

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

## Correlation tracking based on wavelet domain information

Huseyin Levent Ipek, I. Yilmaz, Yasemin C. Yardimci,  
Ahmet Enis Cetin

**SPIE.**

# Correlation Tracking Based on Wavelet Domain Information

H. Levent Ipek<sup>a</sup>, I. Yilmaz<sup>b</sup>, Yasemin Yardimci<sup>\*a</sup> and A. Enis Cetin<sup>\*b</sup>

<sup>a</sup>Informatics Institute, Middle East Technical University, Ankara, 06531, Turkey ;

<sup>b</sup>Dept. of Electrical and Electronics Engineering, Bilkent University, Ankara, 06533, Turkey.

## ABSTRACT

Tracking moving objects in video can be carried out by correlating a template containing object pixels with pixels of the current frame. This approach may produce erroneous results under noise. We determine a set of significant pixels on the object by analyzing the wavelet transform of the template and correlate only these pixels with the current frame to determine the next position of the object. These significant pixels are easily trackable features of the image and increase the performance of the tracker.

**Keywords** : Motion estimation, wavelets.

## 1. INTRODUCTION

Tracking moving objects in video is required in a wide range of applications including forward looking infrared (FLIR) imaging systems as well as other video surveillance systems [1], [2]. In correlation tracking usually a reference image of the object is first estimated from previous frames of video. Then, the reference image template is correlated with offsets within a search window in the current image. The search window is a larger region than the template. The indices maximizing the correlation function determines the shift in the position of the moving target. A similar approach is based on minimization of a cost function  $g(k_1, k_2)$  is first obtained for all  $(k_1, k_2)$  in the search window as follows

---

\* { yardimiy , cetin }@ ee.bilkent.edu.tr ; phone +90 312 2901477 , fax : +90 312 2664192

$$g(k_1, k_2) = \frac{1}{L} \sum_{(n_1, n_2) \in R_t} d(x_r(n_1, n_2) - x_i(n_1 + k_1, n_2 + k_2)) \quad (1)$$

where  $x_r$  is the reference template image,  $x_i$  is the current image,  $R_t$  is the set of pixels forming the object template,  $L$  is the number of pixels in  $R_t$  and  $d(\cdot)$  is the distance measure. The cost function  $d(\cdot)$  can be  $d(\cdot) = (\cdot)^2$  for squared error distance or  $d = |\cdot|$  for an absolute difference based tracking. By minimizing the  $g(k_1, k_2)$  the motion vector of the object is estimated with the usual assumptions that the object only goes through a translation and there is no significant change in lighting conditions between two consecutive frames.

This approach may produce erroneous results under noise. For example, if there are pixels with the same value forming a flat region on the object then the contribution of some of these pixels to the right hand side (RHS) of (1) may be zero in cases where object positions overlap in two consecutive frames in spite of the motion. If such pixels are corrupted by noise, they appear to be moving in all directions in a random manner and this may lead to incorrect results. Therefore it may be important to compute the RHS of (1) in a subset  $R_s \in R_t$  formed by reliable pixels. We call the set  $R_s$  as the set of significant pixels which are determined in wavelet domain. We classify object pixels corresponding to wavelet coefficients exceeding a threshold as significant pixels and include them in  $R_s$ . Other object pixels whose wavelet coefficients are below the threshold are classified as insignificant and they are not used in computing the RHS of (1).

In general large valued wavelet coefficients corresponds high frequency components of the image[3,4,5]. Therefore, the set  $R_s$  includes the edges and the texture information of the object. Edges and the texture of the object produce relatively large wavelet coefficients compared to smooth regions as well. Due to this reason, the proposed tracker is also more robust than edge and corner trackers which only track the changes in the position of the edges and corners of the object.

## 2. ESTIMATION OF SIGNIFICANT PIXELS FOR TRACKING

In our correlation based motion detection algorithm, we first compute the WT (Wavelet Transformation) of the

reference template image and define the wavelet coefficients exceeding the threshold as significant wavelet coefficients. Let  $x_w(l_1, l_2)$  be a wavelet coefficient in a subband of  $x_r$ . In order to determine the set  $R_S$  containing significant pixels each wavelet coefficient is processed as follows

$$x_{wr}(l_1, l_2) = \begin{cases} x_w(l_1, l_2), & \text{if } |x_w(l_1, l_2)| > T_h \\ 0 & , \text{ if } |x_w(l_1, l_2)| < T_h \end{cases} \quad (2)$$

where  $T_h$  is a threshold. In general, the threshold can be selected as  $T_h = c\sigma_v$  where  $c$  is an appropriate constant greater than one, and  $\sigma_v$  is the variance of wavelet coefficients of the particular subband. After we determine each wavelet coefficients  $x_{wr}(l_1, l_2)$  exceeding its threshold we estimate the location of actual image pixels producing them. The significant pixel must ensure that all wavelet coefficients for that pixel ( Low/High, High/Low, High/High) are greater than their own threshold. These image pixels form the set  $R_S$ .

Let us assume that a single stage wavelet analysis (or subband decomposition) is carried out. As a result four quarter size subimages are obtained:  $x_{ll}$ ,  $x_{lh}$ ,  $x_{hl}$  and  $x_{hh}$ . The last three subimages are the wavelet subimages  $wx_r = \{ x_{lh}, x_{hl}, x_{hh} \}$ . If the Haar wavelet decomposition is used then each wavelet coefficient is produced by a two by two region in the original image. For example,  $|x_{lh}(l_1, l_2)| > T_h$  corresponds to a two by two region in the original image:  $x_r(n_1, n_2)$ ,  $n_1 = 2l_1, 2l_1-1$ ,  $n_2 = 2l_2, 2l_2-1$ . In other wavelets the number of pixels forming a wavelet coefficient is larger than four but most of the contribution comes from the immediate neighborhood of  $(n_1, n_2) = (2l_1, 2l_2)$ . Therefore, in other wavelets one can classify  $x_r(n_1, n_2)$ ,  $n_1 = 2l_1 - 1, 2l_1, 2l_1 + 1$ ,  $n_2 = 2l_2 - 1, 2l_2, 2l_2 + 1$  as significant pixels and include them in the set  $R_S$ .

In summary, we compute the WT of the reference image and determine three wavelet coefficients exceeding the threshold. Then, we form the set  $R_S$  by determining the pixels corresponding to these significant wavelet coefficients. We compute Equation (1) over the set  $R_S$  and find the minimizer of (1) where  $d(\cdot)$  is the absolute value function as the motion vector,  $(l_1, l_2)$ . This is achieved by the well known Kanade-Lucas Tracker (KLT) as implemented in [6]. We compared our wavelet domain detector with the well known Harris corner detector. The wavelet domain detector is

implemented as a single stage wavelet decomposition using 7-tap Lagrange wavelets.

### 3. EXPERIMENTAL STUDIES

#### 3.1. Images taken with the synthetic video camera

We generated a synthetic object as a 30x30 square region with mean intensity 0.8. Its texture is created by adding i.i.d. on Gaussian noise samples with mean zero and variance 0.8. The background pixels have mean 0.2 and size 100x100. The object is located at the center of the background pixels and on the next frame it is moved by 1 pixel to the right horizontally and one pixel to the bottom vertically. In both frames the synthetic image is corrupted by a Gaussian noise with mean zero and variance 0.01. The pixel values which have an intensity below zero and above one are saturated to zero and one respectively.

Figure 1 depicts the result of Wavelet Domain Feature Detector. It is observed that all of the 45 features are on the object itself and most of them are tracked correctly. The results of the Harris detector is presented in Figure 2. we also detected 45 features on this image. This corresponds to a threshold value of 67. As seen in Figure 2a most of the features are detected on the background rather than the target and they are almost always tracked erroneously.

We decreased the number of detected features for both detectors by eliminating the features that are in a 15x15 neighborhood of the strongest features. Similar tracking results are obtained for both cases.

#### 3.2. Images taken with the real video camera

We worked on real video sequences as well. The results of the highway video sequences are presented. We implemented our study on five consecutive frames of highway video sequence using both the wavelet domain feature detector and Harris corner detector. The standard deviation of the wavelet coefficients in a subband is estimated from the data and the threshold value for each subband is selected as  $2.5\sigma$  to determine statistically significant pixels for the wavelet domain feature detector. The minimum eigenvalue for the Harris detector is selected as 0.25 times the maximum of the minimum eigenvalues of the features detected in the whole image. The feature window size for the both detectors is 15x15.

Figure 3 shows that both feature detectors detect many significant features on the main moving object (Van). All of these significant pixels are trackable.

We corrupted the highway video frames with the additive white Gaussian noise with mean zero and variance

0.01. The wavelet domain detector can still detect the van with many pixels. Many features are still located on the van in the presence of noise. The few pixels that are not located on the van are tracked correctly with zero motion vector (Figure 4a). The Harris detector, on the other hand, also selects many features on the van in the noiseless case but most features are selected on the background in the presence of noise as seen in Figure 4b. Although it tracks the van correctly, most of the features are picked on the background in noise.

#### 4. CONCLUSION

In this paper a new method for determining trackable features is described. The method is based on using high/low, low/high and high/high wavelet coefficients to determine features to be tracked. Its performance is compared with the well known Harris detector. The wavelet domain feature detector is more likely to select feature points on the target like objects and is robust under noise. Its computational complexity is Order(N) which is less than that of the Harris detector.

#### REFERENCES

- [1] S. Blackman, R. Papoli, *Design and analysis of modern tracking systems*, Artech House, 1999.
- [2] A. M. Bagci, Y. Yardimci, A. Enis Cetin, "Moving object detection using adaptive subband decomposition and fractional lower order statistics in video sequences, *Signal Processing*, Oct. 2002.
- [3] S. G. Mallat and S. Zhong. Characterization of signals from multiscale edges, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710-732, July 1992.
- [4] Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks*, Wellesley - Cambridge Press, Wellesley MA , 1996.
- [5] A. E. Cetin, R. Ansari, 'Signal recovery from wavelet transform maxima,' *IEEE Trans. Signal Processing*, Vol 42, No.1, pp. 194-196, Jan. 1994.
- [6] A. Fusiello, E. Trucco, T. Tommasini, V. Roberto 'Improving Feature Tracking with Robust Statistics,' *Pattern Analysis & Applications*, Vol 2, pp. 312-320, 1999.
- [7] C.G. Harris, M.J. Stephens, "A combined corner and edge detector", proceeding Fourth Alvey Vision Conference, Manchester. pp.147-151, 1988.

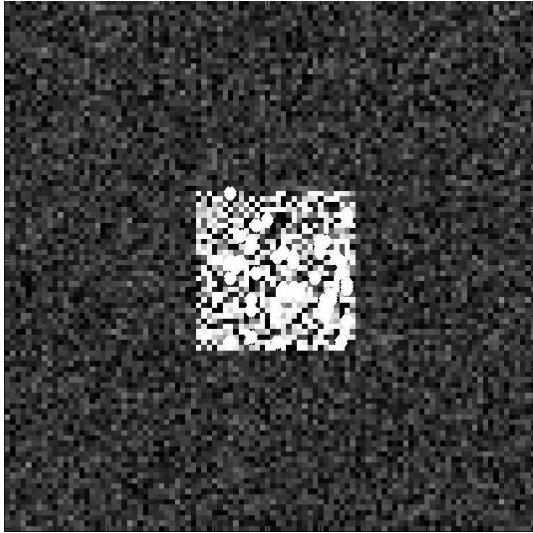


Figure 1 (a)

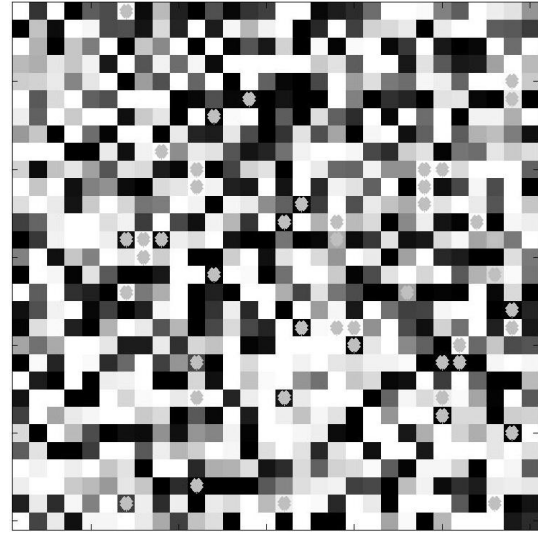


Figure 1 (b)

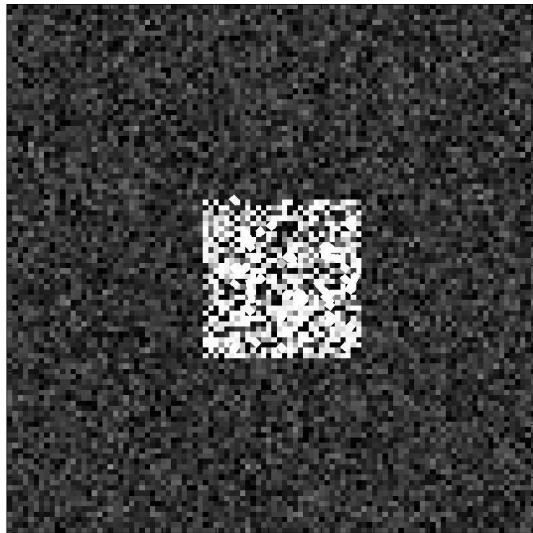


Figure 1 (c)

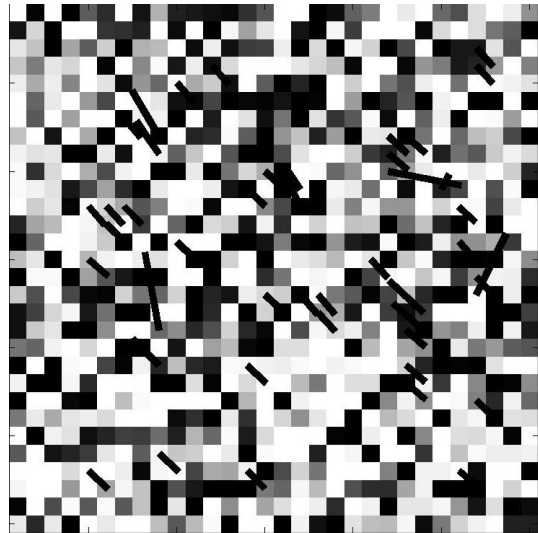


Figure 1 (d)

Figure 1 The results of wavelet domain feature detector

a) Detected features on the image

b) Detected features on the object

c) Needle diagram of tracked features on the image

d) Needle diagram of the features on the object

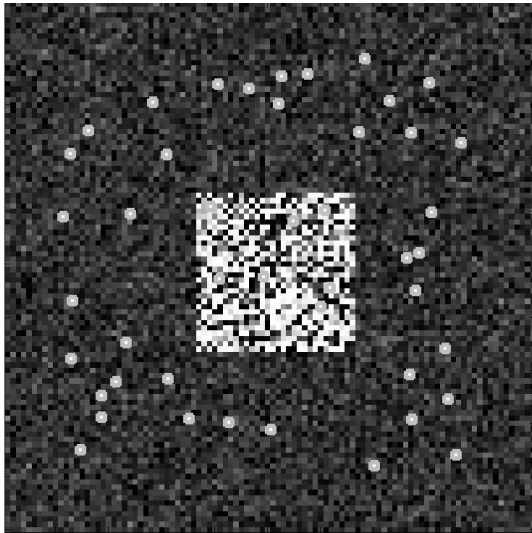


Figure 2 (a)



Figure 2 (b)

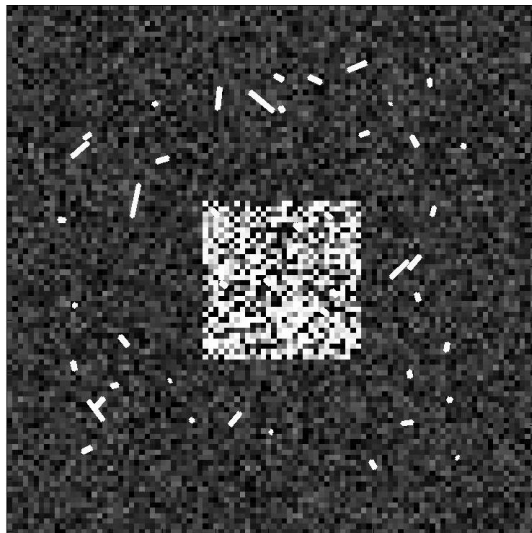


Figure 2 (c)

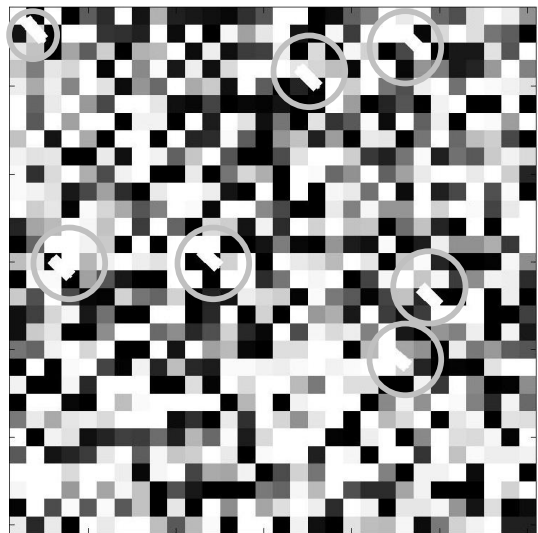


Figure 2 (d)

Figure 2 The results of Harris detector

a) Detected features on the image

b) Detected features on the object

c) Needle diagram of tracked features on the image

d) Needle diagram of the features on the object





Figure 3 (a)



Figure 3 (b)

Figure 3 The feature detection on a real image (Noiseless Case)

- a) The features obtained by wavelet domain detector
- b) The features obtained by Harris Detector

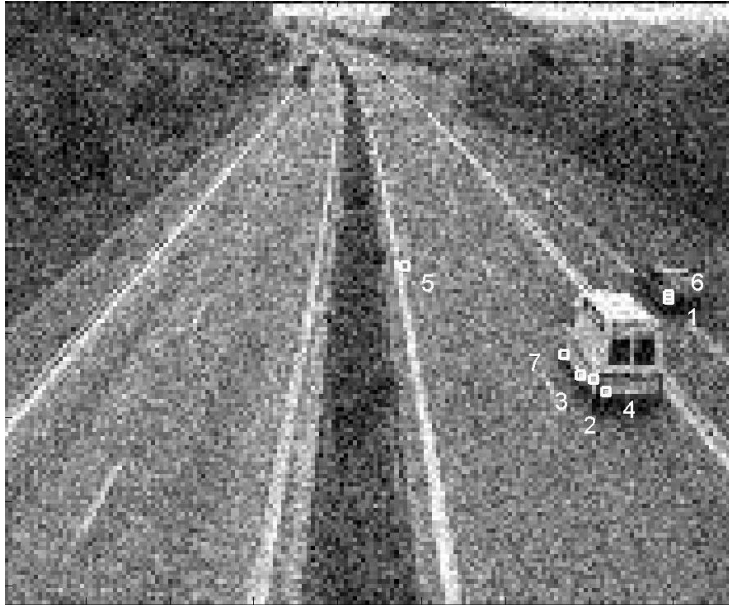


Figure 4 (a)

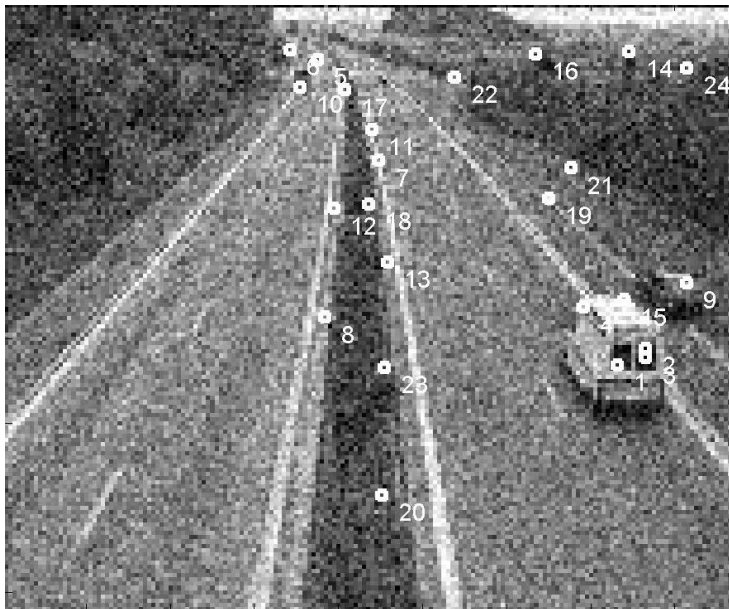


Figure 4 (b)

Figure 4 The feature detection on a real image (Noisy Case ; zero mean and 0.01 variance)

- a) The features obtained by wavelet domain detector
- b) The features obtained by Harris Detector