MOTION SMOOTHING STRATEGIES FOR VIDEO STABILIZATION

Javier Sanchez* and Jean-Michel Morel**

- * Universidad de Las Palmas Gran Canarias
- ** Ecole Normale Supérieure Paris-Saclay

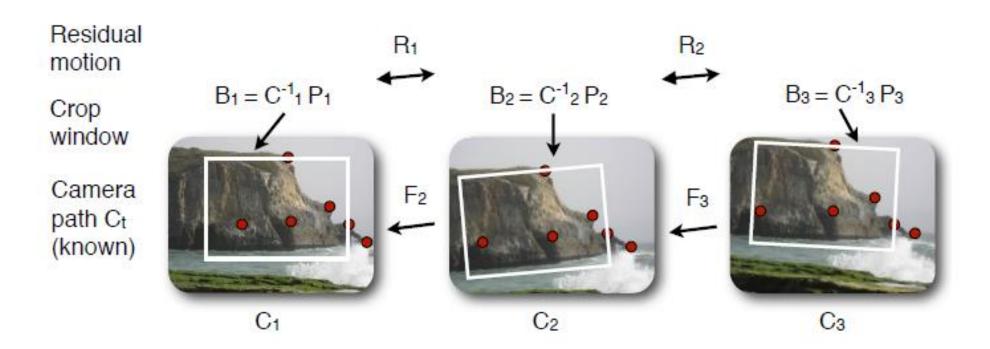
In this work, we propose a unified mathematical analysis and classification of motion smoothing strategies.

We adopt the classic assumption that the apparent motion of the scene is mainly induced by the camera's motion and can therefore be captured through a set of global parametric models.

We show that the best choice yields a *scale-space* of the camera's ego-motion parameters. Displaying this scale space on examples, we show how it is highly characteristic of the camera's path, permitting to detect for example periodic ego-motions like walk or run.

Current stabilization approaches employ key-point feature tracking and linear motion estimation in the form of 2D transformations, or use Structure from Motion (SfM) to estimate the original camera path. From this original shaky camera path, a new smooth camera path is estimated by either smoothing the linear motion models [1] to suppress high frequency jitter, or fitting linear camera paths [2] augmented with smooth changes in velocity to avoid sudden jerks. If SfM is used to estimate the 3D path of the camera, more sophisticated smoothing and linear fits for the 3D motion may be employed [3]. L1-optimal camera paths to generate stabilized. In [4] the method computes camera paths that are composed of constant, linear and parabolic segments mimicking the camera motions employed by professional cinematographers by minimizing the first, second, and third derivatives of the resulting camera path.

- [1] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. Full-frame video stabilization with motion inpainting. IEEE Trans. Pattern Anal. Mach. Intell., 2006.
- [2] M. L. Gleicher and F. Liu. Re-cinematography: Improving the camerawork of casual video. ACM Trans. Mult. Comput. Commun. Appl., 2008.
- [3] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. In ACM SIGGRAPH, 2009.
- [4] M. Grundmann V. Kwatra I. Essa. Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths, preprint.



Questions: How to chose the parametric model

How to smooth the camera path How to fix boundary conditions

How to use the stabilization transform for motion analysis

F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. In ACM SIGGRAPH, 2009. (3D reconstruction by structure from motion)

Formalization

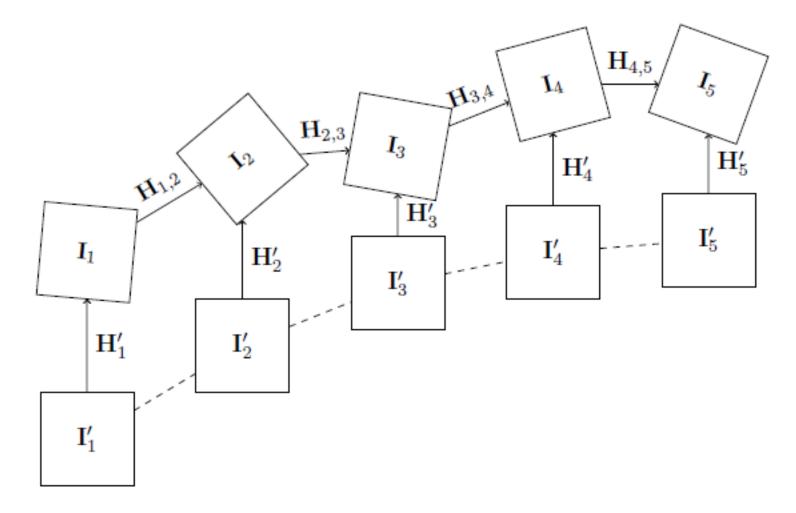


Figure 1. We denote by $\{\mathbf{I}_i\}_{i=1,..N}$ the sequence of images, $\{\mathbf{H}_{j,j+1}\}_{j=1,..N-1}$ the computed transformations between consecutive images, $\{\mathbf{I}_i'\}_{i=1,..N}$ the stabilized sequence and $\{\mathbf{H}_i'\}_{i=1,..N}$ the stabilizing transformations.

Formalization

Transform	Parameters (p)	Н
Translation	(t_x,t_y)	$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix}$
Euclidean	$(t_x,t_y, heta)$	$ \begin{pmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} $
Similarity	(t_x, t_y, a, b)	$\left(\begin{array}{ccc} 1+a & -b & t_x \\ b & 1+a & t_y \\ 0 & 0 & 1 \end{array} \right)$
Affinity	$(t_x, t_y, a_{11}, a_{12}, a_{21}, a_{22})$	0 0 1 /
Homography	$(h_{11}, h_{12}, h_{13}, \cdots, h_{32})$	$\begin{pmatrix} 1 + h_{11} & h_{12} & h_{13} \end{pmatrix}$

Table 1

Planar transformations.

Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths: uses only affine (or similarity, but how to decide?)

Matthias Grundmann Vivek Kwatra Irfan Essa

Formalization

4.1. Compositional approach. Definition 4.1. We define the compositional approach as
the process of calculating the transformations that completely compensate the images with
respect to a given reference frame. The transformations are obtained by composing the relative
motions between successive frames by

(4.1)
$$\mathbf{H}'_{i} = \mathbf{H}_{1,i} := \prod_{j=2}^{N} \mathbf{H}_{j-1,j} = \mathbf{H}_{i-1,i} \mathbf{H}_{i-2,i-1} \cdots \mathbf{H}_{2,3} \mathbf{H}_{1,2}.$$

Definition 4.2. We start from the compositional transition homographies $\mathbf{H}_{1,i}$ defined by (4.1) and more precisely from their eight matrix elements $\mathbf{H}_{1,i}(p,q)$ where p and q are the row and column index respectively. Consider the smooth homographies, $\widetilde{\mathbf{H}}_{1,i}$ which elements $\widetilde{\mathbf{H}}_{1,i}(p,q)$ are obtained through a convolution with a discrete Gaussian function of the series $\mathbf{H}_{1,i}(p,q)$ as

(4.2)
$$\widetilde{\mathbf{H}}_{1,i}(p,q) := (G_{\sigma} * \mathbf{H}_{1,\cdot})_i(p,q) = \sum_{j \in \mathcal{N}_i} G_{\sigma}(j-i)\mathbf{H}_{1,j}(p,q),$$

with $G_{\sigma}(x) := We^{\frac{x^2}{2\sigma^2}}$, $\mathcal{N}_i = \{j : i - k \leq j \leq i + k\}$, and $W := 1/\sum_{i=-k}^{i=k} e^{\frac{i^2}{2\sigma^2}}$ a normalizing coefficient. The compositional smoothing approach is defined by the following respective rectifying transformations and image stabilized sequence

$$\mathbf{H}_i' = \mathbf{H}_{1,i}\widetilde{\mathbf{H}}_{1,i}^{-1} ; \quad \mathbf{I}_i'(\mathbf{x}) := \mathbf{I}_i(\mathbf{H}_i'\mathbf{x}).$$







Figure 3. Compositional approach and the problem of changing the focal length or moving the camera forwards. On the left, the first frame of the sequence; on the center, another frame in the middle of the sequence; on the right, the compensated image for that frame. The problem with the compositional approach is that there is a steady zoom in as the camera moves forward. Thus, compensating the motion by registering the frames up to the first one tends to reduce more and more the current frame. Indeed, this is roughly equivalent to registering the current frame to the first frame.

Definition 4.4.Let $\{\mathbf{H}_{1,i}\}$ be a set of transformations from each frame i in a sequence to the reference frame 1 in a time interval between 1 and N.

We define constant boundary conditions as

(4.9)
$$\mathbf{H}_{1,j} := \begin{cases} \mathbf{H}_{1,1} & \text{if } j < 1 \\ \mathbf{H}_{1,N} & \text{if } j > N \end{cases}.$$

We define Neumann boundary conditions as

(4.10)
$$\mathbf{H}_{1,j} := \begin{cases} \mathbf{H}_{1,-j+1} & \text{if } j < 1 \\ \mathbf{H}_{1,2N-j} & \text{if } j > N \end{cases}.$$

We define Dirichlet boundary conditions in the range [-2N, 3N] as

(4.11)
$$\widetilde{\mathbf{H}}_{1,1} := \mathbf{H}_{1,1} \quad and \quad \widetilde{\mathbf{H}}_{1,N} := \mathbf{H}_{1,N}.$$

(4.12)
$$\mathbf{H}_{1,j} := \begin{cases} \mathbf{H}_{1,j} + 2\mathbf{H}_{1,1} - 2\mathbf{H}_{1,N} & \text{if } -2N \leq j \leq -N \\ 2\mathbf{H}_{1,1} - \mathbf{H}_{1,j} & \text{if } -N \leq j \leq 0 \\ 2\mathbf{H}_{1,N} - \mathbf{H}_{1,N-j} & \text{if } N \leq j \leq 2N \\ \mathbf{H}_{1,j} + 2\mathbf{H}_{1,N} - 2\mathbf{H}_{1,1} & \text{if } 2N \leq j \leq 3N \end{cases}.$$

Definition 4.5. Given the following convolution with a Gaussian function

(4.13)
$$\widetilde{\mathbf{H}}_{i,i+1}(p,q) := (G_{\sigma} * {\mathbf{H}})_i(p,q) = \sum_{j \in \mathcal{N}_i} G_{\sigma}(i-j)\mathbf{H}_{j,j+1}(p,q).$$

The compositional local smoothing approach is defined by the following rectifying transformations,

$$\mathbf{H}_{i}' := \prod_{j=1}^{i} \left(\mathbf{H}_{j,j+1} \widetilde{\mathbf{H}}_{j,j+1}^{-1} \right) = \left(\mathbf{H}_{i,i+1} \widetilde{\mathbf{H}}_{i,i+1}^{-1} \right) \left(\mathbf{H}_{i-1,i} \widetilde{\mathbf{H}}_{i-1,i}^{-1} \right) \cdots$$

$$\cdots \left(\mathbf{H}_{2,3} \widetilde{\mathbf{H}}_{2,3}^{-1} \right) \left(\mathbf{H}_{1,2} \widetilde{\mathbf{H}}_{1,2}^{-1} \right),$$
(4.14)

and the stabilized image sequence by (3.1), $\mathbf{I}'_i(\mathbf{x}) := \mathbf{I}_i(\mathbf{H}'_i\mathbf{x})$.

Definition 4.7. Consider the following compositions with the previous transformations from the current frame

(4.17)
$$\mathbf{H}_{i,j:j< i} = \prod_{l=j}^{i-1} \mathbf{H}_{l+1,l} = \mathbf{H}_{j+1,j} \mathbf{H}_{j+2,j+1} \cdots \mathbf{H}_{i-1,i-2} \mathbf{H}_{i,i-1},$$

with $\mathbf{H}_{l+1,l} = \mathbf{H}_{l,l+1}^{-1}$, and the following as

(4.18)
$$\mathbf{H}_{i,j:j>i} = \prod_{l=i}^{j} \mathbf{H}_{l,l+1} = \mathbf{H}_{j-1,j} \mathbf{H}_{j-2,j-1} \cdots \mathbf{H}_{i-1,i} \mathbf{H}_{i,i+1}.$$

in a neighborhood, $\mathcal{N}_i = \{j : i - k \leq j \leq i + k\}$, around frame i. Define the convolution with a Gaussian function defined in this interval by

(4.19)
$$\widetilde{\mathbf{H}}_{i}(p,q) := \sum_{j=i-k}^{i+k} G_{\sigma}(i-j)\mathbf{H}_{i,j}(p,q).$$

with $\mathbf{H}_{i,i} := \mathbf{Id}$. Then the local matrix-based stabilization is defined by the rectifying transformations $\mathbf{H}'_i := \widetilde{\mathbf{H}}_i^{-1}$ and the stabilized image sequence by (3.1), $\mathbf{I}'_i(\mathbf{x}) := \mathbf{I}_i(\mathbf{H}'_i\mathbf{x})$.

Definition 4.9. Consider a set of points in the current frame i, $\{\mathbf{x}_i^p\}^{p=1,...,N}$, and a trajectory for each point in a temporal neighborhood, $\mathcal{N}_i = \{j : i-k \leq j \leq i+k\}$, given by the following points:

$$(4.23) {\mathbf{x}_{j}^{p}}_{i-k \le j < i} := \left(\prod_{l=j}^{i-1} \mathbf{H}_{l+1,l}\right) \mathbf{x}_{i}^{p} = \mathbf{H}_{j+1,j} \mathbf{H}_{j+2,j+1} \cdots \mathbf{H}_{i-1,i-2} \mathbf{H}_{i,i-1} \mathbf{x}_{i}^{p},$$

on the left and

$$(4.24) \{\mathbf{x}_{j}^{p}\}_{i < j \le i+k} := \left(\prod_{l=i}^{j-1} \mathbf{H}_{l,l+1}\right) \mathbf{x}_{i}^{p} = \mathbf{H}_{j-1,j} \mathbf{H}_{j-2,j-1} \cdots \mathbf{H}_{i+1,i+2} \mathbf{H}_{i,i+1} \mathbf{x}_{i}^{p},$$

on the right. We define the convolution of each trajectory with a Gaussian function by

$$\{\widetilde{\mathbf{x}}_{i}^{p}\} := \left(G_{\sigma} * \{\mathbf{x}_{j}^{p}\}_{j=i-k,..,i+k}^{p=1,..,N}\right)_{i}.$$

The local point-based smoothing approach is then obtained by the rectifying transformations $\mathbf{H}'_i := \widetilde{\mathbf{H}}_i^{-1}$, where the homography $\widetilde{\mathbf{H}}_i$ is calculated from the original points $\{\mathbf{x}_i^p\}$ to the smoothed set $\{\widetilde{\mathbf{x}}_i^p\}$. Finally the stabilized image sequence is obtained by (3.1) as $\mathbf{I}'_i(\mathbf{x}) :=$ $\mathbf{I}_i(\mathbf{H}'_i\mathbf{x})$. Definition 4.11. Consider a fixed set of points in each current frame i, $\{\mathbf{x}^p\}^{p=1,..,N}$, and define the virtual trajectory for each one of these points in a temporal neighborhood $\mathcal{N}_i = \{j : i-k \leq j \leq i+k\}$, defined by $\mathbf{x}_i^p = \mathbf{x}^p$ and

(4.30)
$$\{\mathbf{x}_{j}^{p}\}_{i-k \leq j < i} := \mathbf{x}^{p} + \sum_{l=i}^{i-1} (\mathbf{H}_{l+1,l} \mathbf{x}^{p} - \mathbf{x}^{p}),$$

on the left and

(4.31)
$$\{\mathbf{x}_{j}^{p}\}_{i < j \le i+k} := \mathbf{x}^{p} + \sum_{j=1}^{j-1} (\mathbf{H}_{l,l+1} \mathbf{x}^{p} - \mathbf{x}^{p}),$$

on the right. We define the stabilized position of \mathbf{x}_i^p by a Gaussian convolution, by

(4.32)
$$\{\widetilde{\mathbf{x}}_{i}^{p}\} := \left(G_{\sigma} * \{\mathbf{x}_{j}^{p}\}\right)_{i} = \sum_{j=-k}^{j=k} G_{\sigma}(j) \mathbf{x}_{i-j}^{p}.$$

The linear local point-based smoothing approach is then obtained by the rectifying transformations $\mathbf{H}'_i := \widetilde{\mathbf{H}}_i^{-1}$, where the homography $\widetilde{\mathbf{H}}_i$ is calculated from the original points $\{\mathbf{x}_i^p = \mathbf{x}^p\}$ to the stabilized set at frame i, $\{\widetilde{\mathbf{x}}_i^p\}$. Finally the stabilized image sequence is obtained by (3.1) as $\mathbf{I}'_i(\mathbf{x}) := \mathbf{I}_i(\mathbf{H}'_i\mathbf{x})$.

Crop & Zoom:













First row, several frames of a video with the inscribed rectangle obtained through Algorithm 28; second row, the same frames after the Crop&Zoom process. The coordinates of the crop rectangle are (20.63, 22.28)–(358.54, 180.43), and the computed scale factor is s = 0.760315.

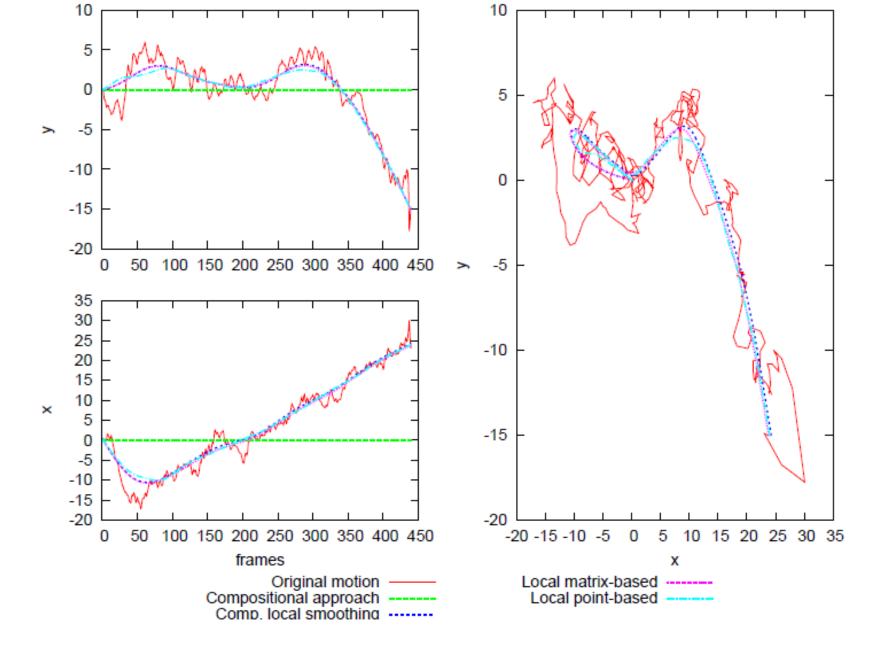


Figure 9. Cournot Building trajectory. These graphics show the trajectory of the central pixel after successively applying the transformations. We show the trajectory for each smoothing strategy. The left graphic show the x component and the right the y component of the tracked point.

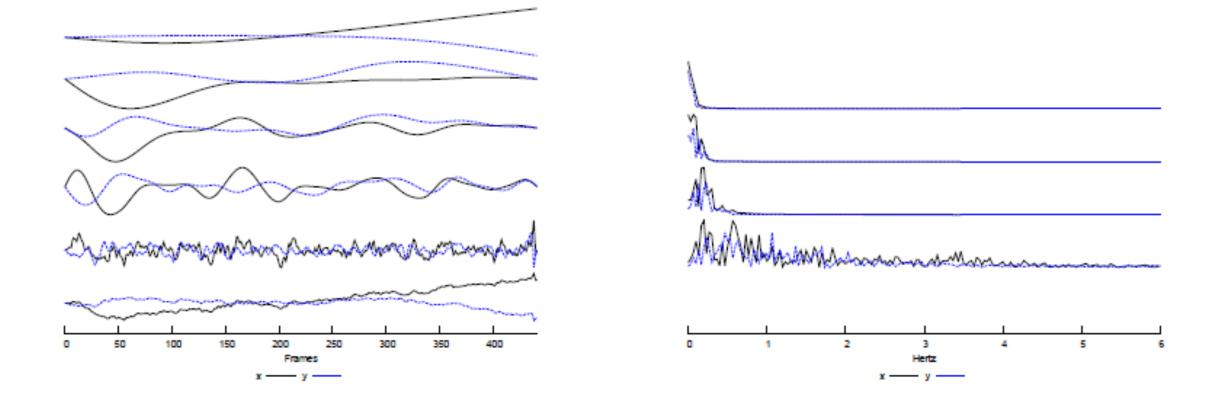


Figure 12. Cournot Building linear scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 9; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy, and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.

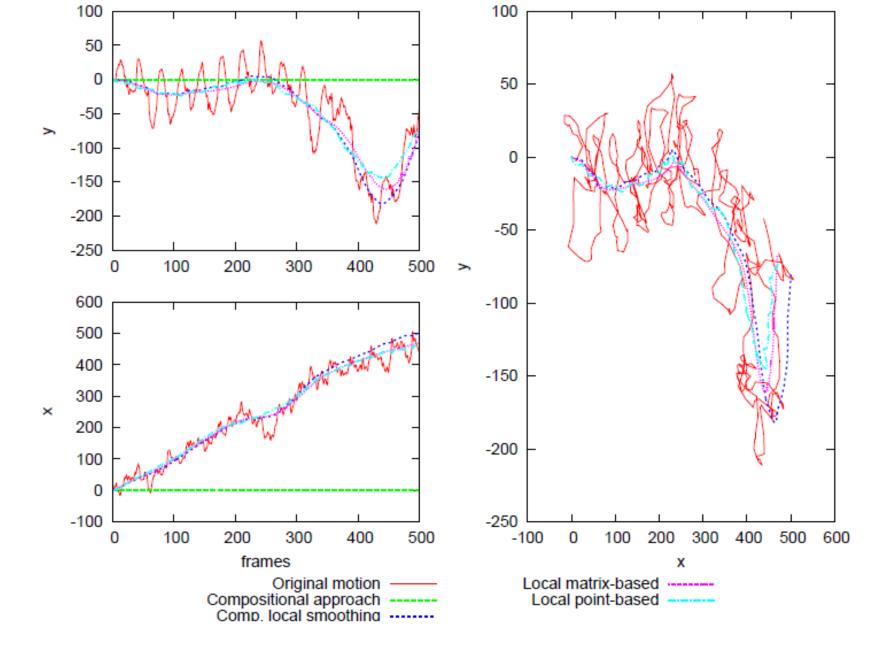


Figure 42. Walking trajectory. These graphics show the trajectory of the central pixel after successively applying the transformations. We show the trajectory for each smoothing strategy. The left graphic show the x component and the right the y component of the tracked point (frames 1000 to 3000).

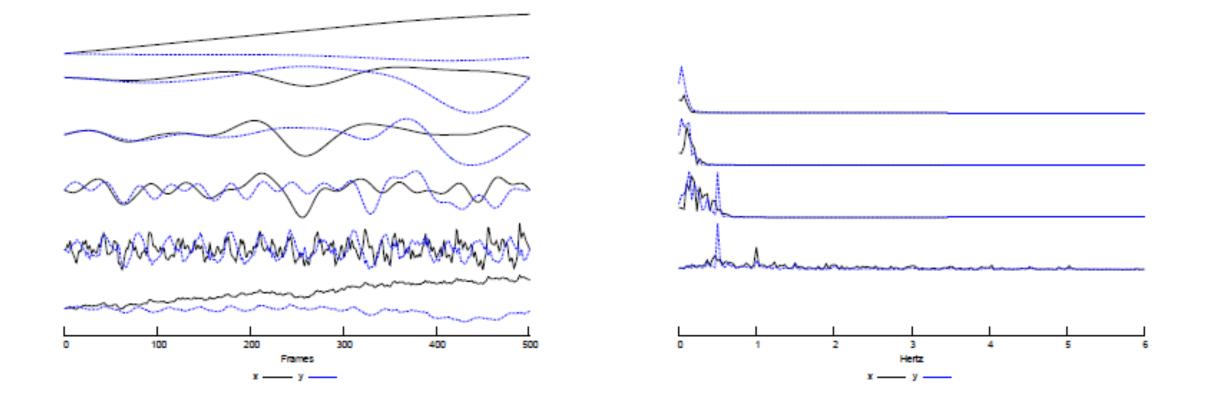
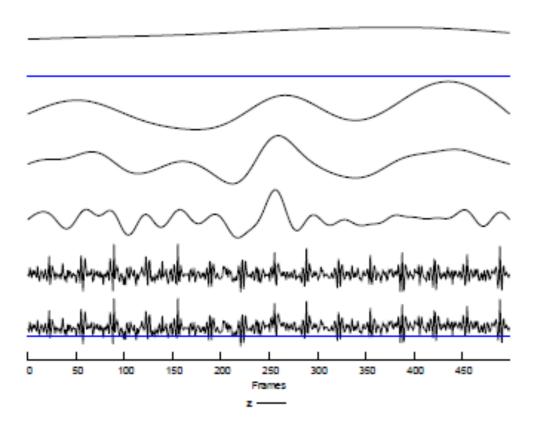


Figure 45. Walking linear scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 9; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy, and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.



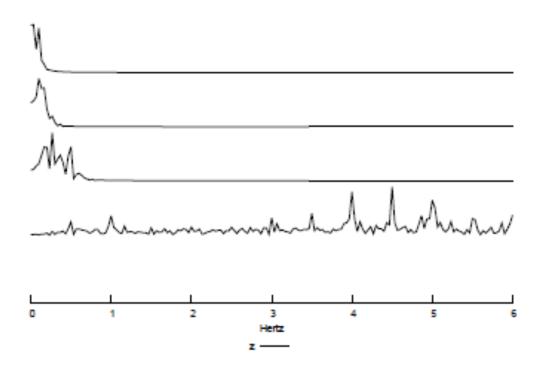
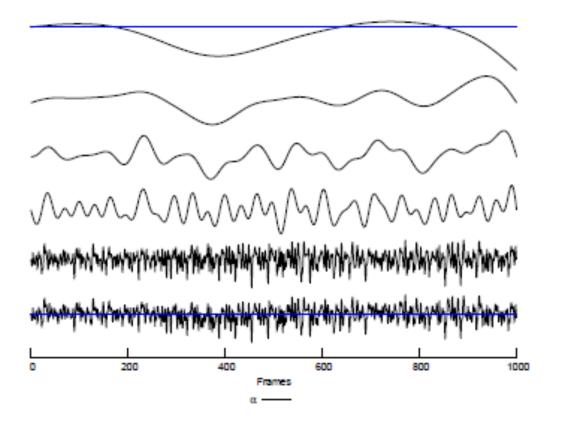


Figure 47. Walking linear scale-space analysis (zoom). We plot $z(i) = Area(\mathbf{H}_i)/(n_x \cdot n_y)$. The average zoom is 1.009399 for the bottom graphic on the left and 1.030396 for the top graphic on the left.



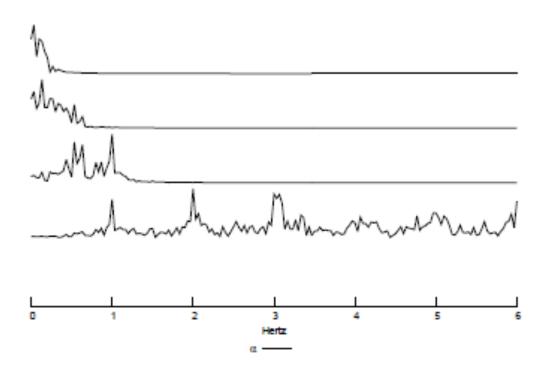


Figure 49. Walking linear scale-space analysis (rotation). We plot $\alpha(i) = angle(\mathbf{H}_i)$, where the angle is obtained from the projection of two points $(0, n_y/2), (n_x, n_y/2)$ through \mathbf{H}_j . The average rotation is 0.000414 for the bottom graphic on the left and 0.001368 for the top graphic on the left.

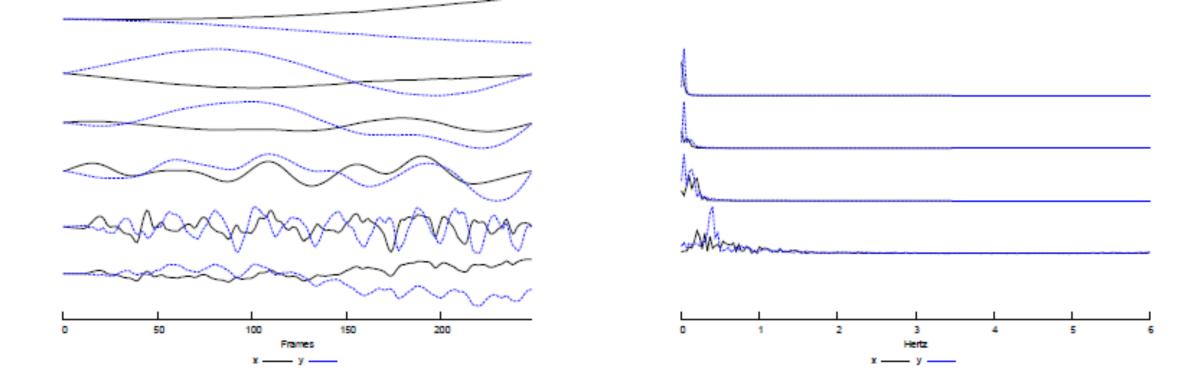


Figure 55. Descending stairs scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 53; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.

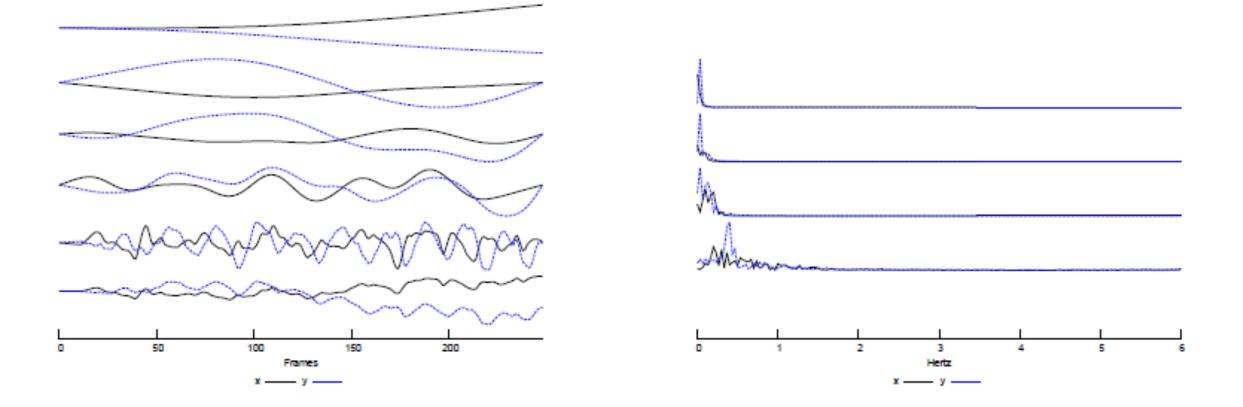
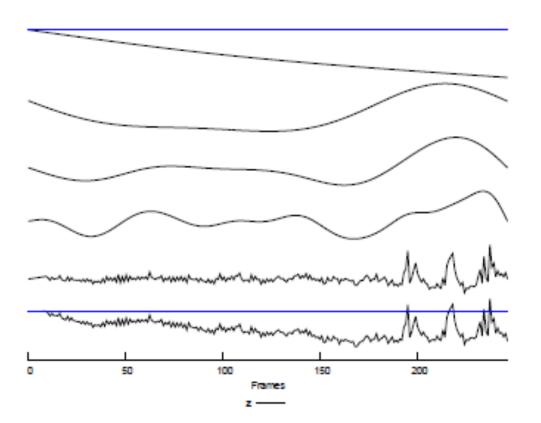


Figure 56. Descending stairs linear scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 9; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy, and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.



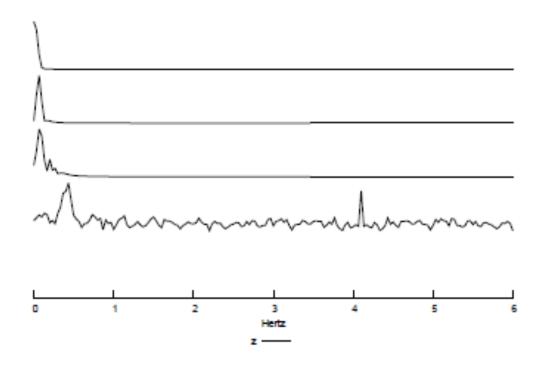


Figure 58. Descending stairs linear scale-space analysis (zoom). We plot $z(i) = Area(\mathbf{H}_i)/(n_x \cdot n_y)$. The average zoom is 0.996083 for the bottom graphic on the left and 0.996177 for the top graphic on the left.

Figure 59. Descending stairs scale-space analysis (rotation). We plot $\alpha(i) = angle(\mathbf{H}_i)$, where the angle is obtained from the projection of two points $(0, n_y/2), (n_x, n_y/2)$ through \mathbf{H}_j . The average rotation is 0.000138 for the bottom graphic on the left and 0.000109 for the top graphic on the left.

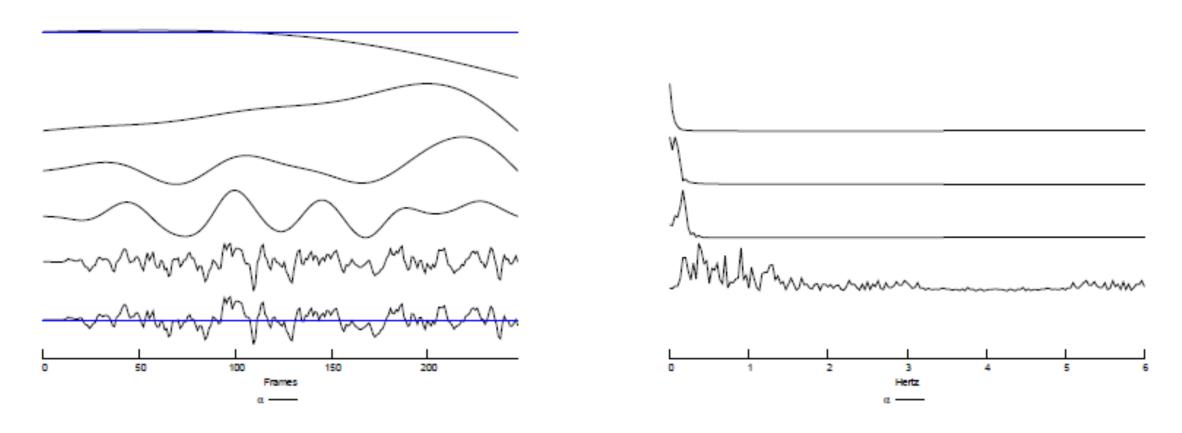


Figure 60. Descending stairs linear scale-space analysis (rotation). We plot $\alpha(i) = angle(\mathbf{H}_i)$, where the angle is obtained from the projection of two points $(0, n_y/2), (n_x, n_y/2)$ through \mathbf{H}_j . The average rotation is 0.000138 for the bottom graphic on the left and 0.000109 for the top graphic on the left.

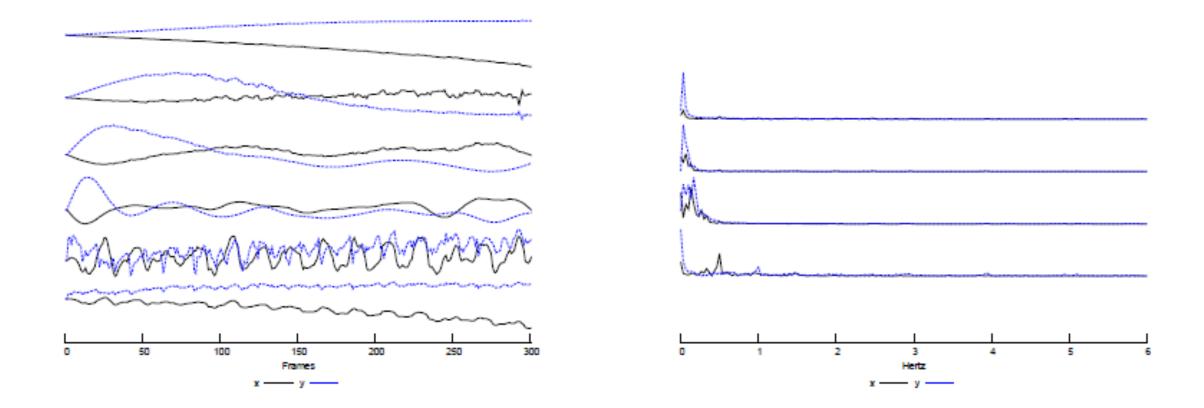


Figure 63. Man running with a GoPro in his head scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 53; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.

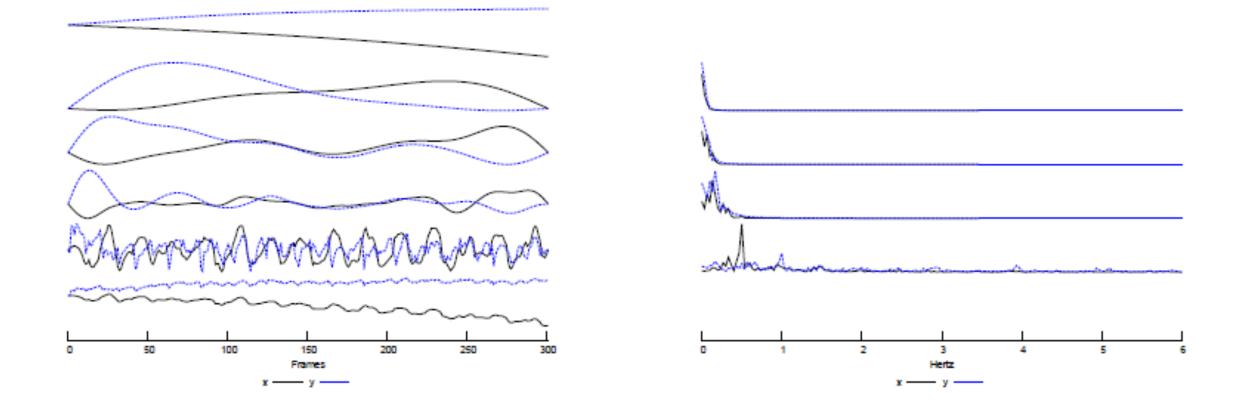
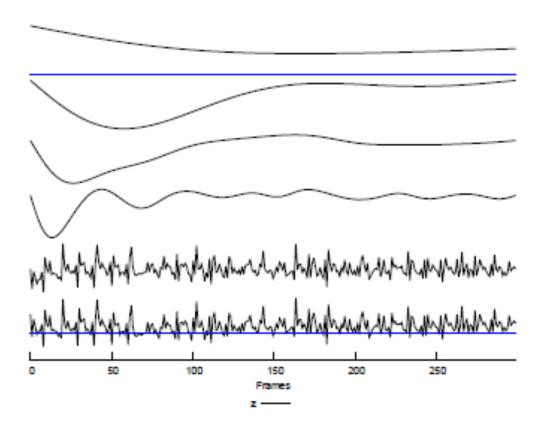


Figure 64. Man running with a GoPro in his head linear scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 9; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy, and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.



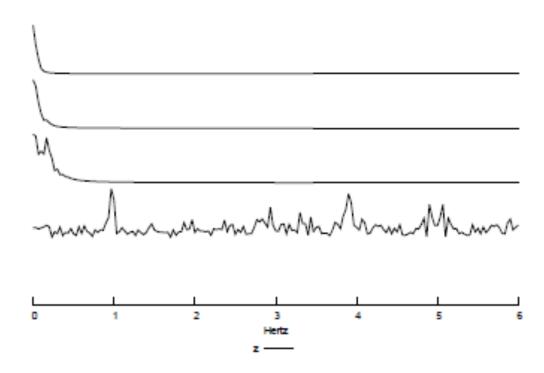
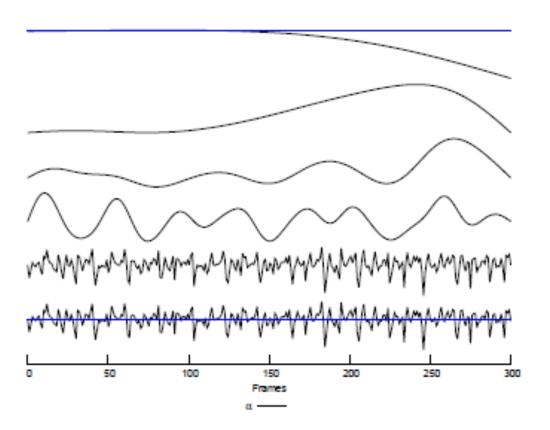


Figure 66. Man running with a GoPro in his head linear scale-space analysis (zoom). We plot $z(i) = Area(\mathbf{H}_i)/(n_x \cdot n_y)$. The average zoom is 1.013419 for the bottom graphic on the left and 1.012940 for the top graphic on the left.



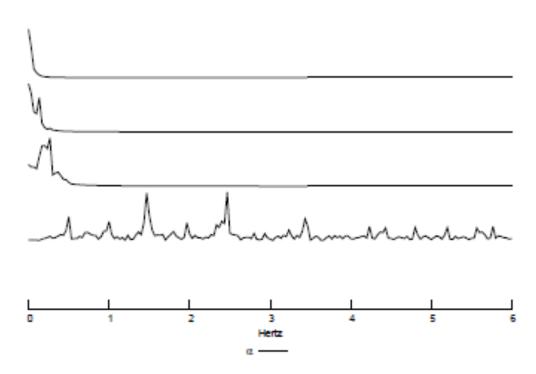
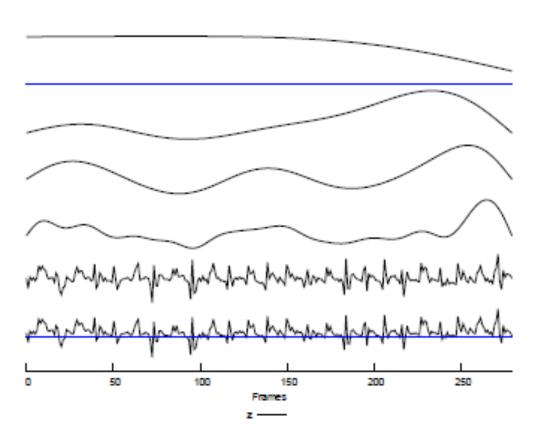


Figure 68. Man running with a GoPro in his head linear scale-space analysis (rotation). We plot $\alpha(i) = angle(\mathbf{H}_i)$, where the angle is obtained from the projection of two points $(0, n_y/2), (n_x, n_y/2)$ through \mathbf{H}_i . The average rotation is 0.000038 for the bottom graphic on the left and 0.000029 for the top graphic on the left.



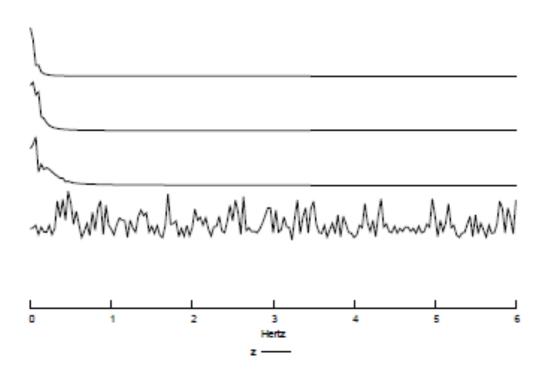
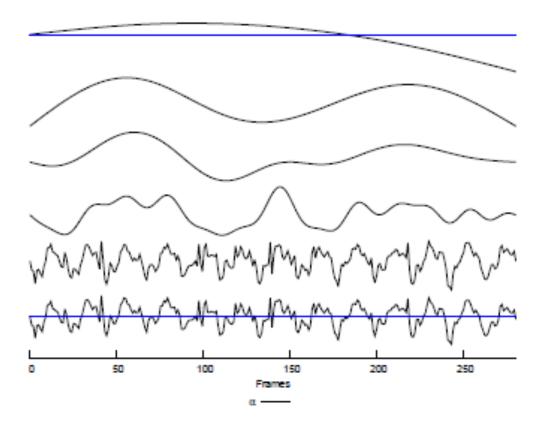


Figure 72. Man running with a GoPro in his chest linear scale-space analysis (zoom). We plot $z(i) = Area(\mathbf{H}_i)/(n_x \cdot n_y)$. The average zoom is 1.026854 for the bottom graphic on the left and 1.038753 for the top graphic on the left.



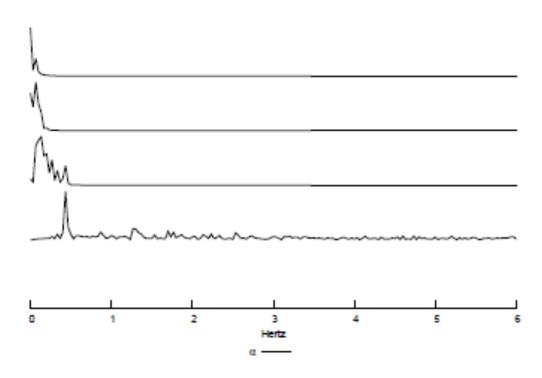


Figure 74. Man running with a GoPro in his chest linear scale-space analysis (rotation). We plot $\alpha(i) = angle(\mathbf{H}_i)$, where the angle is obtained from the projection of two points $(0, n_y/2), (n_x, n_y/2)$ through \mathbf{H}_j . The average rotation is 0.001036 for the bottom graphic on the left and 0.000678 for the top graphic on the left.

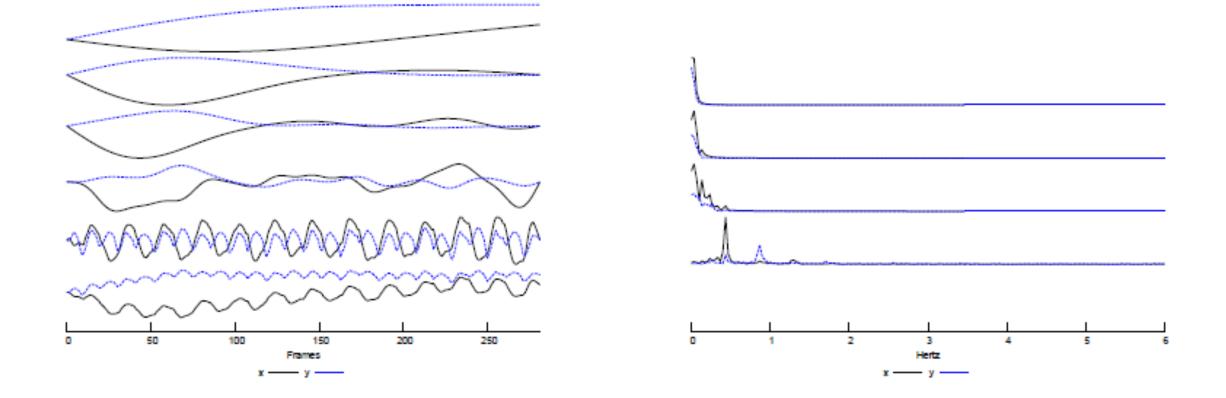


Figure 70. Man running with a GoPro in his chest linear scale–space analysis. Left figure: In the bottom, we show the trajectory of the central point using the original transformations, as in Figure 9; in the following graphic above, we show the difference between the smoothed trajectory obtained with the local matrix–based smoothing strategy, and $\sigma := 10$, minus the original trajectory; the following graphic above depicts the difference between the solution with $\sigma := 20$ and $\sigma := 10$; the third graphic, the difference between the solution with $\sigma := 40$ and $\sigma := 20$; and, in the top graphic, the difference between the solution with $\sigma := 80$ and $\sigma := 40$. Right figure: DFT signals of the corresponding graphics on the left. We show the first six seconds??? of the modulus of the signal spectrum. Each graphic depicts the x and y components of the point in each case.