A FAST VOP EXTRACTION TECHNIQUE BASED ON WAVELET TRANSFORM AND WATERSHED SEGMENTATION

Alireza Tavakkoli
Sharif University of Technology
a_tavakkoli@ee.sharif.edu

Shohreh Kasaei
Sharif University of Technology
skasaei@sharif.edu

Gholamreza Amayeh
Sharif University of Technology
amayeh@ee.sharif.edu

Abstract: *In this paper, a very fast, noise robust and accurate video object plane (VOP) extraction algorithm based on wavelet transform and watershed segmentation techniques is proposed. The proposed algorithm firstly, applies the wavelet transform on incoming frames and uses the approximation coefficient matrix of the first level through out the algorithm. This not only increases the speed of the algorithm but also reduces the camera noise. The proposed method subsequently uses a prediction method to apply the watershed algorithm on those areas of the sub-image that might contain a moving object. This prediction stage increases the speed of the segmentation process efficiently. The experimental results show the superiority of the proposed algorithm when compared with other available schemes.*

Keywords:
Wavelet Transformation, Change Detection Mask, Video Object Plane, Watershed Segmentation.

1 INTRODUCTION

In many applications regarding content-based functionalities one major task, called *VOP* extraction, should be implemented. This task, as it represents a very high level description of the objects in a video, is very difficult to be done fully automatically. Recently many researches have been focused on this field and many algorithms have been developed to extract the *VOPs*, but few of them are fully automatic and their performance is not completely satisfactory; regarding to noise robustness and speed.

As proposed by [1], the *VOP* extraction algorithm can be applied in two different ways; I) using *morphological motion filters* (MMF) and II) computing *change detection masks* (CDMs). The CDM is the difference between two successive frames in a video. For a video with stationary background (not moving background), the most efficient way to extract the *VOPs* is the second way; based on the (CDMs). In order to compute a noise robust CDM, a *maximum connected component* (MCC) labeling method is developed in [1], while [2] and [3] use two statistical approaches to achieve noise robustness. The main drawbacks of these algorithms are that some of them only extract the greatest moving object while the others are so slow that cannot be used in real-time applications.

In order to speed up the *VOP* extraction algorithms, many attempts have been done recently, within which the most efficient ones

are proposed in [4] and [5]. These algorithms are based on the Spatio-Temporal segmentation.
Here, a fast algorithm based on spatio-temporal segmentation is proposed. The spatial segmentation is implemented using the watershed algorithm [6]. For making this method efficient for real-time applications, firstly the algorithm is applied to the approximation coefficients of the first level of the wavelet transform to reduce the computation cost of the algorithm by a factor of four and also to de-noise the image. Finally, a watershed update procedure is proposed and applied only on the selected regions of the frame instead of the whole frame. For extracting the VOPs from these determined segments, a temporal matching procedure based on noise reduced CDM is performed on the segmented frames.
As the proposed algorithm runs on the approximation sub-image of the first level of the wavelet transformation of that frame and then only on a small portion of each sub-image, it is fast enough to be used in real-time applications. Furthermore, it does not need to compute an initial model and to track any of the moving objects in the video. As a result, the computation cost of the proposed method has efficiently reduced while its performance is promising.
In the following, we will introduce the proposed algorithm in Section 3 while Section 4 shows the experimental results of the algorithm and compares it with some of the best available works reported in the literature. In Section 5, a brief conclusion is drawn.

2 PREVIOUS WORKS

As mentioned above, many works have been done in the filed of VOP extraction, but there is no simple, fast and fully automatic algorithm to extract VOPs for real-time applications such as distance learning and video telephony. Algorithms like those proposed in [1, 2, 3] are based on CDM and model tracking. While using the CDMs instead of the MMFs will fasten the algorithm, however using an object tracker like the Hausdorff object tracker introduces almost high computation cost and will make the algorithm too slow to be fitted in real-time applications. Also, these methods are sensitive to noise and thus some noise reduction and noise robustness methods are applied in these algorithms.
Some other VOP extraction methods are applied in [4,5] that are based on spatio-temporal segmentation. These methods firstly use a spatial segmentation algorithm and find all objects in the scene and then by using temporal information the spatially segmented frames are used to extract the VOPs. The main drawback of the algorithm in [4] is its low speed, because it uses watershed segmentation (which has a very high computational cost) in addition to the usage of a region merging procedure which is very time consuming too.

3 PROPOSED ALGORITHM

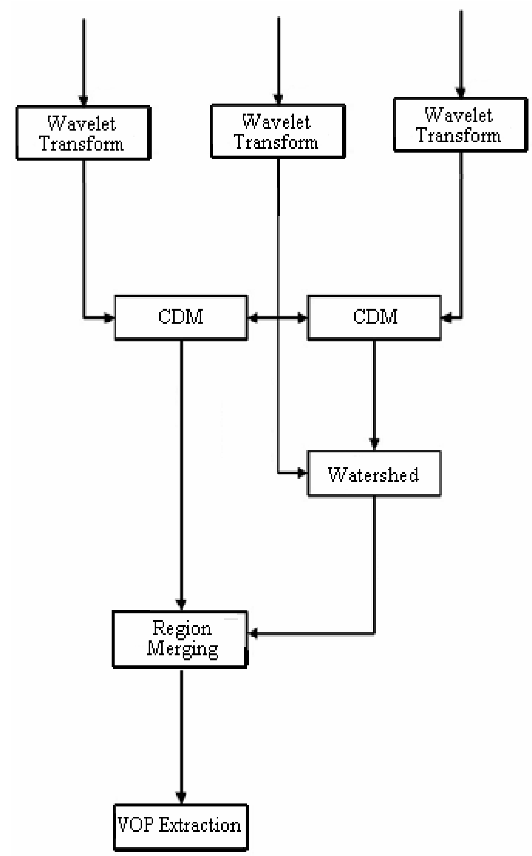

Fig. 1. Block diagram of the proposed method.

In order to reduce the noise sensitivity and also to fasten the algorithm to be applicable in real-time applications, a very fast and noise robust method is proposed based on spatiotemporal segmentation and wavelet transformation. Figure 1 shows the block diagram of the proposed method.

As shown in figure 1, the proposed method is consists of four major parts, the wavelet transformation, the CDM computation, the watershed segmentation, and the VOP extraction. In the proposed method the background and two successive frames are transformed in the wavelet domain and the approximation coefficients of these frames are computed to be used in the rest parts of the algorithm.

The usage of wavelet transform reduces the dimension of input frames as well as the camera noise. This leads to the reduction of the noise and computation cost of the algorithm while the speed of the algorithm will be dramatically increased.

The rest of the algorithm is performed on the average sub-image of level one of the wavelet domain. This consists of computing the CDMs between two successive frames and between each frame and the background frame. Where, the CDM between two successive frames is used to track the moving objects in the image sequence while the other CDM, computed between each frame and the background frame, is used in the region merging stage, discussed later.

For each frame, two major CDMs must be computed. As it will be shown later the CDM between each frame and its previous frame is computed and a simple hypothesis test is performed to smooth the CDM. Then, for the computed CDM, a whole convex rectangle which fits the changing parts is computed and applied to predict the moving parts of the video; to be segmented in the subsequent watershed segmentation stage. By applying this prediction stage, the computation cost of the watershed stage is efficiently reduced by about a factor of ten; depending on the size of the moving objects.

The other CDM, computed in this stage is used for final decision made to extract the exact boundaries of the moving objects, using a region merging in the watershed segmented frames. The test, done in this stage, is more complex than the hypothesis test applied on the CDM computation between two successive frames. This test is based on the Markov properties of the image and the Gaussian properties of noise. By applying this stage a decision is made by considering some parts of the frame as moving objects in the image sequence. More details on this stage are stated in Section 3.4.

The final stage of this algorithm is the VOP extraction stage in which the segmented regions of the frame considered as moving regions are labeled, the resulting objects are then smoothed and the final results are transformed back to the spatial domain; from the wavelet domain. In the following subsections more details about the various stages of the algorithm are given.

3.1 Wavelet Transformation

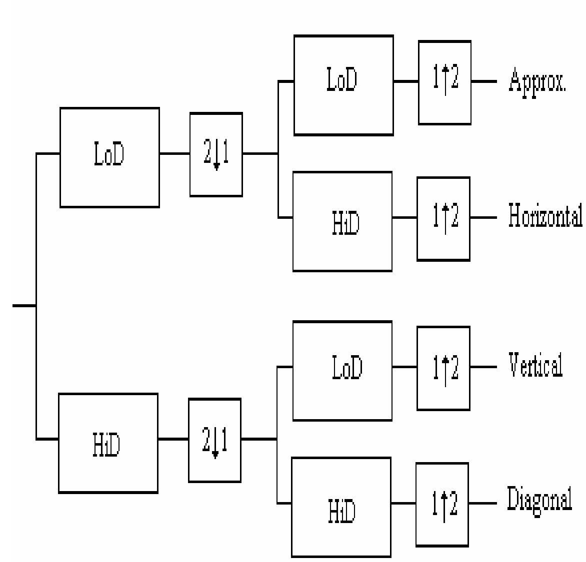

Fig. 2. Block diagram of the wavelet transform.

The block diagram of the wavelet transformation used in the proposed algorithm is shown in figure 2. The wavelet filter used in this stage of the algorithm is the one tap Daubechies filter [7].

As shown in this figure, the rows of the input image are first filtered using a low-pass and a high-pass filter and then are down-sampled by a factor of 2. Each down-sampled row is then further processed using the same procedure on the columns of the resulting sub-images to produce the four output sub-images. Thus, the CA_{j+1} sub-image, called the approximation coefficient matrix is computed by low-pass filtering the frame twice successively. This matrix contains the low-pass components of the frame in both vertical and horizontal directions. As this sub-image contains the average components of the frame, the high-pass components including the camera noise are removed while the dimensions of the frame are reduced by a factor of four too; leading to increase the speed of the whole algorithm.

3.2 Change Detection Masks

The CDM stage of the algorithm is applied using the algorithm discussed in [3]. The major difference between the algorithms in this paper and the one proposed in [3], in computing CDM, is that we remove the noise with the hypothesis test performed on each pixel intensities. Here, we assume that each pixel is changed due to noise. As shown in [3], in this case the intensity of the corresponding pixel in the CDM must obey the Gaussian rule of: Region Merging and Object Extraction:

$$P\left(d_i|H_0\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{d_i^2}{2\sigma^2}\right\} \qquad (1)$$

By computing the variance of the camera noise (which is twice the value of sigma), the noise pixels are removed from the CDM and the CDM will be smoothed. In this stage the convex rectangle which contains the most pixels considered as moving are formed. This rectangle is used in predicting the areas containing the moving objects; on which the watershed segmentation will be performed next.

3.3 Predictive Watershed Segmentation

In this stage of the algorithm, a predictive watershed segmentation algorithm is performed on the frame to segment those regions in each frame which are moving through the image sequence. This stage is called a predictive watershed because the watershed segmentation is performed only on a portion of the frame in which there exist some moving objects. The presence of a moving object is detected by using the previous stage; the CDM.

As a result of this prediction, only a little portion of each frame is considered in the watershed segmentation procedure which leads to a significant increase in the speed of the over all algorithm.

3.4 Region Merging and Object Extraction

After segmenting the parts of each frame that contains a moving object, a region merging algorithm (based on the temporal information of the image sequence) is implemented to determine the regions that belong to a moving object and are not a part of the background.

As mentioned in Section 3.2, when computing the difference between each pixel of one frame in the image sequence and its corresponding pixel in the previous frames, this value will have a Gaussian characteristic; considering that there is no difference between the two frames. According to this fact, if a new object is appeared in one frame and does not exist in another frame, the value of this difference in the location of the newly appeared object does not obey the Gaussian rule shown in equation (1).

So if there is no difference between two frames, which means that there are no newly appeared objects, for a window of size

$w \times w$ the value shown in equation (2) has a χ^2 distribution of the form:

$$D_\chi = \sum_{W} \frac{|Frame_i - Frame_j|^2}{\sigma^2} \qquad (2)$$

considering $Frame_j$ in equation (2) as the background of the sequence (which is the first frame of the sequence), the value of D_χ will be so close to zero if the window does not belong to a newly appeared object in $Frame_i$ It will have an almost greater value if the region belongs to a newly appeared object. This is because of the fact that the density of D_χ around zero is much more than that around the higher values.

As a result of the predictive watershed segmentation stage, those regions which belong to moving/changing parts of the sequence are segmented and finally by performing the region merging stage on these regions, only those regions belonging to the newly appeared objects are extracted.

3.5 VOP Extraction

In this stage, the regions extracted from the previous stages are smoothed with a morphological operation. By using the previous extracted VOPs and considering the fact that the VOP must not change abruptly, those parts of the moving objects which are lost can be recovered. This algorithm considers the moving objects extracted from previous frames and compares them with the objects extracted in the current frame, and refines the object to recover the probably lost parts of the moving objects. This leads to a very fine, accurate, and fast VOP extraction method in an image sequence containing a stationary background.

4 EXPERIMENTAL RESULTS

In this stage, the experimental results of the proposed algorithm are discussed. The algorithm was run on some typical sequences. Here the result of the algorithm when run on the *Hall & Monitor* sequence is shown. As shown in figure 3, the watershed segmentation is performed on only a portion of the whole frame whose dimension is reduced by a factor of four due to the usage of the wavelet transform.

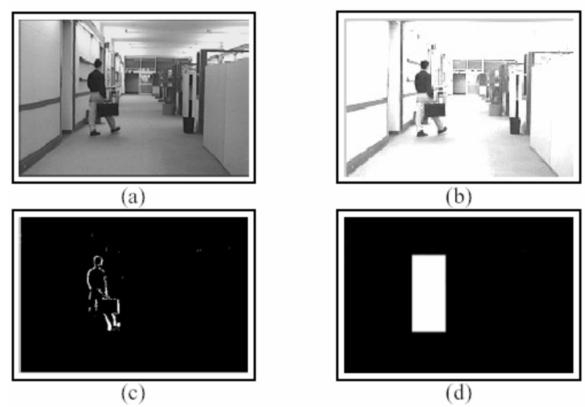

Fig. 3. Results of each stage applied on Frame of the Hall & Monitor sequence: (a) Original Image. (b) Approximation coefficients of the first level of the wavelet transform . (c) The CDM between two successive frames. (d) Region on which the watershed segmentation is applied.

As shown in Table 1, if the watershed algorithm performs on the whole frame, it must consider all pixels which leads to a very time consuming operation. As shown in this table, in this case it should run on 84480pixels resulting in 389 regions. Note that this also decreases the speed of the algorithm when the region merging stage is subsequently implemented on these regions. In contrary, using the proposed predictive watershed algorithm on the approximation level of the wavelet transform dramatically decreases the computation cost of the whole algorithm. This can be seen in table 1, when comparing the 20 resulting regions in the proposed method with the 389 regions if the prediction and wavelet transformation stages where bypassed.

Table 1: Two different segmentation algorithms run on frame 41.

	No. of Pixels	No. of Regions
Whole-Frame Watershed	84480	389
Wavelet Transform	21120	226
Proposed Watershed	656	20

The final VOP extracted from Frame 41 of the sequence is shown in figure 4. As shown in this figure, the VOP extracted with the proposed method is considerably accurate, and extraction of it take dramatically less time than other available algorithms such as algorithm proposed in [1].

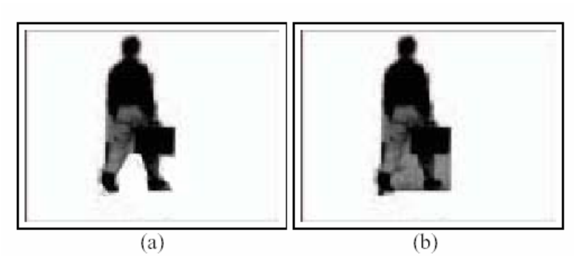

Fig. 4. (a) The VOP extracted by the proposed algorithm. (b) The VOP extracted by [1].

5 CONCLUSION

In this paper, by using the approximation coefficient matrix of the first level of the wavelet transform and applying a CDM along with a hypothesis test on the resulting sub-image, the regions containing the moving objects in the frame are efficiently predicted. Performing the watershed segmentation on the prediction regions (instead of the whole image) increases the speed of the whole algorithm dramatically. Furthermore, as the number of the regions extracted in the prediction region is considerably less than the total number of regions in each frame, the speed of the region merging algorithm, to extract those regions belonging to the moving objects, is incredibly increased too. As show in the experimental result section, the combination of the wavelet transform and predictive watershed segmentation lead to a very fast, noise robust and accurate VOP extraction algorithm for image sequences with stationary background.

ACKNOWLEDGEMENT

This research was partially supported by a grant from ITRC.

REFRENCES

[1] K.N. Ngan, T. Meier, , and D. Chai, *Advanced Video Coding: Principles and Techniques.*, Elsevier, Amsterdam, 1999.

[2] Y. Yong et al., “An object-based segmentation approach for very low bit rate video sequence.,” *Proc. Int. Sym. On Intelligent Multimedia, Video and Speech Processing*, pp. 205–208., 2001.

[3] R. Mech and M. Wollborn, “A noise robust method for 2-d shape estimation of moving objects in video sequences considering a moving camera.,” *Signal Procesing*, vol. 66, pp. 203–217., 1998.

[4] Y. Tsiag and A. Averbuch, “Automatic segmentation of moving objects in video sequences: A region labeliong approach.,” *IEEE Trans. on Circuits and Systems for Video Technology,*, vol. 12, pp. 597–612., 2002.

[5] Yu-Wen Huang Shao-yi Chien and Liang-Gee Chen, “Predictive watershed: A fast watershed algorithm for video segmentation.,” *IEEE Trans. on Circuits and Systems for Video Technology,*, vol. 13, pp. 453–461., 2003.

[6] L. Vincent and P. Soille, “Watershed in digital spaces: an efficient algorithm based on immersion simulation.,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. Vol. 13, pp. 583–598., June 1991.

[7] Daubechies, “Ten lectures on wavelets.,” *SIAM:CBMSNSF.*, vol. Conference series in applied mathematics, Pennsylvania, 1992.