Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Moving Multi-Object Tracking Algorithm Based on Wavelet Clustering and Frame Difference

Li Zeng
Institute of Computing Technology, Chinese Academy of Sciences, and Graduate University of Chinese Academy of Sciences, Beijing, China
zengl@ics.ict.ac.cn

Lida Xu
Department of Information Technology and Decision Sciences
Old Dominion University
VA, USA

***Abstract*—The paper presents an approach to motion objects tracking by combining parallel wavelet clustering with frame difference, which satisfies certain requirements of rapid moving objects detection. By utilizing multi-resolution property of wavelet clustering analysis based on adjacent frame difference results, we can identify arbitrary shape moving objects at different degree of accuracy. Experiment results show that good accuracy of proposed algorithm can be obtained at speeds close to real-time. Applications in real world are also presented which further demonstrated the efficiency and effectiveness of the proposed method.**

***Keywords*—Moving Object Tracking, Wavelet Clustering**

I. INTRODUCTION

Recently, there has been a growing interest in the computer vision community towards content-based video understanding, and in particular toward real-time objects detection, tracking or recognition. Moving multi-object tracking is the process of segmenting the motion objects or regions (also called foreground) from the rest of the image (also called background). As a research branch of computer vision, moving multi-object detection and tracking are of importance since motion perception and interpretation constitute the core of any visual analysis systems, such as robot vision systems, self-driven airplane, radar systems, civil video surveillance systems and Intelligent Traffic Systems.

However, the research fields are still far from being completely solved. The challenges are from the domain of artificial intelligence, and computer vision. [1-3] Some issues are opening such as illumination change, real-time limitation, and noise from background. Among prior work, there are two broad classes method of motion objects detection: approaches based on optical flow analysis, and approaches based on change detection (background subtraction and temporal differencing). The advantage of methods based on optical flow analysis is that these methods can be used to detect independently motion objects in the presence of camera motion. But, the disadvantage is that most optical flow computation methods are computationally expensive. In background subtraction, objects are detected through difference between the current frame and a reference background image. Background subtraction provides the most complete feature data, but is sensitive to dynamic scene changes due to lighting. Temporal differencing is similar to background subtraction but the estimated background is just the previous frame. Motion objects are detected through pixel-wise difference between two or three consecutive frames. Temporal differencing is easily adaptive to dynamic environments, but has the problems in extraction of all relevant feature pixels, and is sensitive to the threshold. That is, some methods have been proposed, but the problems of detection and segmentation of motion objects in complex circumstances are far from maturity.

As a result, we propose a simple method of rapid moving objects tracking at multi-scale by combining wavelet clustering with frame difference algorithm. The proposed approach is efficient in real-time application. The computational complexity of generating clusters in our method is $O(N)$. The results are not sensitive to outliers, noises, and the order of the number of multiply objects to be processed. It takes arbitrary temporal change image $f(x, y, t)$ as a basic function, and then imposes a serial of suited constraint conditions and wavelet transform on the two frame difference images for ultimate clustering analysis and objects recognition. The proposed detection method leads to a robust moving objects detection capability. It does not rely on training models or assumption of the scene, which make it a fast, automatic, and robust way to objects tracking in a variety of video surveillance and real-time intelligent traffic applications. The paper is organized as follows. We first shortly introduce the related concepts and principles of the multiresolution analysis and pyramidal decomposition in section two, which are the essential of WaveCluster. In section three, we discuss the wavelet clustering and parallelization process, and show how to combine frame difference with wavelet transform to illustrate our motion objects tracking method. Section four presents the experiment results of tracking the transient moving cars in real circumstances. Finally in the last section, conclusions are offered.

II. MULTIRESOLUTION ANALYSIS AND PYRAMIDAL DECOMPOSITION

As we all know, wavelet transform is a kind of signal processing technique that decomposes a signal into different frequency subbands. It is a type of signal representation that can give the frequency content of the signal at a particular instant of time by convolving the filter. [4] It has advantages over traditional Fourier and windowed Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. When we look at a signal with a small window, we would notice its small features. Analogically, we would get gross features if we look at this

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

signal with a large window. The result in wavelet analysis is to see both the "forest" and the "trees", so to speak. However, wavelet transform is referred to so many convolve operations due to the property of multi-resolution analysis that the calculation complexity increases exponentially. Multiresolution representations are very effective for analyzing the information of image content. Considering the exellent multiresolution property of wavelet transform, we apply wavelet tranform clustering on after-frame-difference binary images for getting rid of the noises and obtaining arbitrary shape moving objects.

Pyramid decomposition algorithm [5] is considered the most important algorithm to calculate DWT coefficients. It plays the similarly important role in wavelet transform as FFT has been in Fourier Transform. Classical algorithm is basically a collection of cascaded FIR filtering operations by a pair of mirror filters and down-sampling procedure in each scale. Given a signal $\{x(n)\}$, we are expected to generate a set of wavelet coefficients $\{d_j^k\}$ and scaling coefficients $\{c_{jo}^k\}$ from a pair of mirror FIR filters $\{h_o(n)\},\{h_1(n)\}$. The computation of one scale includes two sets of coefficients: the wavelet coefficients $\{d_j^k\}$, which are often the detailed representation of the signal, and the scaling coefficients $\{c_{jo}^k\}$, which are often the coarse representation of signal. Pyramidal decomposition algorithm has been proven to be a useful tool in image segmentation and moving objects detection.

III. WAVELET CLUSTERING AND PARALLEL SEARCH POLICY

Wavelet transform [5,6] has very successful applications in image compression, communication and denoising. It's been broadly classified into the discrete wavelet transform (DWT) and the continuous wavelet transform (CWT). The principal difference between the two is the continuous transform operates over every possible scale and translation, whereas the discrete uses a specific subset of all scale and translation values. As continuous wavelet transform often appears in theory research, we choose the first manner.

Wavelet Transform Based Clustering

Wavelet-based clustering presented by [7,8] is a clustering approach of data mining by partitioning the data space into cells and applying wavelet transform on them, which meets many desirable properties of a good clustering technique. It looks at the multidimensional data space from a signal processing perspective. The data in the d-dimensional data space composes a d-dimensional signal.

Let us briefly recall property of multi-resolution of wavelet analysis. Wavelet transforms are integral transforms allowing a multi-resolution analysis by variations of Δt and Δf in the time-frequency domain. The two-dimension continued wavelet transform of a function $f \in L^2(R)$ with the basis wavelet ψ is define by

$$L_\psi f(a,b) = \frac{1}{\sqrt{c_\psi}} |a|^{-1/2} \int_R f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad [9]$$

It means a signal can be analysis by a wavelet basis function ψ constituting L^2 scalar products of f with dilated and translated versions of ψ .

Take 1-dimensional for example; we compute a coarser approximation of the 1-dimensional input signal c_0 by convolving it with the low-pass filter $\overline{h}$ and down sampling the signal by two filters. All the discrete approximations $c_j, 1 < j < J$ (here J is the maximum possible scale) can thus be computed from c_0 by repeating this process. Figure 1 illustrates the method of 1-dimensional wavelet decomposition.

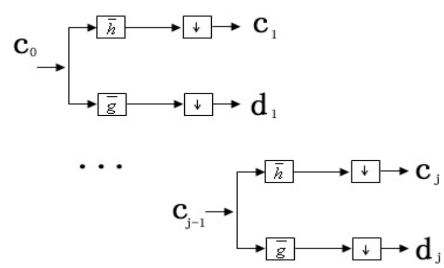

Figure 1. Multi-resolution of wavelet image decomposition

where d_j denotes the difference between c_j and c_{j-1}, and is called detail signal at the scale j . The detail signal d_j can be computed by convolving c_{j-1} with the high pass filter $\overline{g}$, and returning every other sample of output. The wavelet representation of a discrete signal c_0 can therefore be computed by successively decomposing c_j into c_{j+1} and d_{j+1} which $0 \le j \le J$. This representation can provide information about signal approximation and detail signals at different scales. In general, wavelet decomposition and reconstruction can be described as following formulas:

$$\begin{cases} c_k^j = \sum_{l \in \mathbb{Z}} c_l^{j+1} \overline{h}_{l-2k} \\ d_k^j = \sum_{l \in \mathbb{Z}} c_l^{j+1} \overline{g}_{l-2k} \end{cases}$$

$$c_k^j = \sum_{l \in \mathbb{Z}} c_l^j \overline{h}_{k-2l} + \sum_{l \in \mathbb{Z}} d_l^j \overline{g}_{k-2l}$$

where $\{h_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$ denotes the low pass filter, $\{g_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$ denotes the high pass filter. c_k^j is the discrete approximations of c_l^{j+1}, and d_k^j is the detail signal of c_l^{j+1}. The one-dimensional case is easily to generalize for two dimensions or higher.

$$\begin{cases} c_{j,k,m} = \sum_{l,n} \overline{h}_{l-2k}\overline{h}_{n-2m}c_{j+1,l,n} \\ d^1_{j,k,m} = \sum_{l,n} \overline{h}_{l-2k}\overline{g}_{n-2m}c_{j+1,l,n} \\ d^2_{j,k,m} = \sum_{l,n} \overline{g}_{l-2k}\overline{h}_{n-2m}c_{j+1,l,n} \\ d^3_{j,k,m} = \sum_{l,n} \overline{g}_{l-2k}\overline{g}_{n-2m}c_{j+1,l,n} \end{cases}$$

$$\begin{aligned} &c_{j+1,k,m} = \\ &\sum_{l,n} h_{k-2l}h_{m-2n}c_{j,l,n} + \sum_{l,n} h_{k-2l}g_{m-2n}d^1_{j,l,n} + \\ &\sum_{l,n} g_{k-2l}h_{m-2n}d^2_{j,l,n} + \sum_{l,n} g_{k-2l}g_{m-2n}d^3_{j,l,n} \end{aligned}$$

Typical 2-dimensional discrete wavelet analysis applies transform on both rows and columns, i.e. along the axes x and y. In 2-dimensional data space, an image $c_{j+1} = \{c_{j+1,k,m}\}_{k,m}$ is decomposed into an average signal $c_j(LL)$ and three detail signals which are directionally sensitive: $d_j^1(LH)$ emphasizes the horizontal image features, $d_j^2(HL)$ the vertical features, and $d_j^3(HH)$ the diagonal features. c_j can further decomposed into c_{j-1}, d^1_{j-1}, d^2_{j-1} and d^3_{j-1}. By repeating this process, we will get satisfied approximation features and a number of detail features of original 2-dimensional data space.

Discovering arbitrary shape clusters at different degrees of detail is the essential step to our moving objects tracking algorithm. Given a large set of data, the goal of clustering algorithm is often to detect clusters and assign labels to them based on the cluster they belong to. A good clustering approach should be unaffected by outliers (or noise) and should detect them effectively. In addition, clustering algorithms should be insensitive to the order of the input data. Another desirable property for clustering algorithm is the ability to find clusters at different levels of detail which is termed as multi-resolution property. Our motivation for using wavelet transform is drawn from its useful properties: multi-resolution property and effective removal of noise. That is, multiresolution property of wavelet transform can help detecting the clusters at different levels of accuracy. As it will be shown, we can apply wavelet transform which results in clusters at scales from fine to coarse. Clusters in the data after wavelet transform automatically stand out and clear regions around them, so that they become distinct. One thing to remember is that the appropriate scale for choosing clusters can be decided based on the user's needs. Since the noise data do not belong to any of the clusters, and usually their presence causes problems for the current clustering methods, we have to get rid of them. Low-pass filters used in the wavelet transform will help us remove the noise and result in more accurate clusters. Obviously, the discovered clusters in frame-diff images, under some supposed circumstances, indicate to the moving objects we ultimately intend to find.

Parallel Search Policy

The idea of parallel wavelet clustering succeeds to the desirable properties of a good clustering technique, such as speeding up the convolution operation of high-dimensional or very large data space, being insensitive to order of input data to be processed, and assumption about the shape and numbers of clusters. During applying wavelet transform on each dimension of the data space, the required operations for each cell can carried out independent of the other cells. Thus, using parallel processing can speed up transformation the space. The connected component analysis can be speeded up using parallel processing. Each thread in search step runs separately.

Suppose there is a d-dimensional vector, which has d attributes and is considered as a point of d-dimensional space. Then all these points may constitute a d-dimensional signal. After applying wavelet transform, the high frequency parts of the signal correspond to the regions of the data space where there is a rapid change in the distribution of data which is the boundaries of clusters; the low frequency parts of the d-dimensional signal correspond to the region of the cluster itself. Removing low count value cells by applying a threshold on the count values will effectively remove majority of the outliers and help preserving the original shape of the clusters. As a result, we can pick up the dense region of discrete approximation and labeled them to clusters by transforming the signal into frequency domain. Figure 2 denotes the original 2-dimensional data space with noise before wavelet clustering, where the binary image is the difference of two sequent frames, and Figure 3 illustrates the result of 2-d wavelet clustering.

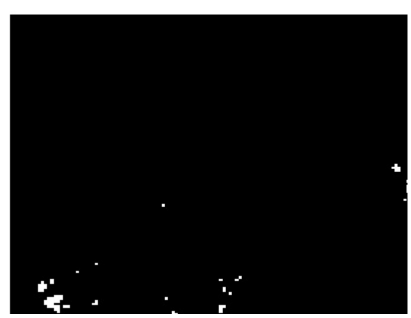

Figure 2. Original data space with noises $(N = 821)$

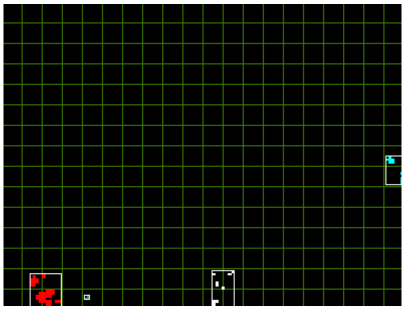

Figure 3. Result of 2-d Wavelet Clustering

$(C = 4, m = 20, \lambda t = 0.25)$

However, the performance of wavelet clustering depends on the values on m (the number of intervals in each dimension) and d (the number of dimensions in the data space) since each dimension will be divided into m intervals which totally build $k = m^d$ grid cells in feature space. When the data is input, the time complexity is only the function of k, and is independent of N (the number of samples). In other words, quantization and dimensionality of the data space are two important issues in wavelet clustering. Generally speaking, wavelet clustering is a very fast method and performs very efficiently on very large databases due to assuming $k \ll N$ in low-dimensional space. Therefore, the overall time complexity is $O(N)$. An example result of parallel search is shown as figure 4.

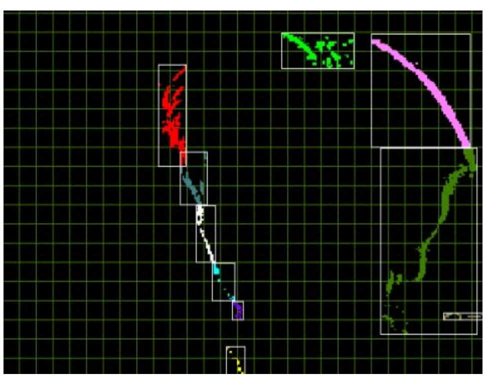

Figure 4. Parallel search the nearest $3^d - 1$ fields

$(C = 10, m = 20, \lambda t = 0.235)$

Each time we consider a new decomposition level, we can ignore some details in the average subbands. Thus, the five main steps of parallel wave clustering can be concluded as follows:

(a). Quantize the original data space; and assign data points into quantized cells

(b): Apply wavelet transform in feature space

(c): Try parallel search the connected components at different resolutions

(d): Assign labels to connected components

(e): Map the data to clusters via mapping tables

In the step (b), Applying wavelet transform on the cells may results in a new data space and hence new cells. This new cells are often produced by isolated points instead of points of clusters. The shape of clusters may be distorted due to convolution operations. In this case, applying reasonable thresholds can get rid of the isolated points (noise), and help keeping the original shape of clusters. Only significant cells in transformed space are saved and processed which has count values greater than a particular threshold τ. Parallel depth-first approach then can be used to find the connected components until all cells are visited. Morphological filtering, connectivity analysis, color and edge analysis are the most common post-processing methods.

IV. EXPERIMENTS AND APPLICATIONS

As mentioned before, some trivial issues may cause serious problems including merging of two or more objects, object shape distortions, such as unclear video data, illumination change [10], poor weather conditions and shadow disturbance. Shadow elimination is an important post-processing step, especially for outdoor image sequence because the moving shadows are easily misunderstood as foreground as they differ from the background but very like the actual moving objects that they belong to.

From the results of reparative experiment, however, our proposed approach is good to lighting changes, repetitive motion of the background, slow or fast moving objects and removal of scene objects, by adjusting the threshold of luminance and saturation. It can handle sudden illumination changes due to fast cloud movements with minor modification, waving tree branches, and spectral reflections such as reflections from car windows. We also found that the proposed moving objects detection algorithm has a high accuracy equals to 97.2% when the thresholds set to $\lambda t = 0.156$, $m = 20$ under some good testing conditions such as on expressway. Figure 5 denotes the image of initial background, in which a red car stayed. Figure 6 illustrates the result of single moving object detection - a white car was identified and captured. Multiple moving objects, a passerby and a bicycle were also tracking in figure 7.

Reliable moving object detection is an essential step in further analysis of image sequence. Tracking motion objects and their trajectories play the central role in visual event detection. We addressed these issues in order to bridge the gap between low-level primitives such as pixel-level features and high-level primitives such as moving objects, actions and events recognition. Figure 8 and Figure 9 illustrate the papilionaceous trajectories of moving car in examination system for vehicle driver's scene-driving skill.

Figure 5. Initial scene background
(Note that the red car was stopping)

Figure 6. Result of single moving object detection
(A transient white car was captured)

Figure 7. Result of moving multi-object detection
(A passerby and a bicycle were tracking)

V. CONCLUSIONS

The proposed moving objects tracking algorithm, which is based on frame difference and wavelet clustering, leads to a robust moving objects detection capability. It does not rely on models or assumptions about objects or the scene, which make it a fast and automatic method of multiply objects tracking for a variety of video surveillance systems and real-time ITS systems. Though this approach meets some desirable property of rapid moving objects detection, tracking technique, it still requires modification to better handle high dimensional datasets and concurrently enhance the efficiency. Our work is only part of a set of related research piece.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (No.60675010), 863 National High-Tech Program (No.2007AA01Z132), National Basic Research Priorities Programme (No.2007CB311004) and National Science and Technology Support Plan (No.2006BAC08B06).

REFERENCES

[1] J. Crowley. A representation for visual information. Robotic Inst. Carnegie-Mellon Univ., Tech. Rep. CMU-RI-TR-82-7. 1987

[2] Y. Ho and A. Gersho, A pyramidal image coder with contour-based interpolative vector quantization. Proceedings of the SPIE Visual Communications and Image Processing Conference IV. Vol. 1199. Nov 1989. pp. 733-740

[3] F. Sun and P. Maragos, Experiments on image compression using morpho9logical pyramids. Proceedings of the SPIE Visual Communications and Image Processing Conference IV. Vol. 1199. Nov 1989. pp. 1303-1312

[4] J.W.Han, M.Kamber, Data Ming: Concepts and Techniques. 2001: Morgan Kaufmann.

[5] Stephane Mallat, A wavelet tour of signal processing. Academic Press, 1999

[6] M. Vishwanath. The recursive pyramid algorithm for the discrete wavelet transform. IEEE Trans. on Signal Processing. Vol 42, no.3, 1994, pp.673-676

[7] Gholamhosein Sheikholeslami , Surojit Chatterjee , Aidong Zhang. WaveCluster: a wavelet-based clustering approach for spatial databases. The International Journal on Very Large Data Bases. Vol. 8 , Issue 3-4 , 2000

[8] Sheikholeslami G, Zhang A (1997) An Approach to Clustering Large Visual Databases Using Wavelet Transform. In: Proceedings of the SPIE Conference on Visual Data Exploration and Analysis IV, San Jose, Calif., pp322-333.

[9] G. Strang, Wavelets. American Scientist, Vol. 82, 1992, p.250-255

[10] Cezar, Rosito, SR Musse. A Background Subtraction Model Adapted to Illumination Changes. Image Processing, 2006 IEEE International Conference on (2006), pp. 1817-1820.

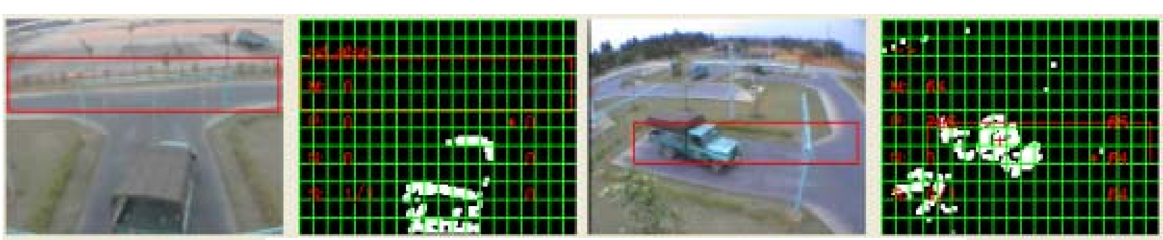

Figure 8. Moving Car Captured in examination system for vehicle driver's scene-driving skill