MATHEMATICAL MODELING AND COMPUTING, Vol. 10, No. 1, pp. 159–170 (2023)

An adaptive wavelet shrinkage based accumulative frame differencing model for motion segmentation

Lahgazi M. J.[1], Hakim A.[1], Argoul P.[2]

[1] *Faculty of Sciences and Technics, Cadi Ayyad University, Marrakesh, Morocco*
[2] *MAST-EMGCU, Université Gustave Eiffel, IFSTTAR, F-77477 Marne-la-Vallée, France*

(Received 30 May 2022; Accepted 9 November 2022)

Motion segmentation in real-world scenes is a fundamental component in computer vision. There exists a variety of motion recognition algorithms, each with varying degrees of accuracy and computational complexity. The most widely used techniques, in the case of static cameras, are those based on frame difference. Those methods have a significant weakness when it comes to detect slow moving objects. Therefore, we introduce in this paper a novel approach that aims to improve motion segmentation by proposing an accumulative wavelet based frame differencing technique. Moreover, in the proposed approach we exploit a combination of several techniques to efficiently enhance the quality of motion segmentation results. The approach's performance on real-world video sequences shows that comparing frames using the 2D wavelet transform increases motion segmentation quality.

Keywords: *motion segmentation; wavelet transform; wavelet shrinkage; adaptive threshold; frame differencing; background subtraction.*

2010 MSC: 68U10, 65T60, 94A08 **DOI:** 10.23939/mmc2023.01.159

1. Introduction

Video sequence analysis has become a very active research field in computer vision due to the growing of its importance in many applications such as surveillance systems, pedestrian detection and obstacle detection for intelligent cars. One of the most important operations is to separate foreground objects from background, this can be performed in different ways depending on the nature of the available data and the captured scenarios, which may present multiple moving or static objects captured from a moving or static or multiple cameras. When the data are a video sequence split into several consecutive frames, the detection of moving objects doesn't require any additional knowledge about the data, and the separation of the background that is assumed to be static, from foreground that contains the objects to be detected, is accomplished by frames differentiation. Motion segmentation is often the first step of a multi-stage computer vision system [1–4] such as car tracking, person recognition or wildlife monitoring. Background subtraction (BS) is a quick way of localizing moving objects in a video captured by a static camera. Besides, it can improve the performance of computer vision systems by deleting the uninteresting sections of the scene, which lowers the probability of having false negatives.

The nature of the approach to use for background subtraction is dictated by the nature of the data. If the video sequence is taken from a moving camera, the most common used techniques are the ones based on optical flow [5, 6]. Optical flow is the pattern of apparent motion of gray level intensities in an image that models objects, surfaces, and edges in a visual scene caused by the relative motion between a viewer and an object. The motivation behind using optical flow for this case is the fact that the background pixels have negligible motion as compared to foreground pixels [7]. In the case where the video sequences are taken from a static camera, the most common approach is to compare frames pixel-wise. In this contest, some methods use the current frame as reference, while others use an image of the background taken in the absence of moving objects, or a statistical background model [8]. The background can be modeled either by a uni-modal Gaussian [9], a Gaussian mixture [10], a non-parametric kernel density function [11], or simply temporal median filtering [12].

In this work, the main aim is to segment moving objects in a video sequence taken from a static camera. Thus, we propose a movement segmentation technique that exploits the wavelet analysis to separate foreground pixels from background pixels. The classical frame comparison methods compare the frames in the pixel domain. In the field of image recovery, recent studies [13,14] have demonstrated the effectiveness of using transform domain, such as wavelet transform, instead of pixel domain in image recovery. This idea inspired us to propose a wavelet shrinkage based frame differencing technique for motion segmentation. Besides, to improve the quality of the proposed model, we integrate other techniques including an adaptive threshold, a specific space of indexes, and the use of a suppression coefficient in the wavelet transform. Our approach makes the movement segmentation more robust against slow moving objects. Which is reported via different experiments conducted on real-world data.

The rest of the paper is organized as follows. In Section 2 we present an overview about several state of the art motion segmentation techniques including those based on the well-known wavelet approach. Section 3 is designed to present a mathematical background about the wavelet transform. Section 4 exposes the proposed motion segmentation technique. Section 5 addresses the experimental and comparatives results. Finally, we conclude the paper in Section 6.

2. Related works

2.1. Motion segmentation techniques:

Motion segmentation techniques are various, and as we mentioned before it is basically depending on the nature of the data [8,15–19]. Since we work on video sequences taken from a static camera, a basic approach is to compare frames with a reference image [8]. However the basic version of the method is limited when the scene presents lighting variations. To overcome the sensibility to lighting change, Kim et al. [15] proposed to use a codebook based algorithm to model background. The background model was created using a clustering technique where the intensities at each background pixel in every frame were clustered into the set of codewords. The proposed method was robust in handling moving background and lighting variation. Stauffer and Grimson [16] proposed the Gaussian mixture model (GMM), where a mixture of K Gaussian functions is used to model every pixel. In their work, the current frame is compared to the background model using every Gaussian in the mixture until a matched Gaussian is found. If the match is found the mean and variance of the matched Gaussian is updated. Elgammal et al. [17] proposed to construct a non-parametric background model using kernel density estimation. To handle small amounts of motion in the background scene, the method matches each background pixel with the nearby pixel location of the background model. Azmat et al. [19] proposed a parallel multi-modal background modeling algorithm which aims to optimize memory access pattern when working on low power GPUs. This algorithm has less computational and memory cost but its accuracy is at the same scale as the Gaussian mixture model.

2.2. Wavelet based motion segmentation techniques

Since its first definition in the beginning of the 1980s by French researchers, especially Grossmann and Morlet [20], the growth of wavelet research in mathematics has been first explosive with significant contributions from numerous authors. A large number of books exits on wavelets theory [21–23] and wavelet analysis has rapidly become a widely used tool in several fields including first signal and image processing, pattern recognition, machine learning, and computer vision. One of its basics properties is to provide a basis that can express the features of a signal or an image in an effective and easy way. For example in [24], the authors proposed a modal identification technique based on wavelet analysis for processing free responses of non-proportionally damped systems, integrated in noise, to directly obtain complex modes. This explains the emergence of the use of wavelet analysis-based techniques in various research fields.

There exist many approaches that use wavelet analysis for motion segmentation [25–29]. Davies et al. [25] introduced a one dimensional wavelet filter based on Haar wavelet filters [26] to detect objects that are very small, very slow or located in a low contrast area of the image. The major limitation of the proposed method was the high computation cost. Khare et al. [27] introduced a method for segmentation of moving object which is based on double change detection technique applied on Daubechies complex wavelet coefficients of three consecutive frames in order to solve foreground aperture and ghosting problems. Al-Berry and Mohammed [28] proposed a technique based on the idea of accumulative frame difference and used 2D Discrete Wavelet Transform (DWT) to enhance the accuracy of the segmentation. Jalal et al. [29] proposed a background subtraction approach exploiting noise resilience capability of wavelet domain combined with local spatial coherence and median filter in the training stage, and introduced a shadow suppression scheme based on directional coefficients of Daubechies complex wavelet transform [26].

3. Mathematical background

A mathematical background about the 2D wavelet transform is presented in this section. In our algorithm, we aim to apply the 2D discrete wavelet transform to an image, which can be done by applying 1D discrete wavelet transform to each row and then to each column. Following this idea, we introduce the 1D wavelet transform, and we present the scheme that leads to the 2D generalization.

3.1. 1D continuous wavelet transform

The 1D continuous wavelet transform replaces the 1D continuous Fourier transform's sinusoidal waves by a family generated by translations and dilation's of a window called a mother wavelet. It takes two arguments: time and scale.

A mother wavelet ψ is usually a function that have a zero-mean oscillation behavior, it oscillates like a wave but is quickly attenuated. Mathematically, ψ is a function of $L^2(\mathbb{R})$ that verifies the admissibility condition:

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(x)|^2}{|x|}\, dx < +\infty, \tag{1}$$

where $\hat{\psi}$ is the Fourier transform of ψ.

Moreover a mother wavelet ψ could produce families of wavelets:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-\tau}{s}\right), \quad s \in \mathbb{R}_+^*, \quad \tau \in \mathbb{R}, \tag{2}$$

where ψ is the mother wavelet, s is the scale parameter, τ is the shifting parameter or the position of the scale window, and $\psi_{s,\tau}$ is the scaled and shifted wavelet.

The factor $\frac{1}{\sqrt{s}}$ guarantees that the wavelet maintains constant energy, since $\frac{1}{s}\int_{-\infty}^{+\infty} \left|\psi(\frac{t}{s})\right|^2 dt = \int_{-\infty}^{+\infty} |\psi(t)|^2 dt$. Thus, the L^2 norm is appropriate when the modulus-squared wavelet transform is wished to reflect the energy of the analyzed signal.

Similar to Fourier transform (FT) which uses sine and cosine as the basis function, wavelet transform (WT) utilizes a family of wavelet functions. Considered the convolution between the signal and a wavelet function, WT decomposes a signal into localized contributions (approximations and details) labeled by scale and shifting parameters. The 1D continuous wavelet transform of a function $f \in L^2(\mathbb{R})$ is defined as follows:

$$g(s,\tau) = \int_{-\infty}^{+\infty} f(t)\, \psi_{s,\tau}^*(t)\, dt, \quad s \in \mathbb{R}_+^*, \quad \tau \in \mathbb{R}, \tag{3}$$

where "*" designates the complex conjugate.

The reconstruction of the original function f is performed with the inverse continuous wavelet transform:

$$f(t) = \int_0^{+\infty} \int_{-\infty}^{+\infty} g(s,\tau)\, \psi_{s,\tau}(t)\, ds\, d\tau, \tag{4}$$

where

$$C = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|}\, d\omega.$$

3.2. 1D discrete wavelet transform

It is advantageous to use a dyadic variant of the 1D discrete wavelet transform that uses special values for the shift b and scale a while defining the wavelet basis by introducing the scale step j and the shift step k: $s = 2^{-j}$ and $\tau = k.2^{-j}$. The dyadic 1D discrete wavelet transform of an analog signal $x(t)$ with finite energy is defined as follows:

$$c(j,k) = \sum_t x(t)\, \psi^*_{j,k}(t), \tag{5}$$

where $\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k)$.

The inverse 1D discrete wavelet transform is defined as follows:

$$x(t) = \sum_k \sum_j c(j,k) \psi_{j,k}(t). \tag{6}$$

3.3. Choice of appropriate mother wavelet

The architecture of the mother wavelet ψ must be designed to generate as many wavelet coefficients $\langle x, \psi_{j,k} \rangle$ as possible that are close to zero. This depends mostly on the regularity of x, the number of vanishing moments of ψ, and the size of its support.

Mathematically, the wavelet ψ has p vanishing moments if

$$\int t^k \psi(t)\, dt = 0 \quad \text{for} \quad 0 \leqslant k < p, \tag{7}$$

which means that ψ is orthogonal to any polynomial of degree $p-1$. In other words, a wavelet has p vanishing moments, if and only if the wavelet scaling function can generate polynomials up to degree $p-1$. If x is regular and ψ has enough vanishing moments then the wavelet coefficients $|\langle x, \psi_{j,k} \rangle|$ are small at fine scales 2^j.

On the other hand, if x has an isolated singularity at a point inside the support of $\psi_{j,k}$ then $\langle x, \psi_{j,k} \rangle$ may have a large amplitude. To minimize the number of high amplitude coefficients, we must reduce the size of the support of ψ, which corresponds to the notion of sparsity of ψ.

The constraints imposed on orthogonal wavelets imply that if ψ has p vanishing moments then its support is at least of size $2p-1$. When choosing a particular wavelet, we thus face a trade-off between the number of vanishing moments and the support size. If x has few isolated singularities and is very regular between singularities, we must choose a wavelet with many vanishing moments to produce a large number of small wavelet coefficients $\langle x, \psi_{j,k} \rangle$. If the density of singularities increases, it might be better to decrease the size of its support at the cost of reducing the number of vanishing moments [30].

3.4. 1D fast wavelet transform

The concept of multi-resolution analysis introduced by Mallat [31] is the basis of the fast discrete wavelet transform (DWT). The algorithm, proposed by Mallat, makes the signal pass through a low-pass filter to implement large-scale analysis and pass through a high-pass filter to implement small-scale analysis.

The DWT can be interpreted as signal decomposition in a set of independent spatially oriented frequency channels. The signal x is passed through two complementary filters and produces two signals: approximation and details. This is called decomposition or analysis. The components can be assembled back into the original signal without loss of information. This process is called reconstruction or synthesis. The mathematical manipulation, which implies analysis and synthesis, is called a discrete wavelet transform and inverse discrete wavelet transform [32].

In practice, the choice of appropriate filter banks, in term of reconstruction, is usually used as a simple way to perform DWT. This is executed using low-pass and high-pass filters associated, respectively, to the scale function, and the wavelet function [33].

3.5. 2D fast wavelet transform

The simplest form of the 2D discrete wavelet transform is the separable 2D DWT [31], which consists in applying the 1D DWT to each row and then to each column. The image is decomposed into four sub-images via the high-pass and low-pass filtering. The image is decomposed along column direction into sub-images to high-pass frequency band H and low-pass frequency band L. Assuming that the input image is a matrix of $m \times n$ pixels, the resulting sub-images become $m/2 \times n$ matrices. At the second step the images H and L are decomposed along row vector direction and respectively produce the high and low frequency band HH and HL for H, and LH and LL for L. Four output images become the matrices of $m/2 \times n/2$ pixels. Low frequency sub-image LL (A_1) possesses high energy, and is a smallest copy of original image (A_0). The remaining sub-images LH, HL, and HH respectively extract the changing components in horizontal (D_{11}), vertical (D_{12}), and diagonal (D_{13}) direction [34]. The scheme in Figure 1 illustrates this procedure.

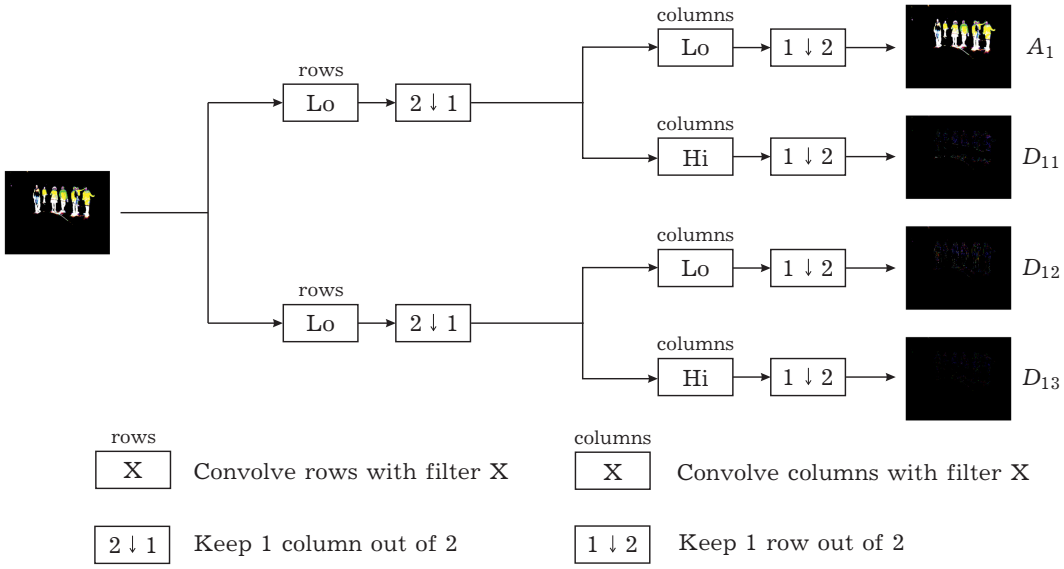


Fig. 1. 2D discrete wavelet transform [34].

The use of wavelet transform defined by both Eqs. (2) and (3) has several advantages. It allows complex information such as images to be decomposed into elementary forms at different positions and scales and subsequently reconstructed with high precision [33]. Moreover, the use of filter banks in the fast wavelet transform, makes it computationally efficient [33]. Another advantage is their ability to capture the energy of a signal in few energy transform values, it does not change the number of pixels required to represent the image and separate the information in a way that similar to the human visual system [35]. They have the capability of revealing aspects of data such as trends, breakdown points, discontinuities in higher derivatives and self-similarity which other signal analysis techniques miss [35].

In this work, we will use wavelet shrinkage in accumulative frame differencing to capture more accurately the change that occurs in pixel intensity from a frame to another.

4. The proposed approach

Our approach consists on change detection technique to detect moving objects in a specific frame. As the intensity of pixel changes, a motion extraction technique must be able to distinguish whether the difference is due to actual movement or something else, such as illumination changes. For this purpose, we will use accumulative frame differentiation. The accumulative frame difference, or so called *memory based technique* has the advantage of holding information about past frames. This enables the motion detector to ignore changes that occur only on few frames and attribute it to noise [28]. The frame we wish to subtract the background from, is used as the reference, and since the density of frames in video sequence is about 25 frames per second, comparison between consecutive frames will not detect a signification information about variation, especially for slow motion. On the other hand, it will be expensive in term of computational cost. To improve the performance, we define a neighborhood of the reference frame and only frames outside it are taken in consideration. We also avoid consecutive frames by defining a step between the frames compared to the reference frame. This helps avoiding redundancy and lightens the computational cost.

Let k be the index of the reference frame, A_k the reference frame, n the step between compared frames, v the radius of the neighborhood of the reference frame. We define the set of picked frames by

$$\mathbb{V} = \{t \in \mathbb{N} \setminus \{[k-v, k+v] \cap \mathbb{N}\}\,,\ t \equiv k[n]\}\,. \tag{8}$$

The choice of the space of indexes is motivated by two reasons. First, the less is the number of taken frames the faster the computation will be. Second, using adjacent frames reduces the efficiency of comparisons, since the comparison of the reference frame with two adjacent frames keeps almost the same information.

The frame differences are computed as follows, for $i \in \mathbb{V}$:

$$WT^k(i) = \left[A_1^k(i), D_{11}^k(i), D_{12}^k(i), D_{13}^k(i)\right] = \mathrm{dwt}(|A_i - A_k| > s), \tag{9}$$

where $A_1^k(i)$ is the matrix of approximation coefficients, and $D_{11}^k(i)$, $D_{12}^k(i)$, $D_{13}^k(i)$ are respectively the detail coefficients matrices in the horizontal, vertical and diagonal directions, and s is a threshold chosen manually.

The signal we are working on is the difference between frames, and important singularities appear mostly on the edges of moving objects, meaning that the majority of singularities are isolated and the spaces between them are regular. Following this logic, we consider *Daubechies* wavelets. This choice can be explained by the fact that Daubechies wavelets are characterized by a minimum size support for a given number of vanishing moments, which corresponds to a maximal number of vanishing moment for a given support.

Once computed, we apply a threshold on the wavelet transform to extract the changes that occur between two compared frames. Generally, the details coefficients are set to zero, and only the approximation coefficients are threshold. Mingbo et al. [36] proved that those components have an important role in the reconstruction of a pure signal and should be kept for the reconstruction. Following this idea, we will enhance the details coefficients matrices by multiplying each of them by its corresponding suppression coefficient [36], and then we will apply a threshold to each of them.

Further, setting a threshold for all the comparisons won't be a good idea since the amount of change between frames is not the same. A better solution will be to use a threshold that will be adaptive to every comparison. For the approximation coefficients $A_1^k(i)$ we will use the threshold proposed by Donoho et al. [37]:

$$T_i^k = \sigma_i^k \sqrt{\log S}, \tag{10}$$

where S is the number of pixels in the image and σ^2 is the variance estimated from the finest level of detail coefficients and defined by

$$\sigma_i^{k\,2} = \left[\,\mathrm{median}\left(|A_1^k(i)|/0.6745\right)\right]^2. \tag{11}$$

For the details coefficients matrices, we start by computing suppression coefficients and multiply every matrix by the corresponding coefficient. The suppression coefficients [36] are computed as follows:

$$sc_n^k(i) = \sqrt{\frac{A_1^k(i)}{A_1^k(i) + D_{1n}^k(i)}}, \quad n = 1, 2, 3. \tag{12}$$

The last step to perform before the inverse transform according to [36] is to discard detail coefficients which correspond to the noise and are less than a proper threshold. For this purpose we used Otsu's method [38] which determines the threshold by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance.

We apply the 2D inverse discrete wavelet transform (idwt) to the threshold transformation to obtain the comparison mask:

$$M^k(i) = \text{idwt}\left(WT_{\text{thresholded}}^k(i)\right). \tag{13}$$

And all the obtained masks are then summed up to obtain a pixel-wise voting map:

$$D_k = \sum_{i \in \mathbb{V}} M^k(i). \tag{14}$$

The final image is obtained by applying an extra threshold tr to eliminate the pixels with a low number of votes:

$$I(i,j) = \begin{cases} A_k(i,j) & \text{if} \quad D_k(i,j) > tr, \\ 0 & \text{if} \quad D_k(i,j) < tr. \end{cases} \tag{15}$$

The final threshold tr is determined empirically, and the image I is then smoothened with a 2D Gaussian smoothing kernel.

Overall, the contribution of our approach is basically based on the combination of multiple and efficient methods in different stages of the algorithm. Thus, it consists of the following points:

1. The use of a specific space of indexes, which reduces the complexity by avoiding redundant comparisons, see Eq. (8).
2. The use of wavelet shrinkage for each of the intermediate comparisons.
3. The use of an adaptive threshold for approximation coefficients. This threshold is adaptive to every comparison between the reference frame and each comparative frame see Eq. (10).
4. The enhancement of the details coefficients by using suppression coefficients [36] before applying OTSU threshold [38].
5. The use of a 2D Gaussian kernel in the final stage to smoothen the final image.

Algorithm 1 summarizes the steps of the approach.

Algorithm 1 Proposed approach

initialization: A_k, tr, n, s
for $i \in \mathbb{V}$
  Compute $A_1^k(i)$, $D_{11}^k(i)$, $D_{12}^k(i)$, $D_{13}^k(i)$ by Eq. (9)
  Apply threshold Eq. (11) to $A_1^k(i)$
  Multiplify $D_{1n}^k(i)$ by $sc_n^k(i)$ Eq. (12)
  Apply Otsu threshold to $D_{11}^k(i)$, $D_{12}^k(i)$, $D_{13}^k(i)$
  Inverse wavelet transform Eq. (13)
  Update D_k by Eq. (14)
Compute the final image I by Eq. (15)
Apply 2D Gaussian smoothing kernel to I

5. Experimental results

To evaluate the proposed approach, we consider PETS2009 [39] as benchmark dataset. PETS2009 provides videos sequences of walking pedestrians captured from different views.

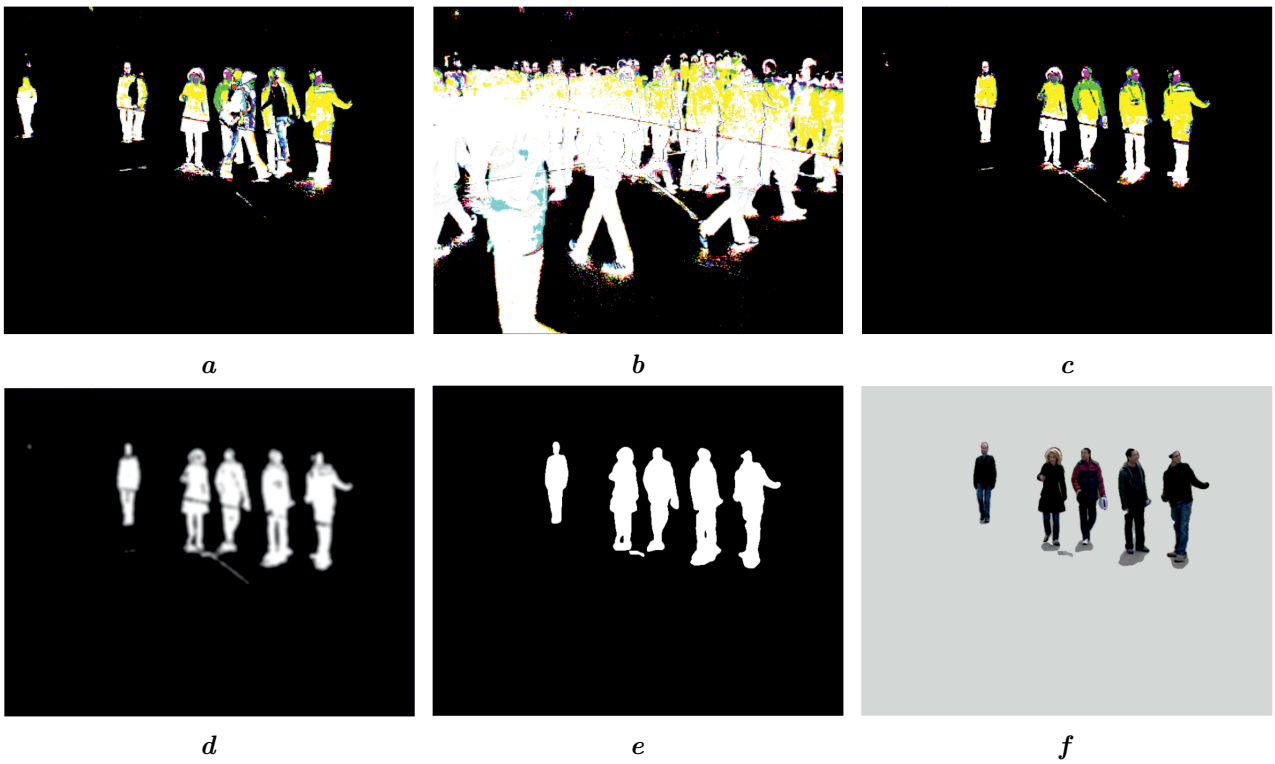


Fig. 2. The intermediate steps of the proposed algorithm.

The steps of our algorithm are illustrated in Figure 2. Figure 2a is an example of a comparison between the reference frame and a picked frame, Figure 2b presents the accumulated comparisons, Figure 2c is the threshold image, Figure 2d is the results of applying the 2D Gaussian smoothing kernel to threshold image, Figure 2e is the obtained motion mask after binarization, and Figure 2f is the final image.

To validate our method, we compare it with two methods, classic memory based frame differencing (CMFD) and the wavelet enhanced accumulative frame differencing (WAFD) [28]. The WAFD uses memory based accumulative frame differences where the final memory image is threshold using wavelet shrinkage, while CMFD does not use wavelet shrinkage, and computes the accumulative frame difference as follows:

$$\mathrm{AFD}(x,y) = \sum_{i=1}^{L-1} FD_i(x,y),$$

where

$$FD_i = \begin{cases} 1 & \text{if} \quad |f_k - f_i| > \tau, \\ 0 & \text{if} \quad |f_k - f_i| \leqslant \tau, \end{cases}$$

k is the index of the reference frame, i is the index of the difference image, τ is a threshold, and L is the number of frames in the sequence.

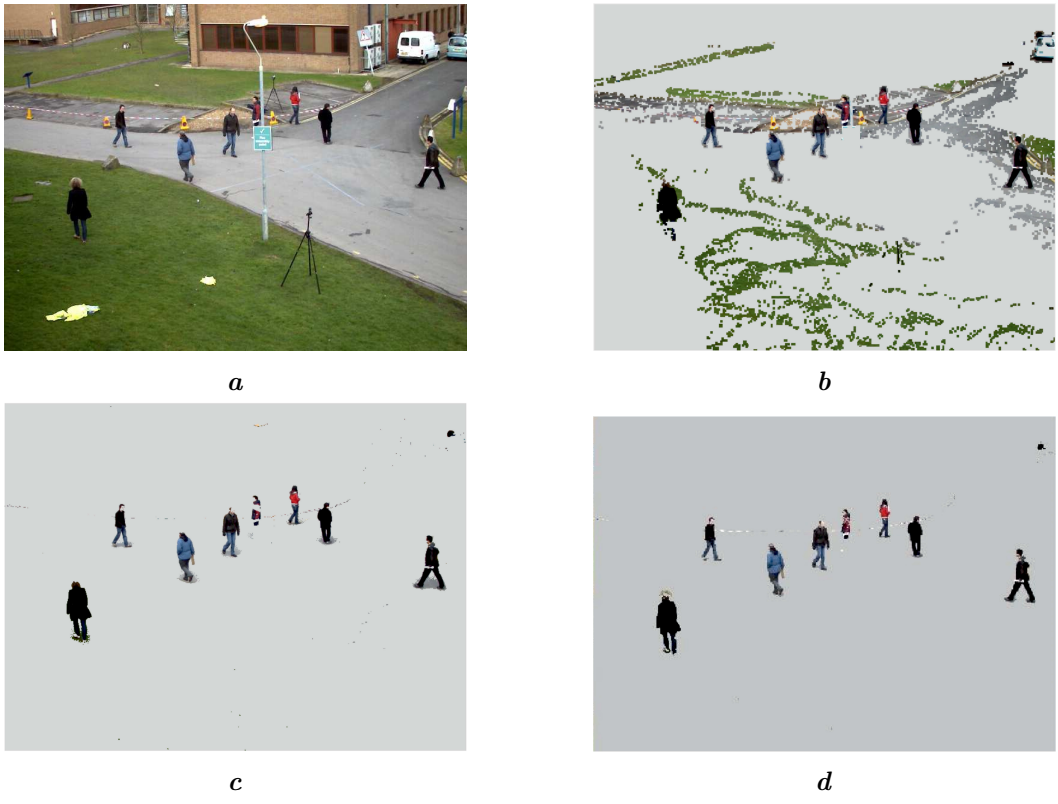
The final image is obtained by the formula

$$\mathrm{FG}(x,y) = \begin{cases} 1 & \text{if} \quad \mathrm{AFD}(x,y)/(L-1) > Th, \\ 0 & \text{otherwise}, \end{cases} \tag{16}$$

where Th is a threshold that was set to 50%.

6. Discussion

We discuss the obtained results of three methods based on the visual quality. In doing so, we consider an image from PETS2009-View001 which is the most similar view to a sequence taken from a surveillance camera. We present in Figure 3 the obtained result of the background subtraction using three methods, where Figure 3a is the reference image, Figure 3b is the result obtained by WAFD, Figure 3c is the result obtained by CMFD, and Figure 3d is the result obtained by our method. From the visual comparison, we observe that our method exhibits more accurate background subtraction with respect to the comparative methods, which fail to segment the motion in several parts of the image.



a ***b*** ***c*** ***d***

Fig. 3. The background subtraction result on PETS2009-View001 [39].

To evaluate the robustness of the proposed approach, another experiment is conducted, in which we compare the obtained motion masks with the ground truth frames. On a frame taken from PETS2009-View008 [39], illustrated in Figure 4, we compare the motion masks obtained by three methods, where Figure 4a is the reference frame, Figure 4b is the ground truth image, Figure 4c is the mask obtained by WAFD, Figure 4d is the mask obtained by CMFD, and Figure 4e is the mask obtained by our method. Obviously, the visual comparison shows that our method achieves a better motion segmentation than the comparative methods.

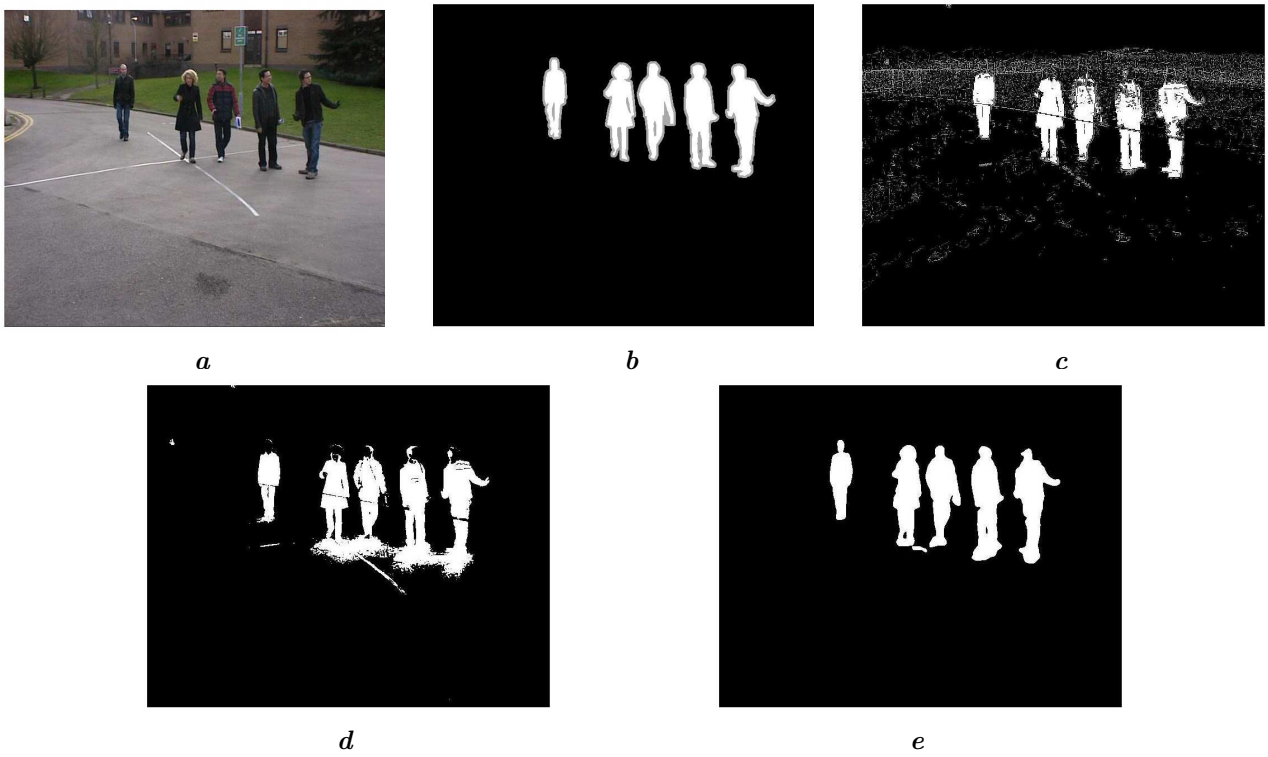
For a deeper comparison, and to further confirm the visual conclusions, we will use the Peak signal-to-noise ratio (PSNR) which is defined as follows:

$$\mathrm{PSNR}(I_{gt}, I_{res}) = 10 \log_{10} \frac{\max(I_{gt})^2}{\|I_{gt} - I_{res}\|_2^2}, \tag{17}$$

where I_{gt} is the ground truth image and I_{res} is the motion mask obtained after background subtraction.

Table 1. PSNR comparison of the three methods.

	WAFD	CMFD	OUR
PSNR	14.4754	14.6262	**19.0605**



a ***b*** ***c*** ***d*** ***e***

Fig. 4. The background subtraction result on PETS2009-View008 [39].

The values of PSNR show that our method exceeded the other two methods and have been able to separate foreground and background pixels more accurately, which confirms the visual conclusions. For a fair comparison, we should mention that the WAFD proposed in [28] exceeds the performance of the CMFD, however in our tests CMFD outperforms WAFD. This can be explained by the fact that we used a different data set and we did not have access the optimal parameters that give the best results for WAFD.

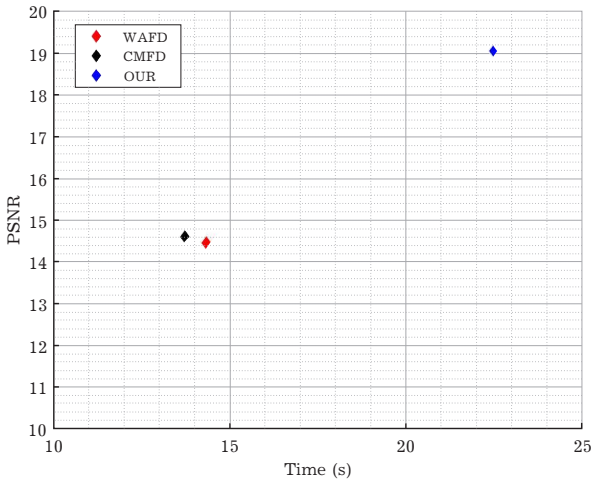


Fig. 5. PSNR versus elapsed time.

In Figure 5 the PSNR is plotted versus the elapsed time for three methods. The comparison shows that even though our method achieves a higher accuracy, two comparative methods are faster. This can be explained by the fact that direct and inverse discrete 2D wavelet transforms are performed at each iteration of the method which requires more computational cost.

7. Conclusion

In this paper a motion segmentation approach based on successive accumulative 2D wavelet comparisons is proposed. The performance of the proposed method was investigated on real-world sequences presenting pedestrians in motion from different views. Then, the results of our algorithm were compared to two reference methods using suitable measures for the investigated scenarios. The proposed approach demonstrated that it improves the motion segmentation precision while costing more in term of computation. Investigation of alternative methods to speed up the process may be one of the perspectives.

[1] Cheng V., Kehtarnavaz N. A smart camera application: DSP-based people detection and tracking. Journal of Electronic Imaging. **9** (3), 336–346 (2000).

[2] Inaguma T., Saji H., Nakatani H. Hand motion tracking based on a constraint of three-dimensional continuity. Journal of Electronic Imaging. **14** (1), 013021 (2005).

[3] Makris D., Ellis T. Path detection in video surveillance. Image and Vision Computing. **20** (12), 895–903 (2002).

[4] Makris D., Ellis T. Learning semantic scene models from observing activity in visual surveillance. IEEE Transactions on Systems, Man, and Cybernetics. Part B (Cybernetics). **35** (3), 397–408 (2005).

[5] Paragios N., Chen Y., Faugeras O. Handbook of Mathematical Models in Computer Vision. Springer New York, NY (2006).

[6] Baghaie A., D'Souza R., Yu Z. Dense Descriptors for Optical Flow Estimation: A Comparative Study. Journal of Imaging. **3** (1), 12 (2017).

[7] Kushwaha A., Khare A., Prakash O., Khare M. Dense optical flow based background subtraction technique for object segmentation in moving camera environment. IET Image Processing. **14** (14), 3393–3404 (2020).

[8] Cucchiara R., Grana C., Piccardi M., Prati A. Detecting moving objects, ghosts and shadows in video streams. IEEE Transactions on Pattern Analysis and Machine Intelligence. **25** (10), 1337–1342 (2003).

[9] Warnell G., Chellappa R. Compressive Sensing in Visual Tracking. IntechOpen (2012).

[10] Song Y., Noh S., Yu J., Park C., Lee B. Background subtraction based on Gaussian mixture models using color and depth information. 2014 International Conference On Control, Automation and Information Sciences (ICCAIS 2014). 132–135 (2015).

[11] Elgammal A., Duraiswami R., Harwood D., Davis L. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of the IEEE. **90** (7), 1151–1163 (2002).

[12] Hoogendoorn E., Crosby K. C., Leyton-Puig D., Breedijk R. M. P., Jalink K., Gadella T. W. J., Postma M. The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. Scientific Reports. **4**, 3854 (2014).

[13] Mohaoui S., Raghay S., Hakim A. Bi-dictionary learning model for medical image reconstruction from undersampled data. IET Image Processing. **14** (10), 2130–2139 (2020).

[14] Ophir B., Lustig M., Elad M. Multi-Scale Dictionary Learning Using Wavelets. IEEE Journal of Selected Topics in Signal Processing. **5** (5), 1014–1024 (2011).

[15] Kim K., Chalidabhongse T. H., Harwood D., Davis L. Background modeling and subtraction by codebook construction. 2004 International Conference on Image Processing, 2004. ICIP'04. **5**, 3061–3064 (2004).

[16] Stauffer C., Grimson W. E. L. Adaptive background mixture models for real-time tracking. Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). **2**, 246–252 (2007).

[17] Elgammal A., Duriswami R., Harwood D., Davis L. S. Background and foreground modelling using nonparametric kernel density estimation for visual surveil. Proceedings of the IEEE. **90** (7), 1151–1163 (2002).

[18] Zhao C., Wang X., Cham W.-K. Background Subtraction via Robust Dictionary Learning. EURASIP Journal on Image and Video Processing. **2011**, 972961 (2011).

[19] Azmat S., Wills L., Wills S. Parallelizing Multimodal Background Modeling on a Low-Power Integrated GPU. Journal of Signal Processing Systems. **88**, 43–53 (2017).

[20] Grossmann A., Morlet J. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. SIAM Journal on Mathematical Analysis. **15** (4), 723–736 (1984).

[21] Carmona R., Hwang W.-L., Torrésani B. Practical Time-Frequency Analysis. **9**, 441–464 (1998).

[22] Chui C. Wavelet analysis and its application. Wavelets. **2**, 725 (1992).

[23] Mallat S. A Wavelet Tour of Signal Processing. The Sparse Way. Academic Press, Inc. (2008).

[24] Carpine R., Ientile S., Vacca N., Boscato G., Rospars C., Cecchi A., Argoul P. Modal identification in the case of complex modes – Use of the wavelet analysis applied to the after-shock responses of a masonry wall during shear compression tests. Mechanical Systems and Signal Processing. **160**, 107753 (2021).

[25] Davies D., Palmer P., Mirmehdi M. Detection and Tracking of Very Small Low Contrast Objects. In M. Nixon and J. Carter (editors), Proceedings of the British Machine Conference, 60.1–60.10 (1998).

[26] Graps A. An introduction to wavelets. IEEE Computational Science and Engineering. **2** (2), 50–61 (1995).

[27] Khare M., Srivastava R., Khare A. Moving object segmentation in Daubechies complex wavelet domain. Signal, Image and Video Processing. **9**, 635–650 (2015).

[28] Al-Berry M. N., Salem M. A.-M., Ebied H. M., Tolba M. F., Hussein A. S. Wavelet-enhanced detection of small/slow object movement in complex scenes. 2016 11th International Conference on Computer Engineering & Systems (ICCES). 172–180 (2016).

[29] Jalal A. S., Singh V. A framework for background modelling and shadow suppression for moving object detection in complex wavelet domain. Multimedia Tools And Applications. **73**, 779–801 (2014).

[30] Sunkara J. Selection of Wavelet Basis Function for Image Compression – A Review. Electronic Letters on Computer Vision and Image Analysis. **18** (1), 1–20 (2019).

[31] Mallat S. G. A theory of multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. **11** (7), 674–693 (1989).

[32] Akansu A., Haddad R. Multiresolution Signal Decomposition. Transforms, Subbands, and Wavelets. Academic Press (2001).

[33] Mallat S. A Wavelet Tour of Signal Processing. The Sparse Way. Academic Press (2009).

[34] Bobulski J. Wavelet Transform in Face Recognition. Springer, Boston, MA (2006).

[35] Bnou K., Raghay S., Hakim A. A wavelet denoising approach based on unsupervised learning model. EURASIP Journal On Advances In Signal Processing. **2020**, 36 (2020).

[36] Chi M., Han X., Xu Y., Wang Y., Shu F., Zhou W., Wu Y. An Improved Background-Correction Algorithm for Raman Spectroscopy Based on the Wavelet Transform. Applied Spectroscopy. **73** (1), 78–87 (2018).

[37] Donoho D. L., Johnstone I. M. Adapting to Unknown Smoothness via Wavelet Shrinkage. Journal of the American Statistical Association. **90** (432), 1200–1224 (1995).

[38] Sezgin M., Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging. **13** (1), 146–168 (2004).

[39] Ferryman J., Shahrokni A. PETS2009: Dataset and challenge. 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. 1–6 (2009).

Адаптивна накопичувальна модель розрізнення кадрів на основі вейвлет-стиснення для сегментації руху

Лахгазі М. Дж.[1], Хакім А.[1], Аргул П.[2]

[1] *Факультет наук і техніки, Університет Каді Айяд, Марракеш, Марокко*
[2] *MAST-EMGCU, Університет Гюстава Ейфеля, IFSTTAR, F-77477 Марн-ла-Валле, Франція*

Сегментація руху в сценах реального світу є фундаментальним компонентом комп'ютерного зору. Існує безліч алгоритмів розпізнавання руху, кожен з яких має різні степені точності та обчислювальної складності. Найпоширенішими методами у випадку статичних камер є методи, засновані на різниці кадрів. Ці методи мають значну слабкість, коли мова йде про виявлення об'єктів, що повільно рухаються. Тому в цій статті подано новий підхід, який спрямований на покращення сегментації руху, пропонуючи метод розрізнення кадрів на основі накопичувального вейвлета. Крім того, у запропонованому підході використовується комбінація декількох методів для ефективного підвищення якості результатів сегментації руху. Ефективність цього підходу на реальних відеопослідовностях показує, що порівняння кадрів за допомогою двовимірного вейвлет-перетворення підвищує якість сегментації руху.

Ключові слова: *сегментація руху; вейвлет-перетворення; вейвлет-стиснення; адаптивний поріг; кадрова різниця; віднімання фону.*