# OPTICAL-FLOW FEATURES EMPIRICAL MODE DECOMPOSITION FOR MOTION ANOMALY DETECTION

Moacir Ponti<sup>\*</sup>, Tiago S. Nazare<sup>†</sup>

ICMC — University of São Paulo São Carlos/SP, 13566-590, Brazil

## ABSTRACT

In video data analysis of dynamic scenes, temporal characteristics of moving objects play an important role in decisionmaking. However, the temporal consistency of typical features used for video interpretation is low due to the overlap of the spectra of informative video signal component and the stochastic variations perturbing it. We propose a novel method for object motion anomaly detection in video designed to overcome this problem. It is based on empirical mode decomposition. We show in experiments on a benchmarking dataset that the deterministic component of an optical flow feature obtained using the proposed method is able to isolate the periodic behaviour of the motion from the stochastic values, facilitating much simpler analysis of the motion patterns and achieving impressive anomaly detection performance.

*Index Terms*— Signal decomposition, surveillance, video processing

# 1. INTRODUCTION

In video recognition applications, in particular when trying to to detect anomalies or unusual events in video surveillance, it is important to capture both spatial and temporal characteristics in order to provide a proper analysis of the scene. In this context, it is common to extract spatial features at a given time-space point by using a three-dimensional bounding box for feature extraction or dictionary learning [1] [2], assuming the bounding box extended for some subsequent frames will be enough to capture the relevant temporal information. Alternatively, one can also model the temporal evolution of independent features using tools such as Markov Chains [3].

Features extracted from a video sequence and stored in time can be interpreted as discrete time series. A discrete time series is a sequence of observations sampled over fixed and isolated time intervals, arranged in chronological order [4].

\* supported by FAPESP grant 15/13504-4

<sup>†</sup> supported by FAPESP grant 15/04883-0

Josef Kittler<sup>‡</sup>

# CVSSP — University of Surrey Guildford, GU2 7XH, UK

This paper focuses on the problem of analysing features extracted from surveillance video, monitoring in particular crowded scenes, by using time series decomposition in order to detect anomalous events, i.e. those events not explained by some model of normal behaviour. Here we consider motion anomalies by codifying motion via optical flow, and detect anomalies in the cases in which an anomalous motion pattern, i.e. with the test statistic above some threshold, appear. By using the time series framework in the video surveillance problem, a series of tools become available for the analysis of features in order to improve anomaly detection.

A time series is **stationary** when its properties are not affected by a change of time, i.e., the joint probability distribution remains the same by shifting any subset of observations over time; **linear** if the sequence of is defined by a linear combination of past observations and noise; **stochastic** when random observations and their relations are present; the probability density functions characterizing the randomness may change over time. Finally, it is **deterministic** if its samples strictly depend on past ones via some map or function.

In discrete control systems [5] a series is monitored in order to detect errors or unexpected behaviours. The use of data in those applications depends on models adjusted on components embedded into observations:

$$x(t) = T(t) + S(t) + \varepsilon(t), \tag{1}$$

in which x(t) represents a time series, T(t) is the trend, S(t) the seasonality and  $\varepsilon(t)$  a stochastic component.

By decomposing a time series into parts and components it is possible to study behaviours and use models that are better suited for each scenario [6, 7]. In this paper we propose the use of an EMD-based analysis in order to decompose optical flow (OF) features extracted from video in order to detect anomalies in a pedestrian dataset [8]. By using the method of Rios and Mello [9] we decompose the optical flow features and find a cut-off point in order to separate the deterministic and stochastic components. Our hypothesis is that the features will have a periodic behaviour for each fixed window since we expect objects to enter and leave a given region periodically — and for that reason by analysing the deterministic component we will have a clearer threshold to flag anomalies, in particular reducing false alarms.

<sup>&</sup>lt;sup>‡</sup>supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration (UDRC) in Signal Processing



**Fig. 1**. Optical flow feature for a fixed window, and its respective deterministic component with examples of events: background, anomalous (identified by green vertical lines), and normal events

## 1.1. Related Work and Contributions

There are many methods for decomposing time series. Spectral analysis methods are the most often used techniques to decompose time series [10], such as Principal Component Analysis, Singular Spectrum Analysis [11], Fourier Transform, Wavelets and Empirical Mode Decomposition (EMD) [12].

PCA and SSA are both linear methods. Fourier Transform approach assumes the series is periodic and stationary. Although Wavelet Transform allows a time/frequency analysis, it also assumes linearity. EMD supports decomposition regardless of linearity, stationarity and stochasticity, by extracting a set of monocomponents called Intrinsic Mode Functions (IMF) [13]. IMFs support the study of instantaneous frequencies and amplitudes using the Hilbert Spectral Analysis, which is useful in signal processing applications [14].

The causality within events is important in video data, as also pointed out by [15, 16]. In particular [15] represented video features as a graph in order to detect anomalies by editing the graph, while [16] focuses on finding subsequent voxels in order to model motion; however, by using one-class SVM classifiers they assume independence among observations, which is inappropriate when trying to analyse motion over time. Unlike other papers which used bounding boxes in order to capture features, we model the time series by selecting IMF components obtained by the EMD [17]. Although the use of 3d bounding boxes is able to flag abnormal events guided by some independent or spatial-dependent model, the temporal consistency is seldom discussed. In [7], the authors also used a temporal decomposition, but for reconstruction objectives, and not using the EMD method.

Our main contribution is a method suited for time series analysis that is not restricted by any assumption, which involves EMD signal decomposition into deterministic and stochastic components. We demonstrate that the representation of OF using the deterministic component is more stable and less ambiguous, producing competitive results in a benchmark dataset. The effect is visually shown in Figure 1 giving an OF feature and its correspondent deterministic component.

# 2. EMPIRICAL MODE DECOMPOSITION (EMD)

The EMD [12] performs a sifting process, in which local maxima and minima of x(t) are computed over time, and used to compose upper and lower envelopes u(t) and l(t), by using cubic splines, and then a mean envelope m(t) = [u(t) + l(t)]/2, which is used as a subtraction function to be applied to the original series x(t). The first candidate for the first monocomponent is  $h_{1,1} = x(t) - m(t)$ . This is used in the place of the original data and the sifting process is repeated until the candidate satisfies the IMF definition: (1) the number of extrema and the number of zero-crossings must differ at most by one; or (2) m(t) = 0 for every t.

After obtaining a candidate k that satisfies the IMF definition, the first IMF monocomponent is set as  $h_1(t) = h_{1,k}(t)$ . The first IMF is then removed from data,  $x(t) - h_1(t)$  and the resulting data is again analysed by the whole process, producing further IMFs, until the last IMF becomes a monotonic function and no further IMF can be extracted. This final component is a residual r(t). Therefore, a time series x(t) is composed of a set of IMF monocomponents plus a residual:

$$x(t) = \sum_{n=1}^{N} h_n(t) + r(t).$$
 (2)

This method makes it possible to analyse a wider class of time series when compared with the other methods cited above, the applicability of which is limited to linear and stationary data. Also, we expect to take advantage of the fact that deterministic influence increases as new IMFs are obtained, due to reinforced sinusoidal characteristics of further IMFs.

## 2.1. IMF analysis for signal decomposition

A method to quantify the mutual information (MI) of consecutive IMFs can be used to detect behaviour modification, allowing to find a cut-off point in which we could separate the deterministic from the stochastic component of the time series. According to [9], higher-frequency IMFs usually exhibit lower mutual information which allows to characterize them as stochastic, whereas MI increases as a the bandwidth of successive IMFs diminishes. This way, deterministic IMFs's phase spectra will share more similar information with other IMFs, once there is a strong entropy among them. Therefore, Fourier Transform can be used as a tool to analyse the spectrum of each IMF (excluding the residue), obtaining a set of complex coefficients in frequency space:

$$C_k(t) = \mathcal{F}(h_k(t)), \tag{3}$$

in which the coefficients  $c_{k,i} \in C_k$  are obtained by:

$$c_{k,i} = \sum_{t=1}^{T} h_k(t) \cdot e^{-i2\pi(i/T)t}, \forall i \in \{1, 2, \cdots, T\}.$$
 (4)

After obtaining the coefficients, the phase spectrum is computed for each IMF using an arctangent function on the ratio of the imaginary and real parts of the coefficients:

$$\theta_k(t) = \arctan\left(\frac{Im[C_k(t)]}{Re[C_k(t)]}\right).$$
(5)

In this paper we use a discrete estimator for the MI based on the method of Darbellay and Tichavsky [18]. It partitions the space of data pairs into finite non-overlapping cells in a quad-tree based approach, generating new partitions until it achieves a conditional independence among cells. It starts with a one-cell partition containing all data pairs, then: (1) the cell is partitioned into 2 halves; (2) the resulting subcells are recursively partitioned if the number of points in each new cell is greater than 4, and the mutual information in each subcell is less then a threshold; otherwise, the MI is estimated from the final subcells, each one contributes to the estimated MI by a proportional amount. For details see [18].

The phase spectra  $\theta_p(t)$  and  $\theta_{p+1}(t)$  — of consecutive IMFs  $h_p(t)$  and  $h_{p+1}(t)$  — are analysed by [18] method, producing a value  $\nu_p$  for each consecutive pair:

$$\nu_p = I(\theta_p(t), \theta_{p+1}(t)) \tag{6}$$

This value indicates how much information is preserved from one component to the next, and it is used to separate the deterministic component as described in the next section.

#### 3. METHOD

In this paper we propose to use the method proposed by Rios and Mello [9] in order to obtain the deterministic component of Farneback optical flow features [19] extracted over time from the UCSD Pedestrian dataset 2 [8]. A threshold is obtained by using the available training data for each  $30 \times 30$ window. An anomaly is flagged if the feature value on a given window is outside the defined threshold.

### 3.1. Preprocessing and Optical Flow (OF) features

Starting with the second frame of each video we compute for each block of the current frame the optical flow (with respect to the previous frame). Two features are used in this paper: the magnitude of the OF vector in the x and y directions. In order to speed-up the optical flow computation, we obtained the background by computing the mean training image. After subtracting the background (both in train and test stages), we defined the regions of interest for which we use the optical flow, as shown in Figure 2.



Fig. 2. Background image and region-of-interest mask

#### 3.2. Deterministic component extraction

The MI among two stochastic IMFs is close to zero because correlation between them is low, we look for a p in which mutual information change is abrupt, and use this as a cut-off point to separate the deterministic component [9]. We compute the mean value:

$$\bar{\mathcal{V}} = \frac{1}{N-1} \sum_{p=1}^{N-1} \nu_p,$$
(7)

and then look for the value  $\nu_z$  that is immediately lower than  $\overline{\mathcal{V}}$  and use it as the cut-off point, therefore obtaining:

$$d(t) = \sum_{n=z+1}^{N} h_n(t) + r(t),$$
(8)

in which d(t) is the deterministic component, and the remaining components are those  $h_n(t)$  with  $n = 1 \cdots z$ .

In order to illustrate the effect of extracting the deterministic component we show the sum of the average magnitude of OF vectors in both directions in Figure 3.

#### **3.3.** Anomaly detection

Using the training data, and for each  $30 \times 30$  block, we compute a training threshold which is the maximum value of the feature for that specific block. Then, in the test stage, when the feature goes over the value, it is considered an anomaly. We can also vary the threshold: allowing for a higher threshold will possibly avoid false alarms, but will also cause the system to fail to detect some anomaly; or reducing the threshold that will cause the opposite effect.



Fig. 3. Average magnitude of the optical flow vectors (x and y), represented by the red blocks, calculated using the original signal (left) and only its deterministic component (right)

## 4. EXPERIMENTS AND RESULTS

We used the UCSD dataset *Ped2*, with 16 training and 12 testing videos, each containing between  $360 \times 240$  frames. The Equal Error Rate (EER) is the standard measure to compare results in this dataset, and it is defined using a relationship between the True Positive Rate (TPR) and False Positive Rate (FPR), so that the EER is the number of misclassified frames for which the FPR = 1-TPR.

In order to illustrate the difference between using the original signal and the decomposed version we show in Figure 4 an example of the variation of the features computed in a fixed window. In this particular example, there are 6 anomalies: using the original OF features, we detect 3 false alarms, and 2 anomalies are missed; using the deterministic OF, it was possible to reduce those errors to 1 anomaly missed.



**Fig. 4**. Example of a time series formed by optical flow magnitude in a fixed window, and its deterministic component. Green vertical lines indicate ground-truth anomalies. The horizontal line is the threshold. A red ball indicates a false alarm, and a black ball an anomaly not detected.

In Figure 5 we examples of detected anomalies, in which red indicate values above the threshold, and blue values on the boundary. It can be seen that the deterministic component was more stable when compared with the original OF.

The EER results are shown in Table 1, comparing the use of original data (OF) with the deterministic component (OFdet). We choose to use a minimalist approach in order to highlight the potential of the decomposition method. For comparison we included the following methods: temporal MDT [8], the mixture of optical flow models (MPPCA) [20] and the local motion histogram (LMH) [21]. Although MDT-tmp achieved the best result, the signal decomposition achieved



Original Optical Flow



Deterministic Component

**Fig. 5**. Anomaly detection based on the original OF feature (left) and deterministic component (right). Blocks highlighted in red correspond to anomalies (outside threshold) and in blue borderline values (5% of the threshold).

a competitive EER using only two features and a thresholdbased anomaly detection, with a significant impact on the error when compared with the original signal.

MDT, MPPCA and LMH are more complex descriptors, involving models and learning stages. In fact, those methods could also benefit from the decomposition, since they could use the deterministic component of features in their model. Our aim was, instead, to show how the decomposition of the optical flow features benefit anomaly detection in video. As can be seen in Figures 3 and 5, the deterministic component provides relevant visual information for surveillance.

Table 1. Frame-based results on UCSD-Ped2 dataset					
Method	OF	OF-det	MDTtmp	MPPCA	LMH
EER (%)	40.2	31.7	27.9	35.8	45.8

# 5. CONCLUSION

In this paper we showed that dynamic behaviour of features extracted from video data can be considered as a time series and decomposed into an informative (deterministic) and an stochastic component to analyse the patterns over time more consistently. The use of the EMD algorithm is proposed for this purpose. The advocated method has been evaluated on the problem of detecting object motion anomalies in a benchmarking video with very promising results. as a reliable tool to decompose signals regardless of linearity, stationarity and stochasticity. We believe that this fact led to the success of the results. In particular, our results show that by separating the deterministic component from the stochastic component, it was possible to reduce the error rate, producing a system that is more stable to recognize motion anomalies.

One limitation of the method is its formulation as a batch computation process, which requires the entire time series for the decomposition. Although there are fast implementations available, in future studies it would be interesting to develop an incremental EMD. The use of decomposed features in conjunction with learning models is also a matter for future work.

## 6. REFERENCES

- Rima Chaker, Zaher Al Aghbari, and Imran N Junejo, "Social network model for crowd anomaly detection and localization," *Pattern Recognition*, vol. 61, pp. 266–281, 2017.
- [2] Mehrsan J Roshtkhari and Martin D Levine, "An online, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1436–1452, 2013.
- [3] Lin Zhao, Xinbo Gao, Dacheng Tao, and Xuelong Li, "Tracking human pose using max-margin markov models," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5274–5287, 2015.
- [4] Robert H Shumway and David S Stoffer, *Time series analysis and its applications*, Springer Science & Business Media, 2013.
- [5] George Box, Gwilym M Jenkins, and Gregory Reinsel, "Time series analysis: Forecasting & control," 1994.
- [6] Renato P Ishii, Ricardo A Rios, and Rodrigo F Mello, "Classification of time series generation processes using experimental tools: a survey and proposal of an automatic and systematic approach," *International Journal* of Computational Science and Engineering, vol. 6, no. 4, pp. 217–237, 2011.
- [7] Nathan Jacobs and Robert Pless, "Time scales in video surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1106– 1113, 2008.
- [8] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.
- [9] Ricardo A Rios and Rodrigo F de Mello, "Applying empirical mode decomposition and mutual information to separate stochastic and deterministic influences embedded in signals," *Signal Processing*, vol. 118, pp. 159 – 176, 2016.
- [10] Ricardo A Rios and Rodrigo F de Mello, "Improving time series modeling by decomposing and analyzing stochastic and deterministic influences," *Signal Processing*, vol. 93, no. 11, pp. 3001–3013, 2013.
- [11] Andrew L Rukhin, "Analysis of time series structure ssa and related techniques," *Technometrics*, vol. 44, no. 3, pp. 290–290, 2002.

- [12] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.* The Royal Society, 1998, vol. 454, pp. 903–995.
- [13] M E Torres, Marcelo A Colominas, Gastón Schlotthauer, and Patrick Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *IEEE International Conference on Acoustics, Speech* and Signal Processing (ICASSP). IEEE, 2011, pp. 4144–4147.
- [14] Udit Satija, Barathram Ramkumar, and M Sabarimalai Manikandan, "A unified sparse signal decomposition and reconstruction framework for elimination of muscle artifacts from ecg signal," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*). IEEE, 2016, pp. 779–783.
- [15] Junseok Kwon and Kyoung Mu Lee, "A unified framework for event summarization and rare event detection.," in CVPR, 2012, pp. 1266–1273.
- [16] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang, "An efficient subsequence search for video anomaly detection and localization," *Multimedia Tools and Applications*, pp. 1–22, 2015.
- [17] Douglas B de Souza, Jocelyn Chanussot, and Anne-Catherine Favre, "On selecting relevant intrinsic mode functions in empirical mode decomposition: An energybased approach," in *IEEE International Conference* on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014, pp. 325–329.
- [18] Georges A Darbellay and Petr Tichavsky, "Independent component analysis through direct estimation of the mutual information," in *ICA*, 2000, vol. 2000, pp. 69–75.
- [19] Gunnar Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.
- [20] Jaechul Kim and Kristen Grauman, "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in *Computer Vision and Pattern Recognition*, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 2921–2928.
- [21] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and Daviv Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions* on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 555–560, 2008.