

Enhancing motion trajectories in noisy video sequences

Alexander Kläser, Srdjan Stankovic, Rainer Herpers

► **To cite this version:**

Alexander Kläser, Srdjan Stankovic, Rainer Herpers. Enhancing motion trajectories in noisy video sequences. 4th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT '04), Dec 2004, Rome, Italy. IEEE Computer Society, pp.99–102, 2004, <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1433697>. <10.1109/ISSPIT.2004.1433697>. <inria-00548544>

HAL Id: inria-00548544

<https://hal.inria.fr/inria-00548544>

Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhancing Motion Trajectories in Noisy Video Sequences

Alexander Kläser^{*†}, Srdjan Stanković[†] and Rainer Herpers^{*‡}

^{*}Department of Computer Science

Bonn-Rhein-Sieg University of Applied Sciences, 53757 St. Augustin, Germany

[†]Department of Electrical Engineering

University of Montenegro, 81000 Podgorica, Serbia and Montenegro

[‡]Centre for Vision Research, Department of Computer Science

York University, Toronto, Canada M3J 1P3

Abstract—An approach for enhancing motion trajectories in video sequences is presented. The proposed approach is applied to projections of image data of video sequences and consists of six enhancement steps. The basic idea is to correlate the positive and the negative derivative part with respect to time to the projected image sequence. Both parts are aligned at their correlation maxima and they are element-wise multiplied to suppress noise. Morphological operators are used to obtain a one pixel wide motion trajectory.

Keywords—Image sequence processing, noise suppression, motion trajectory enhancement, motion estimation, background motion, projection, correlation, morphological operators, SLIDE algorithm.

I. INTRODUCTION

The time-frequency representation is a very important tool for a wide range of applications. Combined with the SLIDE (subspace-based line detection) algorithm [1], a method for the estimation of line parameters, it builds a powerful and very precise method for the estimation of time-varying velocities of moving objects in video sequences. This topic has recently been addressed in [2]–[4]. Therein, the dimensionality of video sequences is reduced by x - and y -axis projection, and a modified version of the SLIDE algorithm is applied together with time frequency-tools. In that way the velocity of one or even of multiple moving objects can be estimated. So far only artificial sequences or simple real sequences with a small amount of noise have been investigated. Therefore a comprehensive testing of this method to real data containing a larger amount of noise was necessary.

Real sequences, however, contain noise originating from the camera sensor as well as from background motion. In this case velocity estimation becomes inaccurate if one applies only the SLIDE algorithm and time-frequency tools. Additional methods have to be employed in order to improve the estimation by enhancing motion trajectories and suppressing noise.

In this contribution time-varying velocity estimation in real video sequences is considered. We propose a sequential trajectory enhancement approach that consists of:

- the computation of the negative and positive derivative part with respect to time to the projected image sequence

- a temporal thresholding of both parts
- the correlation of the negative and positive derivative part followed by an element-wise multiplication, whereby both parts are aligned at their correlation maxima at each time instant
- a thinning step by shifting and element-wise multiplication together with the application of morphological operators

In order to reduce the complexity of the given task, it is assumed that only one moving object is present in the sequence. However, an extension to multiple moving objects is generally possible.

II. AXIS PROJECTION

By using the SLIDE algorithm [1], an image sequence, $i(x, y, t)$, is projected onto its coordinate axes. Its projections for the x - and y -axis are defined according to [2, Eq. 3] as

$$X_{x,t} = X(x, t) = \sum_y i(x, y, t) \quad (1)$$

$$Y_{y,t} = Y(y, t) = \sum_x i(x, y, t). \quad (2)$$

Since working with digital data (pixels and frames), let x , y and t be defined as elements of \mathbb{N}_0^+ . In the sequel, only the projection onto the x -axis will be investigated in more detail for motion analysis in the x -direction. Note that same assumptions hold for the motion analysis in y -direction and can be applied likewise. For more details concerning the SLIDE algorithm, we refer the reader to [1].

III. TRAJECTORY ENHANCEMENT

A. Partial Derivative

Static, non moving background in the projected signal is eliminated by applying the partial derivation with respect to time, $\partial X_{x,t}/\partial t$. For the discrete case we use the Sobel operator as approximation. The Sobel operator is a combination of a weighted averaging filter $[1 \ 2 \ 1]'$ and a symmetrical difference filter $[1 \ 0 \ -1]$. In the presented application, the averaging filter refers to the spatial position, while the difference filter is associated with the position in time. Due to the design of the kernel, the filter result will be smoothed, which is important

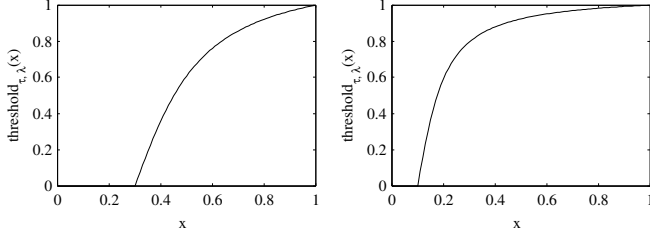


Fig. 1. Illustration of the threshold function with (left) $\lambda = 5$ and $\tau = 0.3$ and (right) $\lambda = 12$ and $\tau = 0.1$.

for the shifting and multiplication step described later. Thus the derivation can be approximated as:

$$\frac{\partial X_{x,t}}{\partial t} \simeq \dot{X}_{x,t} = \left[X_{x-1,t+1} - X_{x-1,t-1} + 2(X_{x,t+1} - X_{x,t-1}) + X_{x+1,t+1} - X_{x+1,t-1} \right]. \quad (3)$$

\dot{X} denotes the discrete approximation of the partial derivative with respect to time, x and t specify the sample position.

It is assumed that the object size is small enough for the object to be present as a whole in the sequence, that its texture is rather uniform and that its contrast to the background is high enough. Then, the partial derivative with respect to time will contain a positive as well as a negative response of the trajectory. If it is additionally assumed that the object's shape and its texture do not vary significantly, both responses will describe a similar trajectory. For later use of this feature, the positive and negative signal part will be distinguished as

$$\dot{X}_{x,t}^+ = \begin{cases} \dot{X}_{x,t} & \dot{X}_{x,t} \geq 0 \\ 0 & \dot{X}_{x,t} < 0 \end{cases} \quad (4)$$

$$\dot{X}_{x,t}^- = \begin{cases} -\dot{X}_{x,t} & \dot{X}_{x,t} \leq 0 \\ 0 & \dot{X}_{x,t} > 0 \end{cases}. \quad (5)$$

B. Temporal Signal Enhancement

Since the noise reduction will be based on the correlation of the negative and positive derivative part, the intensity of the trajectory has to be adjusted in weak and strong parts of the signal independently. Therefore, a non-linear threshold function is introduced. In strong and clear parts, noise can then directly be suppressed since the trajectory can easily be separated from the noise. In weak parts only a small amount of noise will be suppressed, yet the intensity of the whole signal will be adjusted. However, it is assumed that the correlation between noise in both parts is sufficiently less strong than the correlation between trajectories. Therefore, noise can be reduced by a combination of both parts.

We introduce the threshold function as

$$\text{threshold}_{\tau, \lambda}(x) = \frac{1}{\arctan(\lambda(1-\tau))} \begin{cases} 0 & x < \tau \\ \arctan(\lambda(x-\tau)) & x \geq \tau \end{cases}, \quad \text{with } 0 \leq x \leq 1. \quad (6)$$

The function defines a non-linear fading-out (see Figure 1), with τ specifying the threshold value, and λ the rapidness of the fading-out process. In this context $\lambda = 12$ is a reasonable value. For each point in time, τ is calculated as:

$$\tau_t^+ = \max(\sigma_t^+, \vartheta \cdot \sigma^+), \quad (7)$$

where σ_t^+ is the *temporal* standard deviation and σ^+ the standard deviation of the whole sequence. The temporal standard deviation at the time instant t is given by

$$\sigma_t^+ = \sqrt{\frac{1}{|x|-1} \sum_{x=0}^{|x|-1} (\dot{X}_{x,t}^+ - \mu_t^+)^2}, \quad (8)$$

the standard deviation of \dot{X}^+ for the whole video sequence is given by

$$\sigma^+ = \sqrt{\frac{1}{(|t|-1)(|x|-1)} \sum_{t=0}^{|t|-1} \sum_{x=0}^{|x|-1} (\dot{X}_{x,t}^+ - \mu^+)^2}. \quad (9)$$

μ_t^+ is the temporal mean at the time instant t , it is

$$\mu_t^+ = \frac{1}{|x|} \sum_{x=0}^{|x|-1} \dot{X}_{x,t}^+, \quad (10)$$

μ^+ is the mean of \dot{X}^+ for the whole sequence, defined as

$$\mu^+ = \frac{1}{|t| \cdot |x|} \sum_{t=0}^{|t|-1} \sum_{x=0}^{|x|-1} \dot{X}_{x,t}^+. \quad (11)$$

$|x|$ and $|t|$ represent the numbers of samples for x and t , respectively. The equations are likewise defined for \dot{X}^- . In this work $\vartheta = 0.75$ has been proven as meaningful weight.

By using the defined threshold function, the temporally enhanced signals are written as

$$\dot{X}'_{x,t}^+ = \text{threshold}_{\tau_t^+, \lambda} \left(\frac{\dot{X}_{x,t}^+}{\max_{x,t}(\dot{X}^+)} \right), \quad (12)$$

$$\dot{X}'_{x,t}^- = \text{threshold}_{\tau_t^-, \lambda} \left(\frac{\dot{X}_{x,t}^-}{\max_{x,t}(\dot{X}^-)} \right). \quad (13)$$

Since the threshold function expects its input parameter to be ≥ 0 and ≤ 1 , the signals are scaled such that their maximum value is equal to one.

C. Correlation

In this step, the temporally enhanced positive and negative part, \dot{X}'^+ and \dot{X}'^- , are correlated with respect to the x -axis. As mentioned before, it is assumed that the motion trajectory is present in both parts. By stepwise correlation of the signal parts, the correspondence for both motion trajectories is computed. The correlation is determined by a windowed time interval, similar to the idea of the STFT. It can be written as

$$\dot{X}_{\Delta x,t}^{\times} = \sum_{k=-V}^V \sum_{x=0}^{|x|-1} \text{win}_V(k) \cdot \dot{X}'_{x,t+k}^+ \cdot \dot{X}'_{x+\Delta x,t+k}^-, \quad (14)$$

whereby Δx denotes the x -displacement of \dot{X}'^- for each correlation step. The window function win_V is defined as discrete hamming window, with the size $2V + 1$. In terms of accuracy and saving computational power, a reasonable value is $V = 10$. Δx directly corresponds to the size of the moving object.

D. Combination

Based on \dot{X}^{\times} , \dot{X}'^- is aligned to maxima of the correlation at each time instant. It is then element-wise multiplied with \dot{X}'^+ . This step can be denoted as:

$$\dot{X}'_{x,t} = \sqrt{\sum_{\Delta x \in \Omega_t} \dot{X}'_{x,t}^+ \cdot \dot{X}'_{x+\Delta x,t}^-}, \quad (15)$$

with the set Ω_t , containing local maxima – along the axis of Δx , at the instant t – that are in the range of U and greater than a given threshold:

$$\Omega_t = \left\{ \Delta x \mid \dot{X}_{\Delta x,t}^{\times} \text{ is local maximum along } \Delta x\text{-axis} \right\} \cap \left\{ \Delta x \mid \sqrt{\dot{X}_{\Delta x,t}^{\times}} > \sigma^{\times} + \mu^{\times}, \Delta x < U, \Delta x > -U \right\}. \quad (16)$$

σ^{\times} and μ^{\times} are defined according to Equations (9) and (11), respectively, but for $\sqrt{\dot{X}^{\times}}$. By introducing U , random correlations between the signals that occur at a point $\Delta x > U$ or $\Delta x < -U$ are ignored.

E. Shifting and Multiplication

After the alignment step noise generally is reduced, but noise particles might still be present in the signal. Since noise particles are usually just very few pixels large, a further signal enhancement step by shifting and multiplication of the signal itself is introduced. In this way, noise particles are reduced even further. Since the Sobel operator has been used as approximation for the derivation, it can be assumed that the trajectory is smoothed and thus wider than one pixel. Therefore, the trajectory will be preserved even during this shifting and multiplication step which is denoted as:

$$\dot{X}''_{x,t} = (\dot{X}'_{x-1,t} \cdot \dot{X}'_{x,t} \cdot \dot{X}'_{x+1,t} + \dot{X}'_{x,t-1} \cdot \dot{X}'_{x,t} \cdot \dot{X}'_{x,t+1})^{1/2} \quad (17)$$

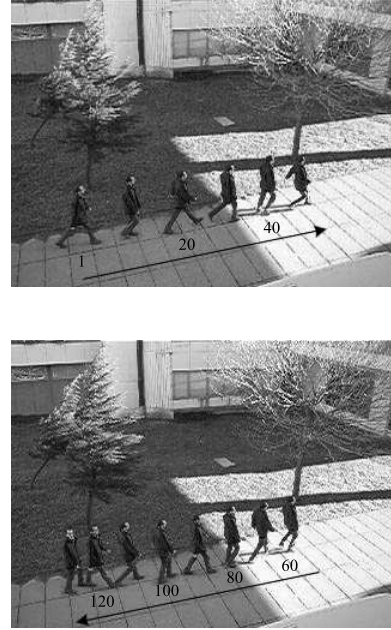


Fig. 2. Sample video sequence with frames (top) 1, 10, 20, 30, 40, 50 and (bottom) 60, 70, 80, 90, 100, 110, 120, 130; (top) the walker is first going from the left to the right, stopping, (bottom) moving backwards to the left and then stopping again. In the background of the sequence trees are waving strongly in the wind.

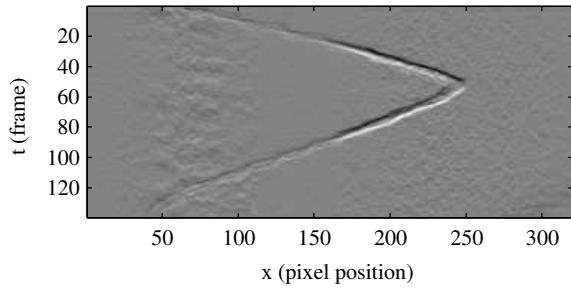
F. Morphological Operators

In order to reduce the motion trajectory to a one pixel thick line, morphological operators are applied to the resulting signal from Equation (17). First a binary closure (denoted as close), then a thinning operation (denoted as thin) are performed until convergence. The final enhanced signal can be introduced as follows:

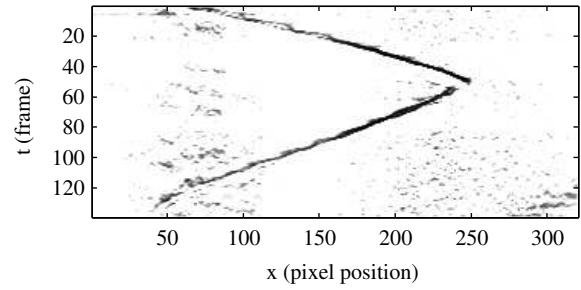
$$\dot{X}_{x,t}^* = \dot{X}''_{x,t} \cdot \text{thin} \left(\text{close} \left(\dot{X}''_{x,t} \right) \right). \quad (18)$$

IV. EXAMPLE

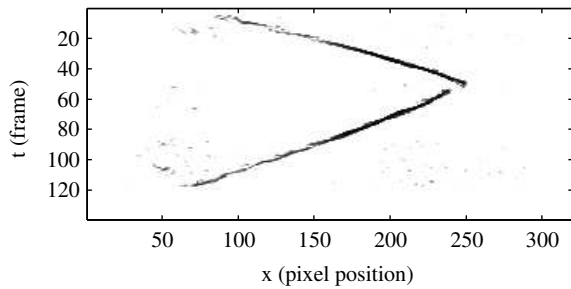
The sample video sequence illustrated in Figure 2 shows a person moving from left to right, stopping (frame 50) and moving backwards to the left. Background motion originates from trees waving in the background. Figure 3(a) depicts the derived x -axis projection of the sequence. Due to the sunlight, the contrast of the walker to the background is much stronger between frames 30 and 80 and hence his motion trajectory is stronger as well. The intensity of the noise between pixel position 50-100 is partly as strong as weak parts of the motion trajectory. In Figures 3(b) and 3(c) the steps of the temporal thresholding and the combination of the derivative parts are illustrated. The combination is based on the outcome of the correlation of both parts, which is given in Figure 4. In the enhanced result (Figure 3(d)), some noise particles are still present, yet overall the method proves to suppress noise and to preserve the trajectory well, even for its weak parts.



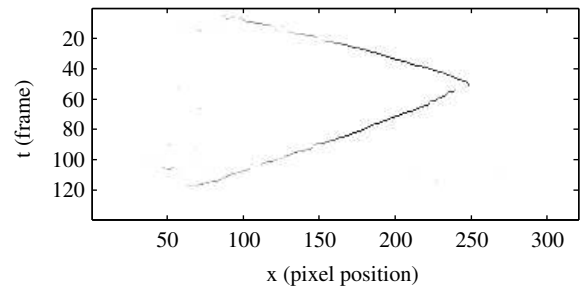
(a) \dot{X} , the partial derivative with respect to time of the x -axis projection of the sequence illustrated in Figure 2 (positive values are displayed in black, negative values in white).



(b) \dot{X}'^+ , the thresholded positive part of the derivative; the motion trajectory as well as some parts of the noise have been enhanced.



(c) \dot{X}' , the combination of the positive and negative derivative parts; the motion trajectory has been preserved while uncorrelated noise has been suppressed.



(d) \dot{X}'' , the resulting signal after shifting and multiplying it by itself as well as the application of morphological operators; the motion trajectory has been thinned out, only some noise particles remain.

Fig. 3. An example for the stepwise noise suppression and trajectory enhancement: (a) the derivative with respect to time, (b) the thresholded positive derivative part, (c) the combination of the positive and negative derivative part, (d) the resulting signal.

V. CONCLUSION

An approach to the suppression of noise and the enhancement of motion trajectories in projections of video sequences has been introduced. In a sample video sequence it could be shown that the applied sequential enhancement methods improves the recovery of motion trajectory well. (Further examples can be found on <http://www.alexklaeser.de/research>.)

ACKNOWLEDGEMENT

A. Kläser gratefully acknowledges the financial support of the DAAD. R. Herpers gratefully acknowledges the financial support of the Bundesministerium für Bildung und Forschung (BMBF-IB), grant WTZ CAN 02/011.

REFERENCES

- [1] H. Aghajan and T. Kailath. Slide: Subspace-based line detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:1057–1073, Nov 1994.
- [2] Igor Djurović and Srdjan Stanković. Estimation of time-varying velocities of moving objects by time-frequency representations. *IEEE Transactions on Image Processing*, 12:550–562, May 2003.

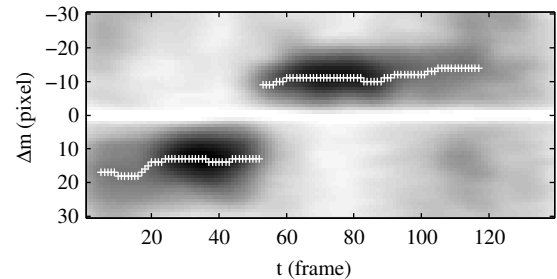


Fig. 4. \dot{X}^\times , the result of the correlation of the positive and negative derivative part (\dot{X}'^+ and \dot{X}'^-) with respect to the x -axis; the sign “+” marks a local maximum above the determined threshold.

- [3] Igor Djurović, Srdjan Stanković, Akira Ohsumi, and Hiroshi Ijima. Estimation of line parameters using slide algorithm and tf representations. In *IEEE International Conference on Electronics, Circuits and Systems, ICECS 2002, Dubrovnik*, pages 1067–1071, Sep 2002.
- [4] Srdjan Stanković and Igor Djurović. Motion parameter estimation by using time frequency representations. *Electronics Letters*, 37(24):1446–1448, Nov 2001.