Pattern Recognition Letters 68 (2015) 48–55

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Unsupervised spatio-temporal filtering of image sequences. A mean-shift specification☆

Simon Mure [a,*], Thomas Grenier [a], Dominik S. Meier [b], Charles R.G. Guttmann [b], Hugues Benoit-Cattin [a]

[a] Université de Lyon, CREATIS, CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1 ; 7 avenue Jean Capelle, Villeurbanne 69621 cedex, France
[b] Center for Neurological Imaging, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, RFB 396, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 4 December 2014
Available online 30 July 2015

Keywords:
Unsupervised spatio-temporal filtering
Mean-shift
Time-series
Feature trajectories
Longitudinal MR acquisitions

ABSTRACT

A new spatio-temporal filtering scheme based on the mean-shift procedure, which computes unsupervised spatio-temporal filtering for univariate feature evolution, is proposed in this paper. Our main contributions are on one hand the modification of the spatial/range domains to appropriately integrate the temporal feature into the mean-shift iterative form and on the other hand the addition of a trajectory constraint in the feature space with the use of the infinity norm. Therefore, only the samples living the same life in the feature space will converge. Major assets of the standard mean-shift framework such as convergence and bandwidth parameters adjustment are preserved. In this paper, we study the relative importance of the bandwidth parameters and the efficiency of the proposed method is assessed on synthetic data and compared to the standard mean-shift framework for spatio-temporal data filtering. The obtained results have allowed us to undertake a first study on real data, which has led to encouraging results in identifying spatio-temporal evolution of multiple sclerosis lesions appearing on T2-weighted MR images.

1. Introduction

Due to the dramatic increase of longitudinal acquisitions in the past decades such as video sequences, global positioning system (GPS) tracking or medical follow-up, many applications for time-series data mining have been developed. Thus, unsupervised time-series data mining has become highly relevant with the aim to automatically detect and identify similar temporal patterns in time-series datasets.

Several time-series clustering methods have been proposed for forecasting based on the study of signal correlations [28], shape attributes [14,25] or evolution models [18]. The work presented in [24] introduces an unsupervised and parameter-free method to mine regimes (patterns) and transitions (discontinuities) in large co-evolving time-series but does not cluster similar evolution. According to [1], the only known methods that can be generalized to multivariate time-series clustering are the ones proposed in the domain of spatial trajectories [4,17,22,33].

However, some works have already been published in the context of video streams analysis where multivariate image sequences are often being processed, e.g. which evolving features are a combination of the spatial and the color information. The major issues in this context are image enhancement and filtering to facilitate detection, tracking and patterns segmentation. Many studies dealing with image sequence filtering or restoration have been conducted since the early 90's (see [5]). For example, [11] briefly described how to extend the mean-shift (M-S) framework to the space/time domain in order to filter video sequences. Approaches using M-S for color video segmentation have also been proposed [9,19,32]. In these works, pixels of each frame were considered as independent samples, i.e. M-S was not used to filter the temporal evolution of the features associated to a pixel. Rather, in these publications M-S was used to filter multi-channel (e.g. RGB) video volumes (2D+t) or features previously derived for each pixel of these volumes. [31] and more recently [23] have proposed new approaches exploiting both the spatial and the temporal redundancies of the data and compared them with the most efficient ones known so far for spatio-temporal video filtering. Nevertheless the performance of these methods is dependent on the choice of a motion estimator, which is the most challenging step.

The M-S technique, which has been proposed by Fukunaga and Hostetler [12] can be used in the context of image filtering and segmentation [8]. M-S has already been applied to spatio-temporal data filtering [3,11] and some works were published on longitudinal MRI processing with M-S [2,6,20]. However, they have neither directly

☆ This paper has been recommended for acceptance by Dr. C. Luengo.

* Corresponding author. Tel.: +33 472 437 545.
E-mail address: simon.mure@creatis.insa-lyon.fr (S. Mure).

http://dx.doi.org/10.1016/j.patrec.2015.07.021
0167-8655/

used the time information, reducing it to a scalar value, nor explicitly formulated how to handle the temporal dimension. In contrast, we claim to explain how to extend M-S to spatio-temporal data by adding a constraint on the evolution of the samples over time. Only the samples sharing a similar evolution pattern of their feature will contribute to the data filtering. The standard M-S principle and our main contribution, the extension of the M-S procedure by specifically taking into account the temporal dimension, are introduced in Section 2. Section 3 presents the validation protocol while Section 4 evaluates our method on both simulated data, with comparison to filtering outputs obtained with M-S, and real data.

2. Methods

2.1. Mean shift procedure

Let us consider $\mathbf{X}$ the image sample set in the D-dimensional spatial/range domain as input samples of the M-S procedure:

$$\mathbf{X} = \{\mathbf{x}_i\}_{i=1\ldots n} \quad \text{with} \quad \begin{array}{l} \mathbf{x}_i \in \mathbb{R}^D\text{: input samples} \\ i = 1, \ldots, n\text{: sample index} \end{array} \tag{1}$$

A sample's new position at iteration $k+1$ is computed with:

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^{n} g\left(d^2\left(\mathbf{x}_i^{[k]}, \mathbf{x}_j^{[k]}, \mathbf{H}\right)\right) \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^{n} g\left(d^2\left(\mathbf{x}_i^{[k]}, \mathbf{x}_j^{[k]}, \mathbf{H}\right)\right)} \tag{2}$$

where d is the Mahalanobis distance between two samples $\mathbf{u}$ and $\mathbf{v}$ of $\mathbb{R}^D$ and is defined as:

$$d(\mathbf{u}, \mathbf{v}, \mathbf{H}) = \sqrt{(\mathbf{u} - \mathbf{v})' \mathbf{H}^{-1} (\mathbf{u} - \mathbf{v})} \tag{3}$$

with $\mathbf{H}$ the bandwidth matrix, squared and positive definite.

The function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is the "neighborhood function", that acts as a weight function of the squared Mahalanobis distance.

This M-S formulation describes the blurring M-S scheme [12], where the updated values of the samples are used at each iteration to reach the density maximum until no significant shift is observed between two consecutive iterations. As shown by Cheng [7] and [10], the blurring procedure is ensured to converge.

2.2. Proposed approach: STM-S

In this section we describe our main contribution: a novel approach allowing spatio-temporal time-series filtering based on the M-S framework. We shall name this procedure spatio-temporal mean-shift (STM-S).

Let us now consider a set of n samples located at the positions $\{\mathbf{x}_{s,i}\}_{i=1..n}$ and a set of feature values evolving over time $\{\mathbf{x}_{t,i}\}_{i=1..n}$. The number of spatial dimensions and time-points are respectively noted S and T. The input data set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1\ldots n}$ is now defined as follows:

$$\mathbf{x}_i = \left[\mathbf{x}_{s,i}' \, \mathbf{x}_{t,i}'\right]' \in \mathbf{X} \quad \text{with} \quad \begin{array}{l} \mathbf{x}_{s,i} \in \mathbb{R}^S\text{: spatial domain} \\ \mathbf{x}_{t,i} \in \mathbb{R}^T\text{: temporal domain} \\ i = 1, \ldots, n\text{: sample index} \end{array} \tag{4}$$

Considering these notations, we propose the following equation to iteratively compute the STM-S filtering of each sample:

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^{n} Sp_{i,j}\left(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}\right) \cdot Ra_{i,j}\left(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}\right) \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^{n} Sp_{i,j}\left(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}\right) \cdot Ra_{i,j}\left(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}\right)} \tag{5}$$

where $Sp_{i,j}(\cdot)$ and $Ra_{i,j}(\cdot)$ are respectively the weighted distances of the spatial and the temporal domains between a sample of interest $\mathbf{x}_i$ and another sample $\mathbf{x}_j$ ($\mathbf{x}_i$ and $\mathbf{x}_j \in \mathbb{R}^{S+T}$):

$$Sp_{i,j}\left(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}\right) = g_s\left(d_s^2\left(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}, \mathbf{H}_s\right)\right) \tag{6}$$

$$Ra_{i,j}\left(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}\right) = g_r\left(d_r^2\left(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}, \mathbf{H}_r\right)\right) \tag{7}$$

The combination of the weighted distances in (5) leads to the weights associated to each sample involved in the STM-S computation. Unlike the standard M-S approach, the distances between the samples are not calculated in the same way for the spatial and the temporal dimensions. In the spatial domain, the Mahalanobis distance $d_s(\mathbf{u}_s, \mathbf{v}_s, \mathbf{H}_s)$ is computed between two samples $\mathbf{u}_s$ and $\mathbf{v}_s$ where $\mathbf{H}_s$ is the spatial bandwidth matrix of size $S \times S$. In contrast, to compute the distance between the features evolving in the temporal domain we use the infinity norm:

$$d_r(\mathbf{u}_t, \mathbf{v}_t, \mathbf{H}_r) = \| \mathbf{H}_r^{-\frac{1}{2}} (\mathbf{u}_t - \mathbf{v}_t) \|_\infty \tag{8}$$

where $\mathbf{H}_r$ is the bandwidth applied on the temporal feature values, a square matrix of size $T \times T$. This distance corresponds to the biggest distance between two samples in the temporal feature space, with values scaled by the bandwidth matrix $\mathbf{H}_r$.

The samples that will be kept for the mean computation (5) are the ones close enough to the sample of interest over their evolution. Thus, the samples participating in the STM-S computation are selected with respect to the closeness of their spatial positions and of their temporal trajectories to the sample of interest. In this work, we use the same profile function g to weight both distances:

$$g_s\left(d_s^2(\cdot)\right) = g_r\left(d_r^2(\cdot)\right) = \begin{cases} 1 & \text{if } d_s^2(\cdot),\ d_r^2(\cdot) \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Eq. (7) ensures that $Ra_{i,j}^{[k]}$ becomes zero if the distance between two trajectories at a given time point exceeds 1. Consequently, trajectories far from the reference will be excluded from the STM-S computation.

A graphical example of the sample selection described above is illustrated in Fig. 1. Although the red sample is included in the spatial neighborhood of the reference sample (the blue one) it will not be kept for its update because the candidate evolution lies outside the reference evolution boundary. On the contrary, the green sample is both spatially and temporally close enough to the reference sample to participate to its update. It is important to note that only one spatial value is kept for each sample because a pixel position is the same in

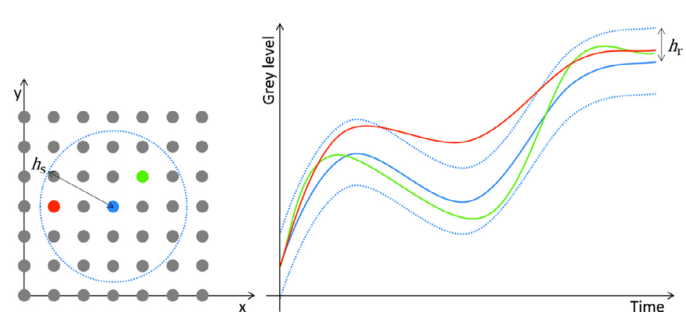

Fig. 1. The left plot shows the spatial features of the samples extracted from a 2D image at the first iteration ($k = 0$). The right plot shows the grey level evolution of three samples, the evolution are associated to the spatial features with the same color. The blue color identifies the reference sample, while the green color and the red color identify two candidate samples. The dotted blue lines represent the boundaries defined by the infinity norm applied on the grey level evolution of the reference sample. Candidate evolution values must lie between these limits during the whole time-course to be taken into account in (5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).