb{R}^d} f_T(s)\ \log f_R(s)\ \mathrm{d}s \qquad (4) \\ &= -H(T) + H_{\times}(T, R) \qquad (5) \end{aligned}$$

where $H$ is the differential entropy and $H_{\times}$ is the cross entropy, also called relative entropy or likelihood.

Not accounting for the knowledge of where a given color was present in the region allows to be more flexible regarding the geometric transformation between the reference region and the target region. However, it increases the number of potential matches and then the risk for the tracking to fail after a few frames. This can be avoided by using a geometry-aware similarity measure.

### 2.3. Similarity measures with strict geometry

Geometry can be added by using transformation $\varphi$, which represents a motion model, to compute the pointwise residual between the reference region and the target region. A function of the residual can serve as a similarity measure, classically in the discrete framework, SSD

$$D_{\mathrm{SSD}}(T, R) = \sum_{x \in \Omega} (T(x) - R(x))^2 \tag{6}$$

or functions used in robust estimation such as SAD or a differentiable approximation of SAD

$$D_{\mathrm{SAD}}(T, R) = \sum_{x \in \Omega} \phi(T(x) - R(x)) \tag{7}$$

where $\phi$ is a smooth approximation of the absolute value function, *e.g.*, $\phi(x) = \sqrt{x^2 + \epsilon^2} - \epsilon$ [18].

The geometric constraint being strictly defined, these similarity measures might be less efficient if the motion model $\varphi$ is not coherent with the actual motion.

### 2.4. Similarity measures with soft geometry

The geometric constraint can be soften by expressing it in the PDF-based framework presented in Section 2.2, *i.e.*,

by adding geometry to the original radiometric data [7]. Formally, the PDF $f_U$ of $U(x)$, $x \in \Omega$, is replaced with the PDF $g_{U,x}$ of $\{U(x), x\}$, $x \in \Omega$. However, as a consequence, the samples get sparser in this extended, high-dimensional space, making the PDF estimation, and therefore the similarity measure estimation, more problematic. A solution consisting in short-cutting the PDF estimation in the KullbackLeibler distance estimation (see Eq. (4)) is proposed in Section 4. Let us first study entropy estimation since entropy appears in the expression of this distance (see Eq. (5)).

## 3. Ahmad-Lin entropy estimator

### 3.1. Expression

The differential entropy of $U$ (see Eq. (5)) can be approximated by the Ahmad-Lin estimator [1]

$$\hat{H}_{\mathrm{AL}}(U) = -\frac{1}{|\Omega|} \sum_{s \in U(\Omega)} \log f_U(s) \tag{8}$$

where $|\Omega|$ is the number of samples of $\Omega$ and $U(\Omega)$ is the set of $U(x)$, $x \in \Omega$. Since the actual PDF $f_U$ is unknown, it must be estimated. A common practice is to use the Parzen windowing method.

### 3.2. Parzen windowing method and limitations

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 set of parameters, as opposed to, *e.g.*, 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 $k(s)/|\Omega|$ falling into the open ball of volume $v$ centered on $s$

$$\hat{f}_U(s) = \frac{k(s)}{v\ |\Omega|}. \tag{9}$$

For a Gaussian window, expression (9) can be rewritten as follows

$$\hat{f}_U(s) = \frac{1}{|\Omega|} \sum_{t \in U(\Omega)} K_\sigma(s - t) \tag{10}$$

where $K_\sigma$ is a multivariate Gaussian kernel with standard deviation, or bandwidth, $\sigma$.

The choice of bandwidth $\sigma$ is critical [14]. The following value has been suggested [15]

$$\sigma = 0.9\ \min(\hat{\sigma}, \hat{p}/1.34)\ |\Omega|^{-1/5} \tag{11}$$

where $\hat{\sigma}$ is the empirical standard deviation of the samples and $\hat{p}$ is the interquartile range. Unfortunately, this kind of estimation (called plug-in because the standard deviation and interquartile range of the samples are *plugged* in the bandwidth estimator) provides a value too large when the underlying PDF has several modes. More generally, the Parzen method suffers from 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 window size: the method cannot adapt to the local sample density. The $k$-th nearest neighbor (kNN) framework provides an advantageous alternative.

## 4. The $k$-th nearest neighbor (kNN) framework

The kNN framework allows to estimate the entropy of a PDF directly from the samples, *i.e.*, without explicitly estimating the PDF. Nevertheless, this entropy estimator derives from the kNN-based PDF estimation method [8], p. 268. For a better understanding of the construction of the entropy estimator, the key results of the kNN-based PDF estimation are presented first in Section 4.1.

### 4.1. 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 centered on the sample (see Eq. (9)). The kNN method is the *dual* approach: the density is related to the size of the window necessary to include the $k$ nearest neighbors of the sample

$$\hat{f}_U(s) = \frac{k}{v(s)\ |\Omega|} \tag{12}$$

where $v(s)$ is the volume of the open ball centered on sample $s$ with a radius of $\rho_k(s)$ equal to the distance to the $k$-th nearest neighbor of $s$ excluding $s$ itself. Let us remind that the samples belong to $\mathbb{R}^d$. Therefore, the volume $v(s)$ can be written as follows

$$v(s) = \rho_k^d(s)\ \underbrace{\frac{2\pi^{d/2}}{d\ \Gamma(d/2)}}_{v_d} \tag{13}$$

where $\Gamma$ is the Gamma function and $v_d$ is the volume of the unit ball in $\mathbb{R}^d$. Then, the kNN density estimate is equal to

$$\hat{f}_U(s) = \frac{k}{\rho_k^d(s)\ v_d\ |\Omega|}. \tag{14}$$

The choice of $k$ appears to be much less critical than the choice of $\sigma$ in the Parzen method. Actually, when the kNN

approach is used for parameter estimation [3] (see Section 4.4), $k$ must be greater than the number of parameters and such that $k/|\Omega|$ tends toward zero when $|\Omega|$ tends toward infinity. A typical choice is $k = \sqrt{|\Omega|}$.

### 4.2. Entropy estimation

Let $\tilde{U}(x)$ denote the joint radiometric/geometric sample $\{U(x), x\}$, $x \in \Omega$. The number of samples $|\Omega|$ will be temporarily denoted by $N_U$. The notation $\rho_k(s)$ will be replaced with $\rho_k(\tilde{V}, s)$ to be able to indicate the data set in which neighbors of $s$ are to be search for.

Based on the Ahmad-Lin entropy estimator (8) and the kNN-based PDF estimation (14), a consistent and unbiased entropy estimator was proposed for $k = 1$ [11]. This work was extended to $k > 1$ with a proof of consistency under weak conditions on the underlying PDF [9]

$$\hat{H}(\tilde{T}) \overset{\text{kNN}}{=} \frac{1}{N_{\tilde{T}}} \sum_{s \in \tilde{T}} \log \xi_k(\tilde{T}, s) \tag{15}$$

where $N_{\tilde{T}}$ is the number of samples of data set $\tilde{T}$ and

$$\xi_k(\tilde{T}, s) = (N_{\tilde{T}} - 1)\ \exp -\psi(k)\ v_d\ \rho_k^d(\tilde{T}, s) \tag{16}$$

where $\psi$ is the Polygamma function $\Gamma'/\Gamma$. Note that estimator (15) does not depend on the PDF $\hat{f}_{\tilde{T}}$. Replacing $\xi_k$ in (15) by its expression (16), the kNN-based estimate of entropy is equal to

$$\hat{H}(\tilde{T}) \overset{\text{kNN}}{=} \log(v_d\ (N_{\tilde{T}}–1)) - \psi(k) + d\ \mu_{\tilde{T}}(\log \rho_k(\tilde{T})) \tag{17}$$

where $\mu_{\tilde{T}}(g)$ is the mean of $g$ for all the values taken over data set $\tilde{T}$

$$\mu_{\tilde{T}}(g) = \frac{1}{N_{\tilde{T}}} \sum_{s \in \tilde{T}} g(s). \tag{18}$$

Informally, the main term in estimate (17) is equal to the mean of the log-distances to the $k$-th nearest neighbor of each sample.

### 4.3. Cross entropy estimation

In the same framework, the cross entropy of two data sets $\tilde{R}$ and $\tilde{T}$ can be approximated by [11]

$$\hat{H}_\times(\tilde{T}, \tilde{R}) \overset{\text{kNN}}{=} \frac{1}{N_{\tilde{T}}} \sum_{s \in \tilde{T}} \log \xi_k(\tilde{R}, s) \tag{19}$$

$$= \log(v_d N_{\tilde{R}}) - \psi(k) + d\ \mu_{\tilde{T}}(\log \rho_k(\tilde{R})). \tag{20}$$

Note again that estimator (20) 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 $\tilde{R}$ of each sample of $\tilde{T}$. Since a sample $s$ of $\tilde{T}$ does not belong to data set $\tilde{R}$, the search for the $k$-th nearest neighbor *excluding* $s$ itself does not in fact exclude any sample of $\tilde{R}$. This is why $N_{\tilde{R}}$ appears in (20) whereas $N_{\tilde{T}} - 1$ appears in (17). Note that $N_{\tilde{R}} = N_{\tilde{T}} = |\Omega|$ .

### 4.4. Kullback-Leibler distance and minimization

**Minimization without derivative.** Since the Kullback-Leibler distance is a difference between a cross entropy and a differential entropy (see Eq. (5)), the kNN estimate of this distance is equal to

$$D_{\text{KL}}(\tilde{T}, \tilde{R}) \overset{\text{kNN}}{=} \log \frac{N_{\tilde{R}}}{N_{\tilde{T}} - 1} + d\ \mu_{\tilde{T}}(\log \rho_k(\tilde{R})) - d\ \mu_{\tilde{T}}(\log \rho_k(\tilde{T})). \tag{21}$$

It has been proven that this estimator is consistent and asymptotically unbiased [12].

Let us remind that $\tilde{R}$ and $\tilde{T}$ are data sets $\{I_{\text{ref}}(x), x\}$, $x \in \Omega$, and $\{I_{\text{tgt}}(\varphi(x)), x\}$, $x \in \Omega$, respectively, where $\varphi$ is a geometric transformation representing the motion of the ROI between the reference frame and the target frame. Therefore, tracking can be performed by minimizing the Kullback-Leibler distance with respect to $\varphi$, or a set of parameters defining $\varphi$. Estimation (21) being defined in the kNN framework, it is not differentiable as is. Its minimization could be performed by an exhaustive search procedure in (a subset of) the space of parameters of $\varphi$. For computational considerations, it will be performed using a suboptimal search procedure known as the diamond search [20].

**Derivative and steepest descent.** Alternatively, one could think of using the Parzen formulation to determine the derivative of the Kullback-Leibler distance and then evaluating the derivative using the kNN framework. This can be done for distance (5). However, the derivative of $D_{\text{KL}}(\tilde{R}, \tilde{T})$ is presented here instead (*i.e.*, with permuted arguments) because it involves the entropy of the reference (which is constant) in place of the entropy of the target. The expression is therefore simpler yet perfectly adapted to demonstrate the validity of the above claim. The development for (5) is similar.

Replacing the Ahmad-Lin approximation (8) in $D_{\text{KL}}(\tilde{R}, \tilde{T})$, one get

$$D_{\text{KL}}(\tilde{R}, \tilde{T}) = \frac{1}{N_{\tilde{R}}} \sum_{x \in \Omega} \Big( \log f_{\tilde{R}}(\tilde{R}(x)) - \log f_{\tilde{T}}(\tilde{R}(x)) \Big) \tag{22}$$

where $f_{\tilde{U}}(s) = \frac{1}{N_{\tilde{U}}} \sum_{y \in \Omega} K_\sigma(s - \tilde{U}(y))$ . Let $\varphi$ be a transformation such that $\varphi(x) = x + M(x)p$ where $M(x)$ is a $2 \times m$-matrix and $p$ is the $m$-vector of the motion parameters (for example, $m = 6$ for an affine motion). The derivative

of distance (22) with respect to $p$ is equal to

$$D'_{\mathrm{KL}}(\tilde{R},\tilde{T}) = -\frac{\alpha}{N_{\tilde{T}}}\sum_{x\in\Omega}\sum_{y\in\Omega} M(y)^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}(\varphi(y)) \\ (I_{\mathrm{ref}}(x) - I_{\mathrm{tgt}}(\varphi(y)))\frac{K_\sigma(\tilde{R}(x)-\tilde{T}(y))}{f_{\tilde{T}}(\tilde{R}(x))} \\ = \alpha\sum_{x\in\Omega}[\mu_{\tilde{R}(x)}(M^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}(\varphi)I_{\mathrm{tgt}}(\varphi)) \\ -\mu_{\tilde{R}(x)}(M^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}(\varphi))I_{\mathrm{ref}}(x)] \qquad (23)$$

where $\alpha$ is equal to $(N_{\tilde{R}}\sigma^2)^{-1}$, $M(y)^{\mathrm{T}}$ is the transpose of $M(y)$, $\mathcal{D}I$ is the $2\times3$-matrix $(\nabla I_Y\ \nabla I_U\ \nabla I_V)$ if $I$ is a color image described in the YUV-space, and $\mu_{\tilde{R}(x)}(f)$ is a weighted mean of $f(s)$ for the samples $s$ of $\tilde{T}$ which belong to the neighborhood of $\tilde{R}(x)$

$$\mu_{\tilde{R}(x)}(f) = \frac{\sum_{y\in\Omega} f(y)\ K_\sigma(\tilde{R}(x)-\tilde{T}(y))}{\sum_{y\in\Omega} K_\sigma(\tilde{R}(x)-\tilde{T}(y))}. \qquad (24)$$

The derivative (23) can be interpreted as a *cross*-mean-shift, *i.e.*, the distance between *(i)* the mean intensity of $I_{\mathrm{tgt}}(\varphi)$ of the samples of $\tilde{T}$ in a window centered at the intensity $I_{\mathrm{ref}}(x)$ of $\tilde{R}(x)$ and *(ii)* the intensity $I_{\mathrm{ref}}(x)$ at this center. This interpretation should be considered up to a weighting by $M^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}$. The mean (24), and therefore the derivative (23), can be computed using the kNN framework, *i.e.*, replacing the $K_\sigma$-weighted mean with a mean based on the $k$ nearest neighbors of $\tilde{R}(x)$ among the samples of $\tilde{T}$

$$D'_{\mathrm{KL}}(\tilde{R},\tilde{T}) \overset{\text{kNN}}{=} \frac{\alpha}{k}\sum_{x\in\Omega}\Bigg[\sum_{y\in\mathrm{kNN}(x)} M(y)^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}(\varphi(y))I_{\mathrm{tgt}}(\varphi(y)) \\ -\left(\sum_{y\in\mathrm{kNN}(x)} M(y)^{\mathrm{T}}\mathcal{D}I_{\mathrm{tgt}}(\varphi(y))\right)I_{\mathrm{ref}}(x)\Bigg] \qquad (25)$$

where $\mathrm{kNN}(x) = \{y, |\tilde{T}(y)-\tilde{R}(x)| < \rho_k(\tilde{T},\tilde{R}(x))\}$. As a consequence, the Kullback-Leibler distance (22) estimated in the kNN framework can be minimized with a steepest descent approach using the derivative (25) also computed in the kNN framework.

## 5. Experimental results

The proposed kNN-based method (referred to as kNN-KL-G where KL stands for Kullback-Leibler and G stands for geometry) was compared with three trackers: a geometry-free version of the proposed method (kNN-KL), SAD (see Eq. 7), and Mean-shift [6, 8]. For the latter, we used the Mean-shift tracker publicly available at [5].

The ROI domain $\Omega$ was a rectangular region (see Figs. 2, 3, and 4 for dimensions). The unknown transformation $\varphi$ was restricted to a translation: two parameters needed to be estimated. As mentioned in Section 4.4, kNN-based distances cannot be minimized using a gradient descent approach. We used a sub-exhaustive search procedure [20] with a search window of $12\times12$-pixels and a pixel precision (for coherence, this procedure was also used for SAD). Parameter $k$ was chosen equal to 3, which satisfies the conditions mentioned at the end of Section 4.1. The robustness to the choice of this parameter is discussed later in the section. The radiometric space used was YUV. No component weighting was done when computing distances in the extended radiometric/geometric space.

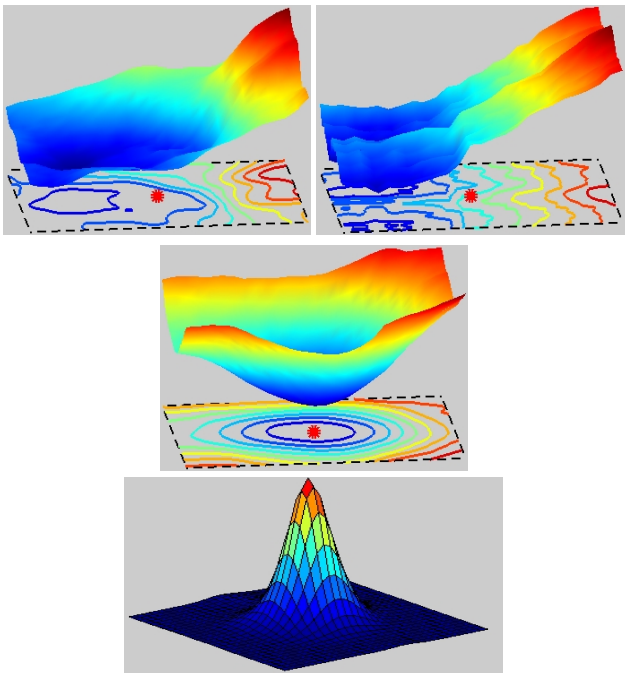

Figure 1. Distance between the reference ROI of sequence "Football" and candidate regions in frame 20 as a function of horizontal and vertical translations: *(from top to bottom)* SAD, kNN-KL, and kNN-KL-G (proposed method). The dashed box is a $12\times12$-square (same size as the search window). The red spot at its center represents the correct translation. The SAD minimum is shifted, kNN-KL has two local minima, and the minimum of kNN-KL-G seems accurate. The last figure is the PDF of the pointwise motion between the reference ROI and the target ROI obtained with kNN-KL-G. For each pixel of the reference, this motion was computed as the space displacement to the nearest neighbor in the extended radiometric/geometric space among the samples of the target. The domain of definition is a $35\times35$-square (to be compared with $\Omega$, a $43\times43$-square) centered around the null translation.

Tracking was performed with $I_{\mathrm{ref}}$ being fixed equal to, say, $I_1$ while $I_{\mathrm{tgt}}$ was successively equal to $I_t$, $t = 2, 3, 4\ldots$ When searching for the ROI in frame $I_t$, the search window was centered around the position of the ROI computed in frame $I_{t-1}$.

Sequence "Car" is an aerial car chase which is part of

the VIVID tracking testbed [5]. It is composed of $640\times480$-frames. In our experiments, tracking was performed on 150 frames (see Fig. 2). Note that the car is partially occulted by trees from frame 36 to frame 116. kNN-KL eventually lost the ROI and ended up tracking the second car which has colors similar to the ROI. This is probably due to the fact that it is based on radiometry only. Mean-shift performed quite well although the tracking shifted upward when occlusion occurred in order to avoid including the green colors of the trees in the color PDF. SAD and kNN-KL-G tracked the car accurately. Concerning SAD, the translation model being fairly well respected within the ROI, taking the pointwise residual makes sense while the use of the absolute value is robust to the outliers arising from the occlusion.

Sequence "Crew" is composed of $352\times288$-frames. Two faces were tracked on 80 frames (see Fig. 3). Note that this sequence has two kinds of variation of luminance: a continuous increase as the crew walks out of a dark area, and strong peaks due to camera flashes. kNN-KL-G tracked the faces successfully. The other methods lost progressively the ROI.

Sequence "Football" is composed of $352\times288$-frames. Tracking was performed on 20 frames (see Fig. 4). Note that this sequence is characterized by fast motions and, consequently, motion blur. Moreover, part of the public has colors similar to colors that can be found in the ROI. In some frames, this area of the public is right above the ROI. This is probably the reason why kNN-KL stayed stuck in this region. Finally, as the player runs, he turns and almost faces the camera toward the end of the sequence. Therefore, the translation model is not appropriate. This can explain why SAD, which relies on a strict translation model, lost the ROI in the first frames. Mean-shift succeeded to track the ROI approximately. However, it could not avoid being attracted by the public. The geometric constraint of kNN-KL-G allowed to avoid being attracted by the public area (where the color spatial distribution is different from that of the ROI) while being soft enough to deal with the mismatch between the translation model and the actual motion. The resulting tracking is accurate.

To support the conclusions about the results on sequence "Football", the distance between the reference ROI and candidate regions in frame 20 was computed as a function of the translation parameters for SAD, kNN-KL, and kNN-KL-G (see Fig. 1). The red spot at the center of the plane represents the correct motion. The SAD minimum is shifted as a result of the inappropriateness of the translation model between frame 1 and frame 20. kNN-KL has several local minima as there are several possible matches when accounting only for radiometry. By adding geometry, kNN-KL-G allows to find a minimum corresponding to the correct motion. Also note that the kNN-KL-G criterion seems strictly convex in a large window around the minimum. This property is interesting for the convergence of optimization algorithms (diamond search in our case). The last plot in Fig. 1 represents the PDF of the pointwise motion between the reference ROI and the target ROI obtained with kNN-KL-G. For each pixel of the reference, this motion was computed as the space displacement (*i.e.*, the distance after projection onto the geometric subspace) to the nearest neighbor in the extended radiometric/geometric space among the samples of the target. The PDF is not a Dirac delta function, illustrating the fact that the translation model was not correct.

To evaluate the robustness of the kNN-KL-G method to the choice of parameter $k$, tracking was performed on sequence "Football" with various values of $k$ that respect the conditions in Section 4.1. The result for $k$ equal to 3 was taken as a reference and the average shift during tracking was measured. For $k$ equal to 10, the average shift was 0.2 pixels; for $k$ equal to 20, the average shift was 0.7 pixels; for $k$ equal to $\sqrt{|\Omega|} = 43$, the average shift was 1.1 pixels. The method is therefore very robust to the choice of $k$.

*Note: videos of the tracking results are available as supplemental material.*

## 6. Discussion and future works

This paper presents a general framework for estimating statistical measures of information. We focused on a distance derived from entropy as proofs of consistency and unbiasedness exist [9, 11, 12].

The kNN-based PDF estimate has two advantages for dealing with high-dimensional data. First, it relies on a non-parametric approach which uses variable size kernels to adapt to the local density of samples. Second, it allows to derive expressions of PDF-based measures (such as entropy or the Kullback-Leibler distance) without computing explicitly the PDFs.

Developments on tracking seem to show that robustness requires the use of more information than just color (*e.g.*, image gradient or motion [4]). In this paper, geometry was added as proposed in [7] while overcoming the difficulty of estimating multivariate PDFs using the kNN framework.

Future works will focus on extending the kNN framework to other statistical measures such as mutual information [17] and the Bregman divergence [2]. The ROI approach should also be extended to accurate motion segmentation.

## References


[1] I. Ahmad and P. E. Lin. A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Inform. Theory*, 22(3):372– 375, 1976.

[2] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh. Clustering with bregman divergences. *J. Mach. Learn. Res.*, 6:1705–1749, 2005.

[3] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, and L. Pronzato. A minimum-entropy procedure for robust motion estimation. In *ICIP*, Atlanta, GA, October 2006.

[4] T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. In *Computer Analysis of Images and Patterns*, volume 2756 of *LNCS*, pages 353–360, Groningen, The Netherlands, 2003.

[5] R. Collins, X. Zhou, and S. K. Teh. An open source tracking testbed and evaluation web site. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, Breckenridge, CO, http://www.vividevaluation.ri.cmu.edu/, 2005.

[6] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR*, pages 142–151, Hilton Head Island, SC, 2000.

[7] A. Elgammal, R. Duraiswami, and L. S. Davis. Probabilistic tracking in joint feature-spatial spaces. In *CVPR*, pages 781–788, Madison, WI, 2003.

[8] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., 1990.

[9] M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. N. Inverardi. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametr. Stat.*, 17(3):277–297, 2005.

[10] P. Huber. *Robust Statistics*. John Wiley and Sons, 1981.

[11] L. Kozachenko and N. Leonenko. On statistical estimation of entropy of random vector. *Problems Infor. Transmiss.*, 23(2):95–101, 1987.

[12] N. Leonenko, L. Pronzato, and V. Savani. A class of Renyi information estimators for multidimensional densities. PASCAL archive: http://eprints.pascal-network.org/archive/00001031/, 2005.

[13] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *ECCV*, volume 2350 of *LNCS*, pages 661–675, Copenhaguen, Denmark, 2002.

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

[15] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[16] R. Venkatesh Babu, P. Pérez, and P. Bouthemy. Robust tracking with motion estimation and local kernel-based color modeling. *Image Vis. Comput. In Press*, 2007.

[17] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *Int. J. Comput. Vis.*, 24(2):137–154, 1997.

[18] J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *J. Math. Imaging Vis.*, 14(3):245–255, 2001.

[19] C. Yang, R. Duraiswami, N. A. Gumerov, and L. Davis. Improved fast gauss transform and efficient kernel density estimation. In *ICCV*, Nice, France, 2003.

[20] S. Zhu and K.-K. Ma. A new diamond search algorithm for fast block-matching motion estimation. *IEEE Trans. Image Process.*, 9(2):287–290, 2000.

Figure 2. Tracking on sequence "Car": frames 1, 30, 60, 90, 120 and 150 (relative to the reference frame). kNN-KL-G (proposed method): green; kNN-KL: cyan; Mean-shift: red; SAD: white. The car is partially occulted by trees from frame 36 to frame 116. $\Omega$: $95 \times 47$-rectangle.

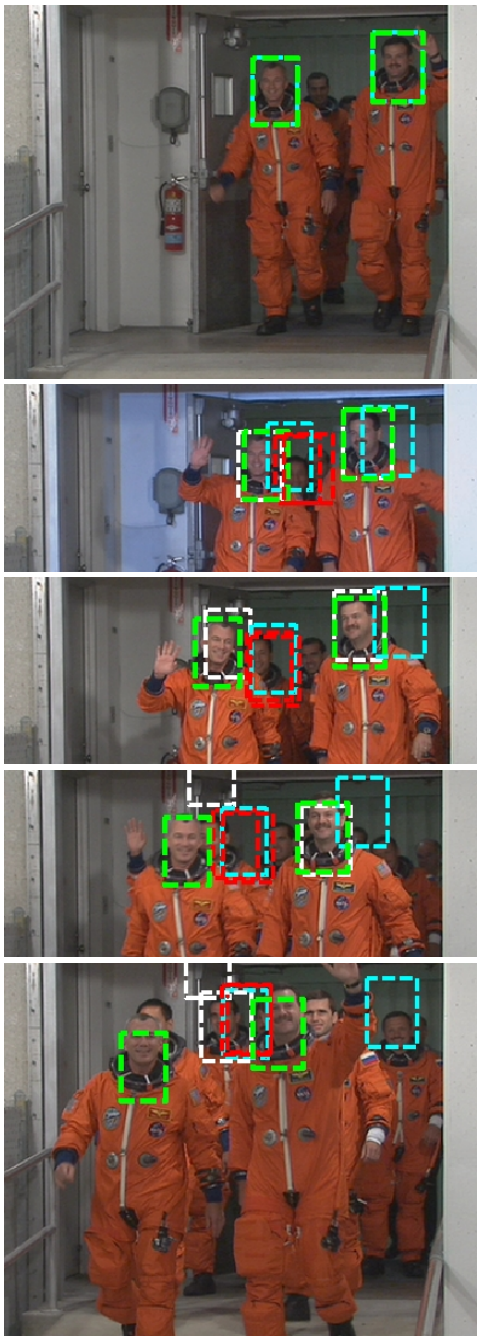

Figure 3. Tracking on sequence "Crew": frames 1, 20, 40, 60 and 80 (relative to the reference frame). kNN-KL-G (proposed method): green; kNN-KL: cyan; Mean-shift: red; SAD: white. Note that there are two kinds of intensity changes in the sequence: a slight, continuous intensity increase and some strong and brief intensity peaks due to camera flashes. $\Omega$: $33 \times 52$-rectangle.

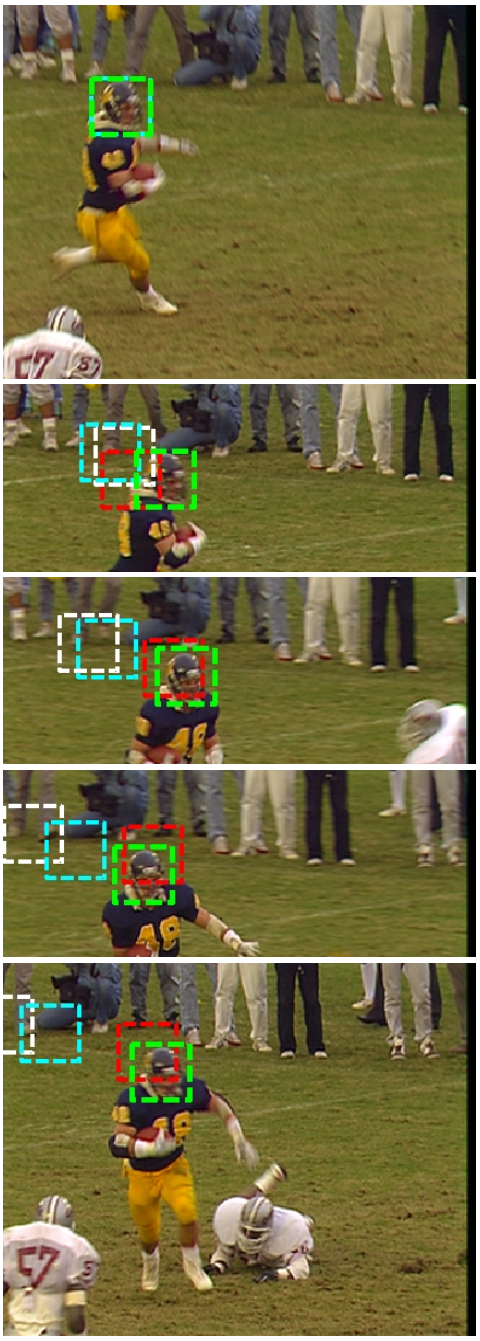

Figure 4. Tracking on sequence "Football": frames 1, 5, 10, 15 and 20 (relative to the reference frame).
kNN-KL-G (proposed method): green; kNN-KL: cyan; Mean-shift: red; SAD: white.
This sequence is characterized by a fast motion generating motion blur. $\Omega$: $43 \times 43$-square.