Motion Detection using Wavelet-SIFT Features

S. Mohamed
Electronics & Communications Dept.
Faculty of Engineering, Cairo University
Giza, Egypt.

M. Hesham
Engineering Math. & Physics Dept.
Faculty of Engineering, Cairo University,
Giza, Egypt,

M. Fekri
Communication & Electronics Dept.,
Faculty of Engineering, Cairo University,
Giza, Egypt,

Abstract—**Object tracking from video becomes a very important topic in the field of computer vision. The problem of matching is an important issue in object tracking applications. Therefore, this paper presents an algorithm to improve the matching stage through using wavelet transform and scale invariant feature transform (SIFT). This algorithm extracts the wavelet LL coefficients, then extracts the SIFT features from the frames and finally calculates the matching between the extracted features from the consequent frames. It is applied on many consequent frames and under different circumstances and compared with different procedures such as wavelet, sift, wavelet with sift. The results demonstrate how the percentage of matching is improved, and then compared it to the other methods and how it is reflected on the performance of simple object tracking technique.**

Keywords—Moving region detection;motion; SIFT; Wavelet Transform;frame difference

I. INTRODUCTION

The object-tracking topic gains considerable attention from researchers because it contributes to many important applications like video compression, surveillance, robot and vehicle localization, and traffic monitoring and human computer interface. Object tracking is the process of finding the location and dynamic configuration of one or more moving objects in each frame (image) of a video. Object tracking is the task of pursuing one or more objects in a scene, from their first appearance to their exit [1].

Many algorithms are proposed in the literature for object tracking. These algorithms can be distinguished based on the way that they are used to handle the issues like segmentations, object representation, image features, occlusion, and the motion modeling.

The motion detection is a very complicated problem. It faces many challenges because of its association with the surrounding changes, which vary from one application to another. For example, noise in the image sequences, changing appearance patterns of the object, changing in scene illumination, object-to-object and object-to-scene occlusions, non-rigid object structures and camera motion[1].

Although many algorithms have been performed in recent years, the subject is still challenging. Some techniques give attention to the preprocessing and feature selection step. This step will reflect the matching performance and give noticeable improvement in the matching feature step.

II. RELATED WORK

Kummar and Sathidevi [2]use wavelet and colour Scale Invariant Features Transform(SIFT) features to propose image matching technique. They extract wavelet coefficients of LL and HH subbands of the image, then apply color SIFT to extract feature descriptors in the wavelet domain. The performance of this algorithm is tested on many pairs of frames at different conditions and it gives good performance compared with Hue-SIFT.

Ashish and Tiwary[3]proposed algorithm symmetric complex wavelet transform for denoising and deblurring the images. It applies the soft-threshold function on multiple levels, which is based on standard deviation, absolute median and absolute mean of wavelet coefficients. It removes the signal independent, signal dependant and color noise from the signal. They test the algorithm and found that the performance is better than the other method, which uses the real wavelet[3].

Jalal and Tiwary[4] propose an algorithm to overcome the problems of illumination variation, change of appearance or pose and noise in the traditional methods. They use a structural similarity index in complex wavelet transform to solve these problems. The algorithm extracts the features of the object in the initial frame, the extracted features is the structural similarity index in complex wavelet domain. Then the similarity measured is used to find the object in the current frame. The results of the proposed algorithm show that it has good performance in the noisy environment and significant variations in the object's pose and illumination[4].

Khansari, Rabiee, Asadi, and Ghanbari,[5] propose an algorithm through using the undecimated wavelet feature and texture analysis. The algorithm selects the region of the intended object in the reference frame, and then generate an undecimated wavelet coefficient to construct the feature vector. The feature vector is used by texture analysis to find the best match in the next frames. The texture analysis determines the direction and the speed of the object motion. This algorithm is robust to Gaussian noise and in partial or full occlusion. The experiment's results give good performance in the crowded area[5].

The proposed Motion detection algorithm employs both SIFT and WT features in a grayscale image. Firstly, the wavelet decomposition is applied to perform multi-resolution analysis on a frame. Therefore, the noise in an image will be reduced. Scale Invariant Feature Transform (SIFT) is then used to extract the features of each frame in the wavelet domain. The features are invariant to image scale, rotation, noise and change in illumination. The extracted features of a

frame are compared with the features of the previous frame to detect moving regions. If there is, no motion then matching between these frames is very close.

This paper has developed a new technique for (preprocessing, feature detection, and extraction blocks of any motion detection) motion detection through digital image sequences in wavelet domain. Section 2 gives an overview of wavelet transform and SIFT-Features. Section 3 presents our new algorithm. In Section 4, the experimental results are presented. Finally, the conclusion is demonstrated in Section 5.

III. OVERVIEW

A. Wavelet Transform

WT can provide analysis on several timescales of the local properties of complex signals that can present non-stationary Zones. They lead to a huge number of applications in various fields, such as geophysics, astrophysics, telecommunications, imagery and video coding. They are the foundation for new techniques of signal analysis and synthesis such as compression and de-noising [6].

Discrete WT (DWT) can be represented after decomposition of an image into four different frequency ranges as shown in Fig. 1and Fig. 2[3], [7].
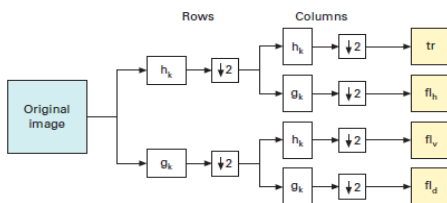

Fig. 1. Sub-band decomposition of Image

Fig. 2. Procedure for the application of the wavelet decomposition

B. SIFT

SIFT is proposed to extract features that are invariant to rotation, scaling and partially invariant to changes in illumination and affine transformation of images. it is used to perform matching of different views of an object or scene[2], [8]. Steps of extracting SIFT features are described as following:

1) Detection of scale-space extrema: this stage detects the keypoints using a cascade filtering approach. Keypoints identify locations and scales that can be repeatedly assigned under differing views of the same object.

2) Keypoint localization: the selection of the keypoints is according to their stability.

3) Orientation assignment: one or more orientation is assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, therefore providing invariance to these transformations[9].

4) Keypoint descriptor: the local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination[9].

IV. THE PRPOSED ALGORITHM

The proposed algorithm is intended to improve the matching, which is reflected on the motion detection from image sequence. The main block diagram for motion detection is shown in Fig.3 and the proposed algorithm is demonstrated in Table 1. It extracts discrete Wavelet transform (WT) coefficients of the images, LL sub band coefficients are used to extract the feature descriptors(SIFT) of images, then compare the descriptors of current frame with the previous through calculation of Euclidian distance between the features to detect the motion. WT acts as a preprocessing stage, and this removes noise and blurring of the input images, which improves the matching percentage. The image information is preserved by LL component and the edge information is preserved by HH [2].

The SIFT features are extracted from the test frame ,then compared with the SIFT features of the reference frame or the background frame[2]. The matching here is computed through obtaining the distance between key point descriptors. Euclidean distance is used to obtain distance d(X,Y) between two vectors X_i and Y_i according to " (1)," [2]:

$$d(X,Y) = \sqrt{\sum_i (X_i - Y_i)^2} \qquad (1)$$

Motion detection is performed through measuring the matching between features descriptors in the consequent frames according to " (2) ". Euclidian distance is used to measure the distance between the features. Therefore, the clutter in the background increases the error of matching between the matched features. So, the best way to select the matched featured is comparing the distance of first with the second closest one, and according to the certain ratio in (3),the featured is either accepted or discarded [9]–[11].

$$\mathbf{m}atchingpercentage = \frac{totalnumberofmatches}{numberoffeaturesintestimage} (2)$$

Video sequence

↓

Convert the frames to Grayscale

↓

Preprocessing

↓

Extract Features

↓

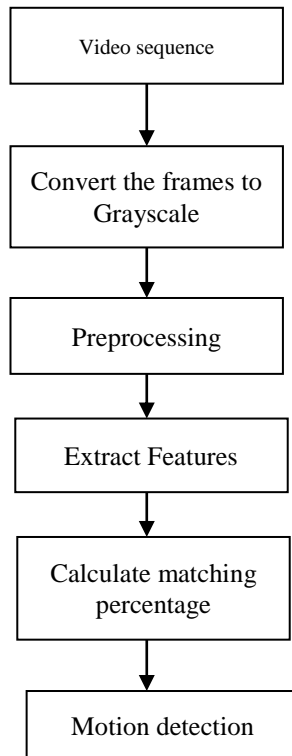Calculate matching percentage

↓

Motion detection

Fig. 3. Block diagram of the procedure

Table 1 Proposed algorithm

The procedure of motion detection

Read frames F_i and F_{i+1}.

Compute wavelet coefficients of F_i and F_{i+1}.

Select LL sub band Coefficients from Wavelet sub bands.

Compute the SIFT features D_i of LL_i sub band coefficients.

Compute the SIFT features D_{i+1} of LL_{i+1} sub band coefficients.

Calculate the matching percentage M between D_i and D_{i+1}.

If M< Threshold

Calculate the difference (net) between the LL_i and LL_{i+1}.

Obtain the position of moving object p=(x,y).

end

V. EXPERIMENT AND RESULT

In this section, the experiments select the best family of wavelet and level, which is suitable for the database through calculating the matching percentage according to (2).Then, it is used with the SIFT features to detect the motion as demonstrated in the following experiments. The experiments are applied on the database of PETS 2001, pets4(camera1), the database for outdoor people and vehicle tracking and it was taken by omnidirectional and moving camera[12]. The database's frames are gray images and with size 71KB/frame, diminutions 768x576 pixel, and JPEG format. The software, which used to execute the experiments are VLFeat toolbox and wavelet toolbox on matlab program. The computations are done through the computer with a microprocessor core I7 and 8 G RAM.

a) Experiment 1

This experiment extracts the SIFT features directly from the frames of the video, and then measures the percentage of matching between the consequent frames as shown in Table 2. The results in Table 2 show the big difference between the frames; however, this is not correct because the motion in these frames is too small to make such a large difference in matching.

Table 2 SIFTmatching percentage

Frames numbers	The percentage of SIFT matching %
1,2	60.2
1,3	62.8
1,4	58.7
1,5	54.8
2,3	59.7
3,4	56.1
4,5	54.62
5,6	58.26
6,7	56.8
7,8	60.02
8,9	57.1
9,10	58.19

b) Experiment 2

Different wavelet families with different family members and different levels are applied on the dataset to get the best one for it. After some experiments, the best family members and level of each family is obtained. The types of families, which are used in this experiment, are Daubechies, Symlets and Coiflets. This experiment obtains wavelet coefficients and SIFT features of the frame, and gets the similarity between the frames through measuring the Euclidian distance. Fig. 4 shows that the DB2 level 2 is more suitable for the data because the transition is proportional to the rate of motion in the frames and the matching percentage is greater than other members in daubechies family. Symlets (sym2) in level2 give results that are more accurate than sym4 in level2 as shown in Fig.5. The Coiflets family (Coif1 and Coif2) in level two gives identical results as evident in Fig.6 shows that the performance of DB2, 2 and Sym2,2 is closer than Coif2,2 and proportional with the motion that was happening in these frames. From this experiment, the DB2, 2 is suitable for the dataset and used in the motion detection procedure.
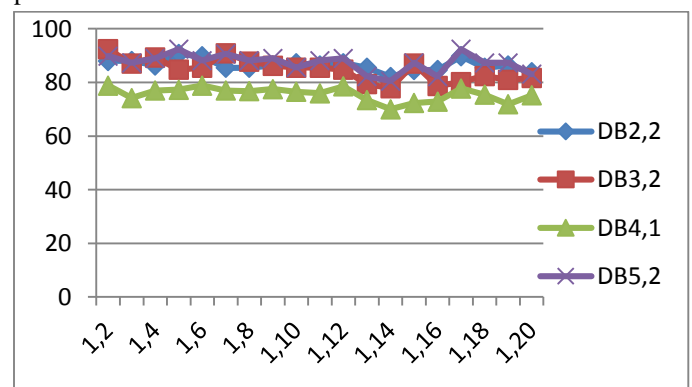

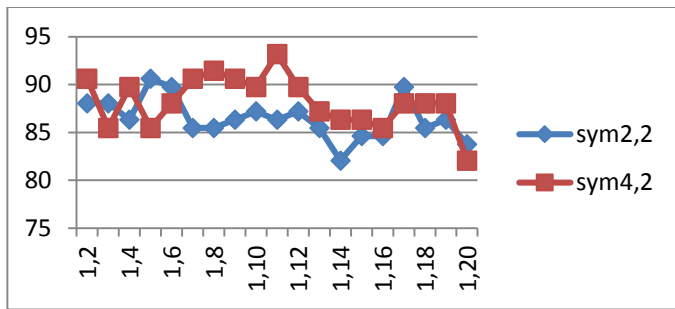Fig. 4. Comparison between Daubechies family in different member and level

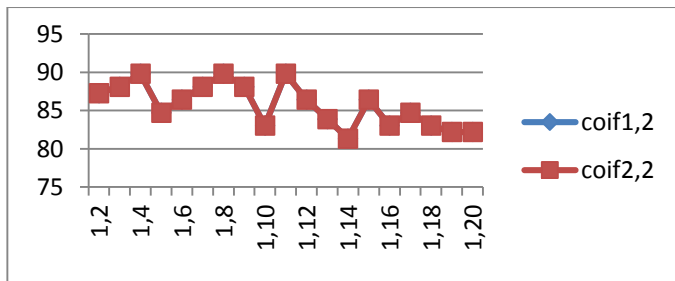Fig.5 Comparison between Symlets family (sym2 and Sym4) in level2

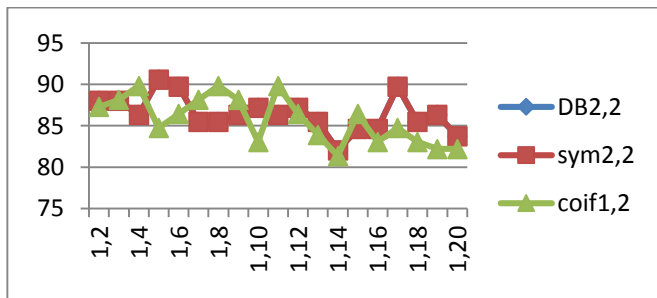Fig.6 Comparison between Coiflets family (coilf1 and coilf2) in level2

Fig.7 Comparison between Wavelet families (DB2, Sym2 and Coif1).

Fig.8 Appearing of moving object from frame 12 to 15
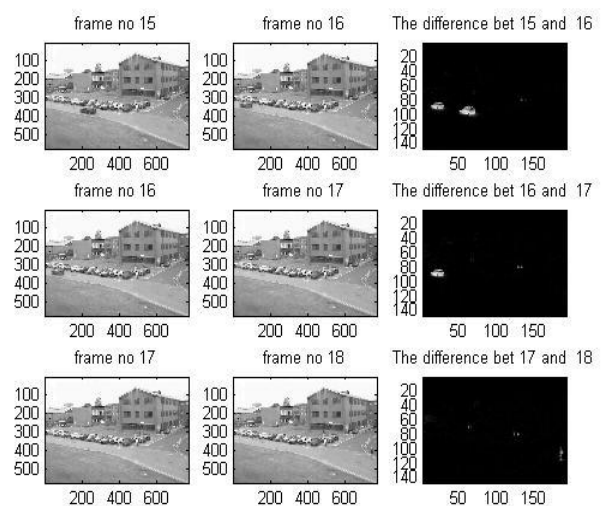
Fig. 9. Appearing of moving object from frame 15 to 18

c) Experiment 3

This experiment shows the motion detection of an object in video and its path starting from its appearance in the frame no 13 until it disappearance in the frame no 17 as shown Fig. 8and Fig.9.

VI. CONCLUSION

WT helps in enhancing an image by removing noise, which reflects the extracted features and motion detection of moving objects in a video as demonstrated in experiment 1 and 2 where the matching percentage increased by almost 20 percent. The SIFT features extracted in the wavelet domain reduce the processing time to detect the motion in frames. The SIFT features, which are used in the experiments, are based SIFT, so we will try to use a reduced version form SIFT such as SIFT-64, SIFT-32 and SIFT-16 to improve the time of processing. As well, there are other features such as PCA-SIFT and SURF will be used to improve the performance comparable to the time.
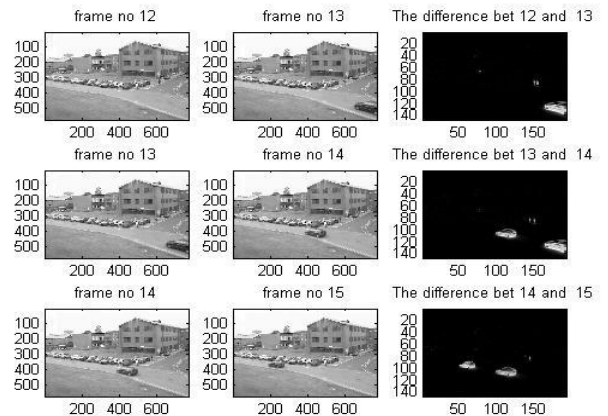
REFERENCES

[1]   A. S. Jalal and V. Singh, *The State-of-the-Art in Visual Object Tracking. Informatica 36:227–248*. 2012.

[2]   N. A. M. Kumar and P. S. Sathidevi, "Image Match Using Wavelet-Colour SIFT Features," in *2012 7th IEEE International Conference on Industrial and Information Systems (ICIIS)*, 2012, pp. 1–6.

[3]   ASHISH Khare and US Tiwary, "Symmetric Daubechies Complex Wavelet Transform and its Application to Denoising and Deblurring," *WSEAS*, vol. 2, no. 5, pp. 738–745, 2006.

[4]   A. S. Jalal and U. S. Tiwary, "A Robust Object Tracking Method Using Structural Similarity in Daubechies Complex Wavelet Domain," in *Pattern Recognition and Machine Intelligence*, S. Chaudhury, S. Mitra, C. A. Murthy, P. S. Sastry, and S. K. Pal, Eds. Springer Berlin Heidelberg, 2009, pp. 315–320.

[5]   M. Khansari, H. R. Rabiee, M. Asadi, and M. Ghanbari, "Object Tracking in Crowded Video Scenes Based on the Undecimated Wavelet Features and Texture Analysis," *EURASIP J Adv Signal Process*, vol. 2008, pp. 102:1–102:18, Jan. 2008.

[6]   M. Misiti, Y. Misiti, and G. Oppenheim, *Wavelets and their Applications*. UK: ISTE, 2007.

[7] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[8] D. G. Lowe, "Object recognition from local scale-invariant features," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, 1999, vol. 2, pp. 1150–1157 vol.2.

[9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] M. M. El-gayar, H. Soliman, and N. meky, "A comparative study of image low level feature extraction algorithms," *Egypt. Inform. J.*, vol. 14, no. 2, pp. 175–181, Jul. 2013.

[11] L. Juan, O. Gwun, L. Juan, and O. Gwun, "A Comparison of SIFT, PCA-SIFT and SURF," *International Journal of Image Processing*, vol. 3, no. 4, pp. 143–152, 2009.

[12] "http://ftp.pets.rdg.ac.uk/pub/PETS2001/." .