2-adic Fuzzy Partitions and Multi-Scale Representation of Time Series

Irina Perfilieva[1] David Adamczyk

University of Ostrava, 30.dubna, 22, Czech Republic,
irina.perfilieva,david.adamczyk@osu.cz,
WWW home page: https://ifm.osu.cz/irina-perfiljeva/13263/

Abstract: We are focused on a new method of time series analysis that is based on the extraction of representative keypoints. We use the multi-scale theory based on the use of non-traditional kernels derived from the theory of F-transforms. The sequence of kernels corresponds to what is called as 2-adic fuzzy partitions. This leads to simplified algorithms and comparable efficiency in the selection of keypoints. We reduce the number of representative keypoints and enhance robustness of their selection. We also propose a new keypoint descriptor and test it on matching financial time series with high volatility.

1 Introduction

A new method of time series analysis based on the selection of representative keypoints with subsequent reverse reconstruction is proposed and tested. Keypoints serve as indicators of the area in which a functional object (time series, image, etc.) has clearly expressive features compared to other nearly flat areas. The way of extracting keypoints and related features is similar to the processing of data by neural networks. Indeed, the latter is focused on stable feature extraction that is invariant with respect to various geometric transformations.

In the related domain, image processing, features are associated with keypoints and their descriptors. Both are used to be identified with local areas of the image that correspond to its content (not related to the background). Thus, the processing is computationally consuming and depends on many space-environment conditions: illumination, position, resolution, etc. In [10], it has been shown that invariant local features (Harris corners and their rotation invariant descriptors) can contribute to solving general problems of image recognition. However, the Harris corner detector is sensitive to changes in image scale. This was eliminated in [6] where the descriptor known in the literature as SIFT was proposed. The method of SIFT has inspired many modifications: SURF, PCA-SIFT, GLOH, Gauss-SIFT, etc., (see [5] and references therein), aimed at improving efficiency in various senses: reliability, computation time, etc. However, the main stages and their semantics have been preserved.

Our contribution to this topic is as follows: we use the basic methodology of SIFT and its modifications, but select non-traditional kernels derived from the theory of F-transforms [7]. This allows to simplify the scaling and selection of key points, as well as reduce their number and enhance robustness. We also propose a new keypoint descriptor and test it on matching financial time series with high volatility.

The main theoretical result that we arrive at here is that the Gaussian kernel as the predominant in the scale-space theory can be replaced with the same success by a special symmetric positive semi-definite kernel with a local support. In particular, we show that generating function of a triangular-based uniform fuzzy partition of $\mathbb{R}$ can be used for determining such kernel. This fact allows us to base upon the theory of F-transforms and its ability to extract features (keypoints) with a clear understanding of their semantic meaning [8].

2 Briefly about the theory of scale-space representations

We start with a brief overview (see in [4]) of the mentioned theory because it explains the proposed methods. Perhaps the quad-tree methodology is the first type of multi-scale representation of image data. It focuses on recursively dividing an image into smaller areas controlled by the intensity range. The low-pass pyramid representation then facilitated multi-scaling in such a way that the image size decreased exponentially compared to scale level.

Koenderink [3] emphasized that scaling up and down the internal scope of observations and handling image structures at all scales (in accordance with the task) contribute to a successful image analysis. The challenge is to understand the image at all relevant scales at the same time, but not as an unrelated set of derived images at different levels of blur.

The basic idea (in Lindeberg [4]) how to obtain a multi-scale representation of an object is to embed it into a one-parameter family of gradually smoothed ones where fine-scale details are sequentially suppressed. Under fairly general conditions, the author showed that the Gaussian kernel and its derivatives are the only possible smoothing kernels. These conditions are mainly linearity and shift invariance, combined with various ways of formalizing the notion that structures on a coarse scale should correspond to simplifications of corresponding structures on a fine scale.

A scale-space representation differs from a multi-scale representation in that it uses the same spatial sampling at all scales and one continuous scale parameter as the generator. By the construction in Witkin [11], a scale-space representation is a one-parameter family of derived signals constructed using convolution with a one-parameter family of Gaussian kernels of increasing width.

Formally, a scale-space family of a continuous signal is constructed as follows. For a signal $f : \mathbb{R}^N \to \mathbb{R}$, the scale-space representation $L : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ is defined by:

$$\begin{aligned} L(\cdot,0) =& f(\cdot), \\ L(\cdot,t) =& g(\cdot,t) \star f, \end{aligned} \tag{1}$$

where $t \in \mathbb{R}_+$ is the scale parameter and $g : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ is the Gaussian kernel as follows:

$$g(x,t) = \frac{1}{(2\pi t)^{N/2}} \exp - \sum_{i=1}^{N} \frac{x_i^2}{2t}.$$

The scale parameter t relates to the standard deviation of the kernel g, and is a natural measure of spatial scale at the level t.

As an important remark, we note that the scale-space family L can be defined as the solution to the diffusion (heat) equation

$$\partial_t L = \frac{1}{2} \nabla^T \nabla L, \tag{2}$$

with initial condition $L(\cdot,0) = f$. The Laplace operator, $\nabla^T \nabla$ or Δ, the divergence of the gradient, is taken in the spatial variables.

The solution to (2) in one-dimension and in the case where the spatial domain is R is known as the convolution ($\star$) of f (initial condition) and the fundamental solution:

$$L(\cdot,t) = g(\cdot,t) \star f, \tag{3}$$

$$g(x,t) = \frac{1}{(\sqrt{2\pi t})} \exp - \frac{x^2}{2t}. \tag{4}$$

The following two questions arise: is this approach the only reasonable way to perform low-level processing, and are Gaussian kernels and their derivatives the only smoothing kernels that can be used? Many authors [4, 11, 3] answer these questions positively, which leads to the default choice of Gaussian kernels in most image processing tasks. In this article, we want to expand on the set of useful kernels suitable for performing scale-space representations. In particular, we propose to use kernels arising from generating functions of fuzzy partitioning.

3 Space with a Fuzzy Partition as a Universe with Closeness

In this section, we introduce space that plays an important role in our research. A space with a fuzzy partition is considered as a space with a proximity (closeness) relation, which is a weak version of a metric space. As we indicated at the beginning, our goal is to extend the Laplace operators to those that take into account the specifics of spaces with fuzzy partitions. Then, the next goal is to show that the diffusion (heat conduction) equation in (2) can be extended to spaces with closeness, where the concepts of derivatives are adapted to nonlocal cases. Both things allow us to use the theory of scale-space representation and propose on its basis a new method for localizing key points.

Let us first recall the basic definitions of all related concepts.

3.1 Fuzzy partition

Definition 1: Fuzzy sets $A_1, \ldots, A_n : [a,b] \to \mathbb{R}$, establish a *fuzzy partition* of the real interval $[a,b]$ with nodes $a = x_1 < \ldots < x_n = b$, if for all $k = 1, \ldots, n$, the following conditions are valid (we assume $x_0 = a$, $x_{n+1} = b$):

1. $A_k(x_k) = 1$, $A_k(x) > 0$ if $x \in (a_k, b_k)$, $a \le a_k < b_k \le b$;
2. $\bigcup_{k=1}^{n} (a_k, b_k) = (a,b)$;
3. $A_k(x) = 0$ if $x \notin [a_k, b_k]$;
4. $A_k(x)$ is continuous on $[a_k, b_k]$.

The membership functions $A_1, \ldots, A_n$ are called *basic functions* [7].

Definition 2: The fuzzy partition $A_1, \ldots, A_n$, where $n \ge 2$, is *h-uniform* if nodes $x_1 < \cdots < x_n$ are h-equidistant, i.e. for all $k = 1, \ldots, n-1$, $x_{k+1} = x_k + h$, where $h = (b-a)/(n-1)$, and there exists an even function $A_0 : [-1,1] \to [0,1]$, such that $A_0(0) = 1$, and for all $k = 1, \ldots, n$:

$$A_k(x) = A_0\left(\frac{x - x_k}{H}\right), x \in [x_k - H, x_k + H] \cap [a,b],$$

where $H > h/2$.

A_0 is called a *generating function* of a uniform fuzzy partition [7]. Generating function A_0 is *normalized* if

$$\int_{-1}^{1} A_0(x)dx = 1.$$

Remark 1. Rescaled generating function $A_H(x) = A_0(x/H)$ generates the corresponding kernel $A_H(x-y)$.

3.2 Discrete Universe with Closeness

In this section, we introduce a finite space with a binary relation of *closeness* and then show how closeness can be related to a uniform fuzzy partition.

The best formal model of a space with closeness is a weighted graph $G = (V,E,w)$ where $V = \{v_1, \ldots, v_\ell\}$ is a finite set of vertices, and E ($E \subset V \times V$) is a set of weighted edges so that $w : E \to \mathbb{R}_+$. The edge $e = (v_i, v_j)$ connects two vertices v_i and v_j, and then the weight of e is $w(v_i, v_j)$ or just w_{ij}. Weights are set using the function $w : V \times V \to \mathbb{R}_+$, which is symmetric ($w_{ij} = w_{ji}, \forall 1 \le i, j \le \ell$), non-negative ($w_{ij} \geq 0$) and $w_{ij} = 0$ if $(v_i, v_j) \notin E$. The notation $v_i \sim v_j$ denotes two adjacent vertices v_i and v_j with an existing edge connecting them. Function w sets the closeness on V.

There are many models of closeness in the literature with the default options: Gaussian and uniform distributions [1]. Below, we propose a new model based on a uniform fuzzy partition.

We use the above given (graph) terminology and assume that the set of vertices V is identified with the set of indices $V = \{1, \ldots, \ell\}$ and that the corresponding interval $[1, \ell]$ is 1-uniform fuzzy partitioned with normalized basic functions $A_1^H, \ldots A_\ell^H$, so that $A_k^H(x) = A_H(x-k)$, $k = 1, \ldots, \ell$, $A_H(x) = A_0(x/H)$, $H \geq 1$, and A_0 is the generating function.

Definition 3: Graph $G_H = (V, E, w_H)$ is fuzzy weighted, if $V = \{1, \ldots, \ell\}$, and the weight function $w_H : V \times V \to [0,1]$, is determined by a 1-uniform fuzzy partition $A_1^H, \ldots A_\ell^H$, of $[1, \ell]$, where $H \geq 1$, so that $w_H(v_i, v_j) = A_i^H(j)$, $i, j = 1, \ldots, \ell$.

4 Discrete (Non-local) Laplace operator

In this section, we aim to develop elements of functional analysis on spaces with closeness in order to be able to introduce an operator with properties similar to the Laplacian. We recall the definition of (non-local) Laplace operator as a differential operator given by the divergence of the gradient of a function (see [2]). On spaces with closeness, the generalized version knows as Laplace-Beltrami operator is used. The definition is based on the self-adjoint property, which leads us to definitions of the corresponding Hilbert spaces.

Let $G = (V, E, w)$ be a weighted graph model of a space with closeness, and let $f : V \to \mathbb{R}$ be a real-valued function. Let $H(V)$ denote the Hilbert space of real-valued functions on V, such that if $f, h \in H(V)$ and $f, h : V \to \mathbb{R}$, then the inner product $\langle f, h \rangle_{H(V)} = \sum_{v \in V} f(v)h(v)$. Similarly, $H(E)$ denotes the space of real-valued functions defined on the set E of edges of a graph G. This space has the inner product $\langle F, H \rangle_{H(E)} = \sum_{(u,v) \in E} F(u,v)H(u,v) = \sum_{u \in V} \sum_{v \sim u} F(u,v)H(u,v)$, where $F, H : E \to \mathbb{R}$ are two functions on $H(E)$.

The difference operator $d : H(V) \to H(E)$ of f, is defined on $(u,v) \in E$ by

$$(df)(u,v) = \sqrt{w(u,v)}\,(f(v) - f(u)). \tag{5}$$

The directional derivative of f, at vertex $v \in V$, along the edge $e = (u,v)$, is defined as:

$$\partial_v f(u) = (df)(u,v). \tag{6}$$

The adjoint to the difference operator $d^* : H(E) \to H(V)$, is a linear operator defined by:

$$\langle df, H \rangle_{H(E)} = \langle f, d^* H \rangle_{H(V)}, \tag{7}$$

for any function $H \in H(E)$ and function $f \in H(V)$.

Proposition 1: The adjoint operator d^* can be expressed at a vertex $u \in V$ by the following formula:

$$(d^* H)(u) = \sum_{v \sim u} \sqrt{w(u,v)}\,(H(v,u) - H(u,v)). \tag{8}$$

The divergence operator, defined by $-d^*$, measures the network outflow of a function in $H(E)$, at each vertex of the graph.

The weighted gradient operator of $f \in H(V)$, at vertex $u \in V$, $\forall (u, v_i) \in E$, is a column vector:

$$\nabla_w f(u) = (\partial_v f(u) : v \sim u)^T = (\partial_{v_1} f(u), \ldots, \partial_{v_k} f(u))^T.$$

The weighted Laplace operator $\Delta_w : H(V) \to H(V)$, is defined by:

$$\Delta_w f = -\frac{1}{2} d^*(df). \tag{9}$$

Proposition 2 [2]: The weighted Laplace operator Δ_w at $f \in H(V)$ acts as follows:

$$(\Delta_w f)(u) = -\sum_{v \sim u} w(u,v)(f(v) - f(u)).$$

This Laplace operator is linear and corresponds to the graph Laplacian.

Proposition 3 [9]: Let $G_H = (V, E, w_H)$ be a fuzzy weighted graph, corresponding to the 1-uniform fuzzy partition of $V = \{1, \ldots \ell\}$. Then, the weighted Laplace operator Δ_H at $f \in H(V)$ acts as follows:

$$(\Delta_H f)(i) = -\sum_{i \sim j} A_i^H(j)(f(j) - f(i)) = f(i) - F^H[f]_i,$$

where $F^h[f]_i$, $i = 1, \ldots, \ell$, is the i-th discrete F-transform component of f, cf. [7].

5 Multi-scale Representation in a Space with a Fuzzy Partition

Taking into account the introduced notation, we propose the following scheme for the multi-scale representation L_{FP} of a signal $f : V \to \mathbb{R}$, where $V = \{1, \ldots, \ell\}$ and subscript "FP" stands for an 1-uniform fuzzy partition determined by parameter $H \in \mathbb{N}$, $H \geq 1$:

$$\begin{aligned} L_{FP}(\cdot, 0) &= f(\cdot), \\ L_{FP}(\cdot, t) &= F^{2^t H}[f], \end{aligned} \tag{10}$$

where $t \in \mathbb{N}$ is the scale parameter and $F^{2^t H}[f]$ is the complete vector of F-transform components of f. The scale parameter t relates to the length of the support of the corresponding basic function. As in the case of (1), it is a natural measure of spatial scale at level t. To show the relationship to the diffusion equation, we formulate the following general result.

Proposition 4: Assume that two time continuously differentiated real function $f : [a,b] \to \mathbb{R}$, and $[a,b]$ is h-uniform fuzzy partitioned by $A_1^H, \ldots, A_n^H$ and $A_1^{2H}, \ldots, A_n^{2H}$, where basic functions A_i^H (A_i^{2H}), $i = 1, \ldots, n$, are generated by $A_0(x) = 1 - |x|$ with the nodes at $x_i = a + \frac{b-a}{n-1}(i-1)$. Then,

$$F^{2H}[f]_i - F^H[f]_i \approx \frac{H^2}{4} f''(x_i). \tag{11}$$

The semantic meaning of this proposition in relation to the proposed scheme (10) of multi-scale representation L_{FP} of f is as follows:

The F-transfrom (FT)-based Laplacian of f (11) can be approximated by the (weighted) differences of two adjacent convolutions determined by the triangular-shaped generating function.

6 Experiments with Time Series

6.1 Reconstruction from FT-based Laplacians

To demonstrate the effectiveness of the proposed representation, we first show that an initial time series can be (with a sufficient precision) reconstructed from a sequence of FT-based Laplacians. Below, we illustrate this claim on a financial time series with high volatility. With each value of $t = 1, 2, \ldots$ we obtain the corresponding FT-based Laplacian as the difference between two adjacent convolutions (vectors with F-transform components), so that we obtain the sequence

$$\{L_{FP}(\cdot, t+1) - L_{FP}(\cdot, t) \mid t = 1, 2, \ldots\}$$

The stop criterion is closeness to zero of the current difference. We then compute the reconstruction by summing all the elements in the sequence. Figure 1 shows the step-by-step reconstruction and the final reconstructed time series. The latter is plotted on the bottom image along with the original time series to give confidence in a perfect fit. The estimated Euclidean distance is 89.6.

In the same Figure 1, we show one MLP reconstructions of the same time series with the following configurations: 4 hidden layers with 4086 neurons in each layer (common setting) and learning rate 0.001. It is obvious that the proposed multi-scale representation and subsequent reconstruction are computationally cheaper and give results with better reconstruction quality. To confirm, we give estimates of the Euclidean distances between the original time series and its reconstructions: (from a sequence of FT-based Laplacians) against 159.3 (using MLP).

6.2 Keypoint Localization and Description

Keypoint localization. The localization accuracy of key points depends on the problem being solved. When analyzing time series, the accuracy requirements are different from those used in computer vision to match or register images. Time series focuses on comparing the target and reference series in order to detect similarities and use them to make a forecast. Therefore, the spatial coordinate is not so important in contrast to the comparative analysis of local trends and their changes in time intervals with adjacent key points as boundaries.

Taking into account the above arguments, we propose to localize and identify keypoints from the second-to-last scaled representation of the Laplacian before the latter meets the stopping criterion. We then follow the technique suggested in [6, 5] and identify the keypoint with the local extremum point of the Laplacian corresponding to the selected scale. As in the cited above works, we faced a number of technical problems related to the stability of local extrema, sampling frequency in a scale, etc. Due to the different spatial organization of the analyzed objects (time series versus images), we found simpler solutions to the problems raised. For example, in order to exclude extrema close to each other (and therefore they are very unstable), we leave only one representative, the value of which gives the best semantic correlation with the characteristic of this particular extremum.

Below, we give illustrations to some processed by us time series. They were selected from the cite with historical data in Yahoo Finance. We analyzed the 2016 daily adjusting closing prices using international stock indices, namely Prague (PX), Paris (FCHI), Frankfurt (GDAXI) and Moscow (MOEX). Due to the daily nature of the time series, they all have high volatility, which is additional support for the proposed method. In Figure 2, we show the time series with stock indices PX (Prague) and its last three scaled representations of the Laplacian, where the latter satisfies the stopping criterion. Selected (filtered out) keypoints are marked with red (blue) dots.

Keypoint Description. Due to the specificity of time series with high volatility, we propose a keypoint descriptor as a vector that includes only the Laplacian values at keypoints from two adjacent scales and in the area bounded by an interval with boundaries set by adjacent left/right keypoints from the same scale. In addition, we normalize the keypoint descriptor coordinates by the Laplacian value of the principal keypoint. As our experiments with matching keypoint descriptors of different time series show, the proposed keypoint descriptor is robust to noise and invariant with respect to spatial shifts and time series ranges. The last remark is that the quality of matching is estimated by the Euclidean distance between keypoint descriptors.

To illustrate the assertion about robustness and invariance, we show (Figure 3) with the results of matches between principal keypoints of time series with stock indices PX (Prague), FCHI (Paris), and GDAXI (Frankfurt). In all cases, the stock indices PX were considered as tested and compared against the reference stock indices FCHI and GDAXI.

Conclusion

We are focused on a new fast and robust algorithm of image/signal feature extraction in the form of representative keypoints. We have contributed to this topic by showing that the use of non-traditional kernels derived from the theory of F-transforms [7] leads to simplified algorithms and comparable efficiency in the selection of keypoints. Moreover, we reduced their number and enhanced robustness. This has been shown at the theoretical and experimental levels. We also proposed a new keypoint descriptor and tested it on matching financial time series with high volatility.

Acknowledgment

The work was supported from the project ERDF/ESF "Centre for the development of Artificial Intelligence Methods for the Automotive Industry of the region" (No. CZ.02.1.01/0.0/0.0/17-049/0008414).

Additional support of the grant project SGS18/PrF-MF/2021 (University of Ostrava) is kindly announced.

References

[1] Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation, **15**(6) (2003) 1373–1396.

[2] Elmoataz, A., Lezoray, O., Bougleux, S.: Discrete regularization on weighted graphs: A framework for image and manifold processing, IEEE Transactions on Image Processing, **17** (2008) 1047 - 1060.

[3] Koenderink, J. J.: The structure of images, Biological Cybernetics, **50** (1984) 363 - 370.

[4] Lindeberg, T.: Scale-space theory: a basic tool for analyzing structures at different scales, Journal of Applied Statistics, **21** (1994) 225- - 270.

[5] Lindeberg, T.: Image matching using generalized scale-space interest points, Journ. Mathematical Imaging and Vision, **52**(1) (2015) 3-36.

[6] Lowe, D. G.: Distinctive image features from scale-invariant key-points, Int. J. Computer Vision, **60** (2004) 91 - 110.

[7] Perfilieva, I.: Fuzzy transforms: Theory and applications, Fuzzy sets and systems **157** (8) (2006) 993 - 1023.

[8] Molek, V., Perfilieva, I.: Deep Learning and Higher Degree F-transforms: Interpretable Kernels Before and After Learning, Int. Journ. Computational Intelligence Systems, **13**(1) (2020) 1404-1414.

[9] Perfilieva, I., Vlasanek, P.: Total variation with nonlocal FT-Laplacian for patch-based inpainting, Soft Computing **23** (2019) 1833 - 1841.

[10] Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval, IEEE Trans. Pattern Analysis and Machine Intelligence, **19**(5), (1997) 530-535.

[11] Witkin, A. P.: Scale-space filtering, Proc. 8th International Joint Conference on Artifcial Intelligence, IJCAI'83, **2** (1983) 1019 - 1022.

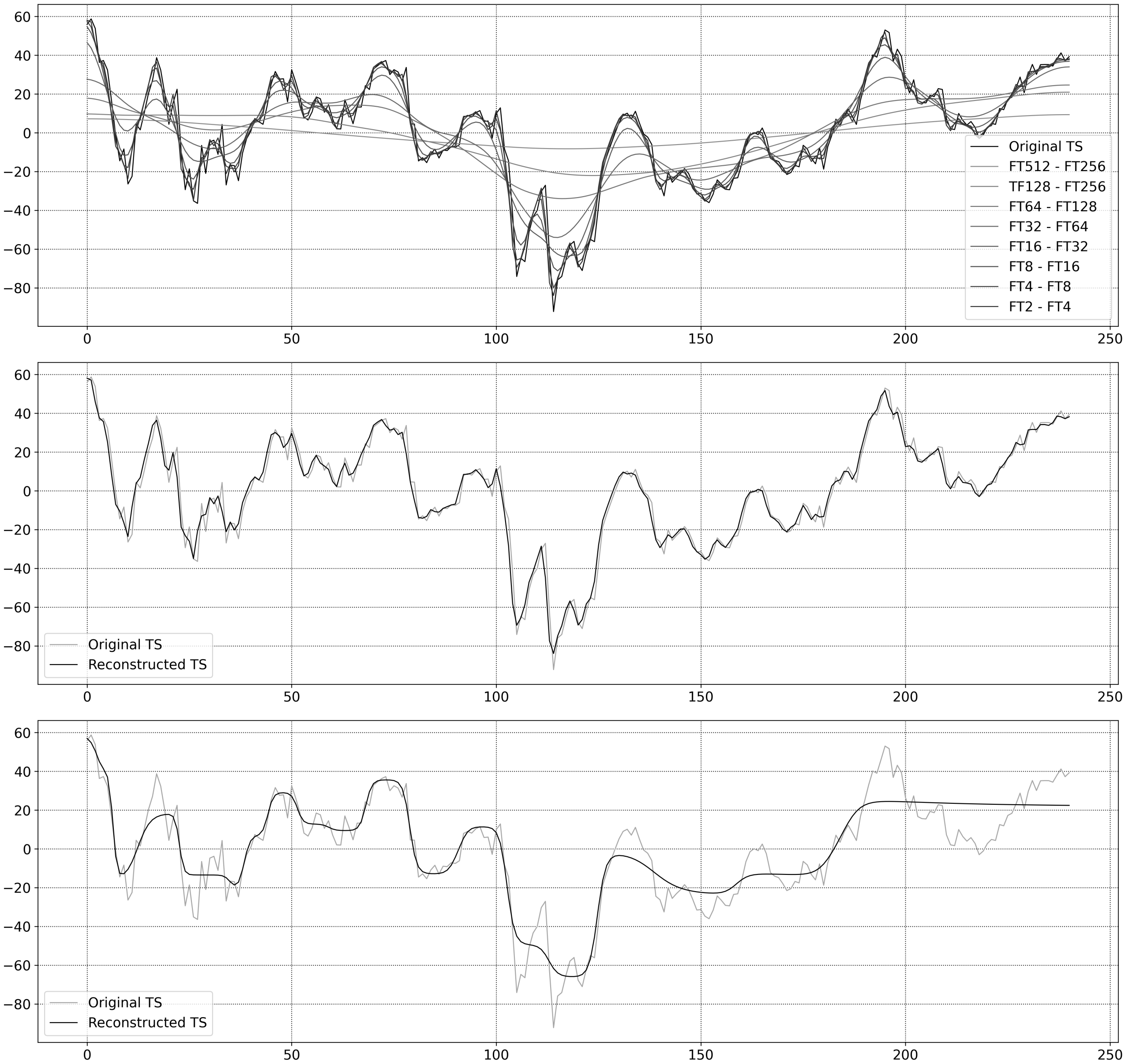

Figure 1: *Top.* The sequence of reconstruction steps, where with each value $t = 1, 2, \ldots$ we improve the quality of the reconstruction by adding the corresponding Laplacian to the previous one. *Middle.* The original time series (gray) against its reconstruction (black) from a sequence of FT-based Laplacians with the distance 89.6. *Bottom.* The original time series (gray) against its MLP reconstruction (black) with the distance 159.3.

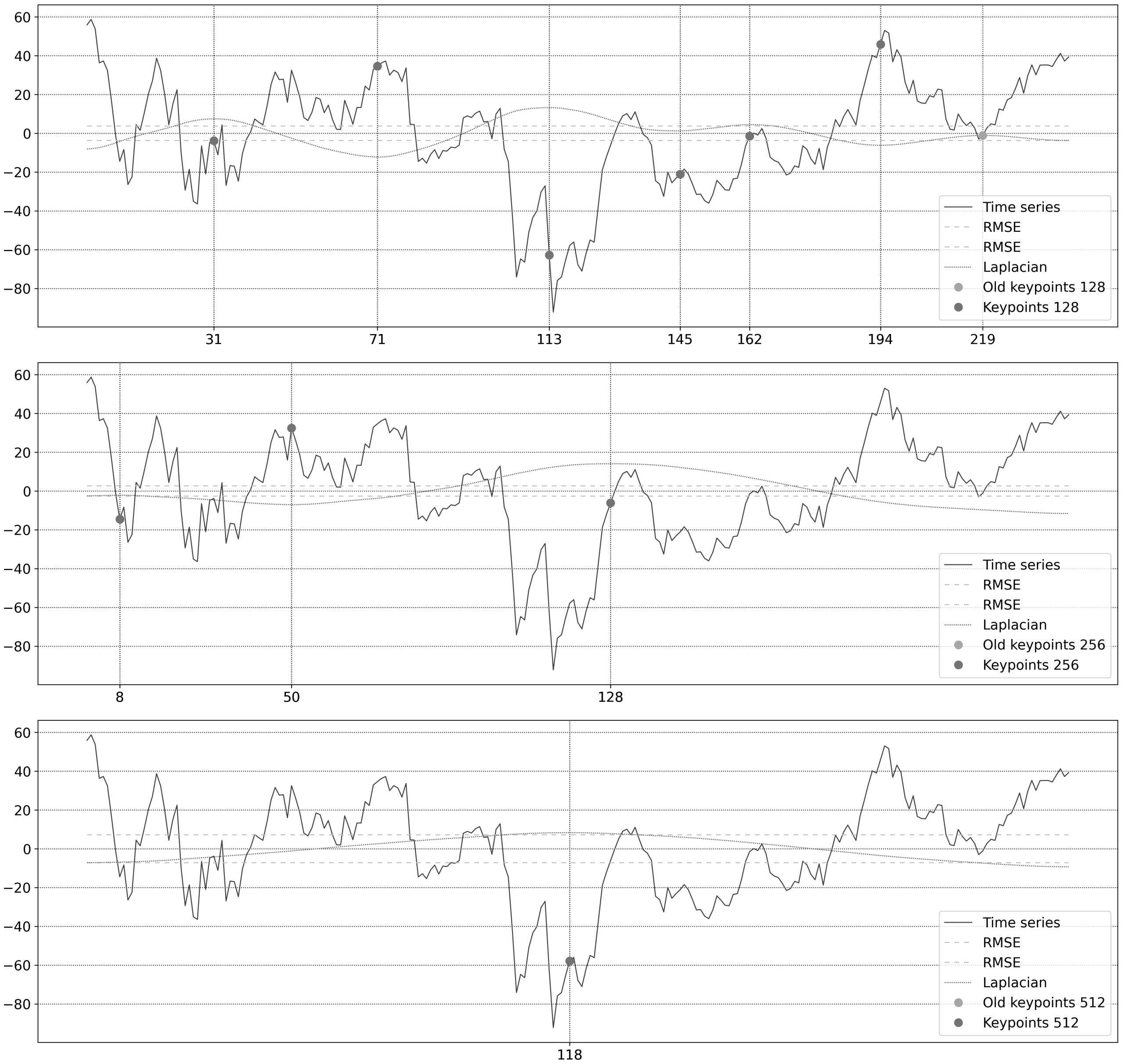

Figure 2: The time series with stock indices PX (Prague) and its last three scaled representations of the Laplacian, where the very last *bottom graph* satisfies the stopping criterion. Selected (filtered out) keypoints are marked with bold dots.

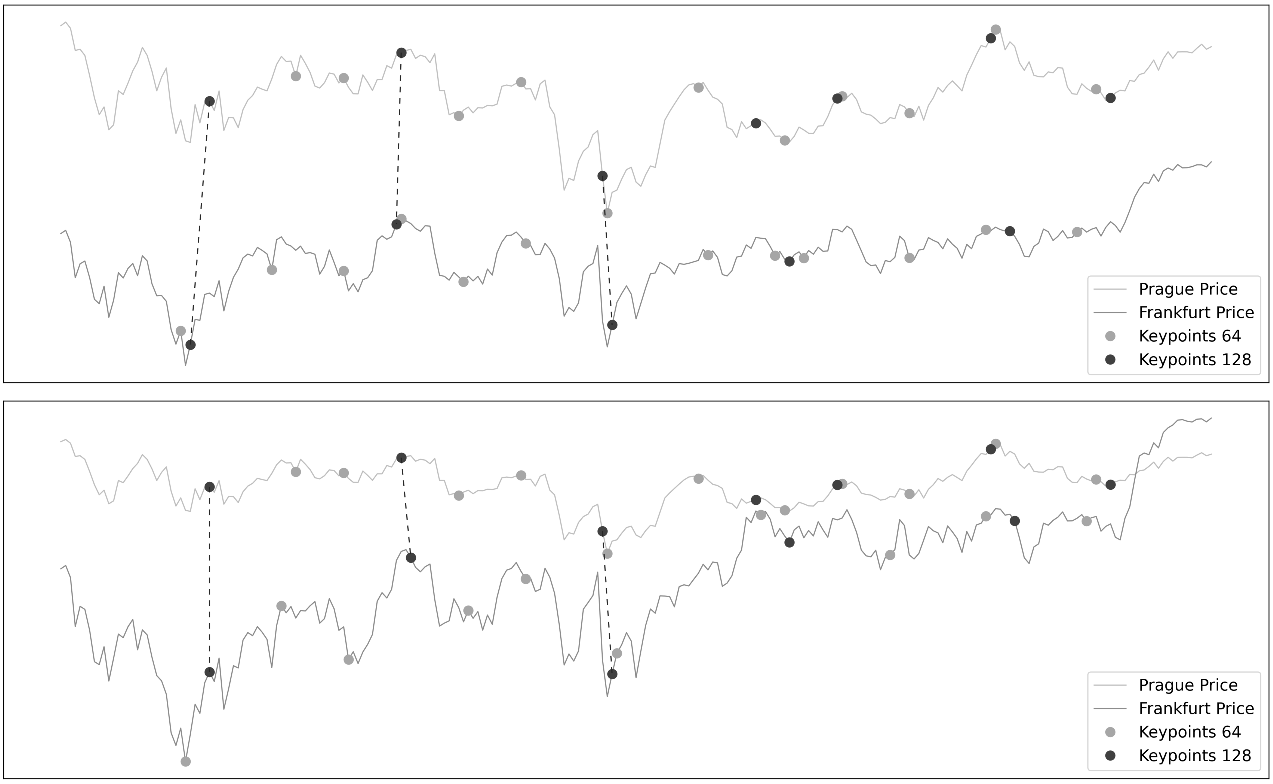

Figure 3: The results of matches between principal keypoints of time series with stock indices PX (Prague), FCHI (Paris), and GDAXI (Frankfurt). The stock indices PX are considered as tested and compared against the reference stock indices FCHI *top image* and GDAXI *bottom image*.