An Adaptive Wavelet Transform Application to Multiple Targets Tracking in the Air

Rui Wang(Member, IEEE), Jingwen Xu
Key Laboratory of Precision Opto-mechatronics Technology, Ministry of Education,School of Instrumentation Science and Opto-electronics Engineering,Beihang University
Beijing 100191,China
Email:wangr@buaa.edu.cn

Yan Zhao
6th Research Laboratory
Beijing Institute of Remote Sensing Equipment
Beijing 100191,China
Email:azz0623@sina.com

***Abstract*—Motion estimation (ME) has been extensively applied in the computer vision, including vision-based target tracking in the air. For attaining robust performance in noise and handling obscuration or the parallel moving nearby target interferences trajectories, the ME algorithm based on the spatio-temporal continuous wavelet transform (CWT) with a band pass velocity Kalman Filter in the transform domain is proposed, which is designed to conduct the efficient multiple target tracking in an adjustable spatio-temporal processing block. The CWT allowing for the definition of three energy densities over a subset of the CWT parameter space, which accompanied with Kalman Filter has been employed to capture motion information over multiple frames and proved excellent in velocity selectivity. To best handle interferences among multiple nearby targets, a simpler and robust solution called as self-adaptive rotation of coordinates and a practical functional relation between the target radius, speed and the scale parameter is developed. The presented novel joint processing technique using expectation-maximization based Gaussian mixture estimation, together with a global nearest neighborhood algorithm to perform data association， achieves maintaining kinematic trajectory of every targets either in linear or nonlinear motion. Examples with synthetic data and real data taken in air surveillance are given to demonstrate the effectiveness of these proposed strategies.**

Keywords—Motion estimation; CWT; kalman; self-adaptive coordinate; expectation-maximization

I. INTRODUCTION

Motion estimation (ME) represents an intriguing field of research in computer vision. It is of primary importance for applications like object tracking in defense scenarios, motion compensated video compression,and automatic traffic monitoring [1]. The numerous methods proposed over the years to estimate motion parameters can be divided into two categories. The first group of methods process data in spatial domain, and the other in frequency domain [2]. In the spatial domain, motion can be estimated with feature based or block-matching algorithm approaches which is notoriously sensitive to noise and is easy to cause instabilities in tracking. Another approach for ME is based on frequency domain processing of the available data, and the transform-based methods are inherently robust to such inaccuracies.The frequency warping is applied to be an enhanced motion estimation scheme in order to improve the block matching accuracy. Furthermore,the multi-resolution representation of wavelet transform [3,4] which involves a different trade off in time/space frequency is very useful for the analysis of image/video signals. So the application of the continuous wavelet transform (CWT) on video signals for motion estimation is imperative in order to integrate target signal energy against random background noise. A joint spatio-temporal multiframe information processing algorithm based on the CWT is introduced in [5]-[7] where the energy densities are derived for motion parameters estimation. Applications of this strategy will benefit from a motion estimation algorithm that is adaptive to time varying object signatures and robust to sensor noise as well as temporary occlusions.Although the definition of three energy densities by integrating over a subset of the CWT parameter space and the principle of ME with energy densities by sequentially optimizing a state vector composed of velocity, position, and size parameters are described in detail in [5] and [6], few literatures take the issues that target moving in a higher speed and the interferences of parallel moving nearby targets into consideration in energy density optimization, which could easily result in the failure of target movement extraction.

The main focus of this paper is to deal with the ME algorithm based on the spatio-temporal continuous wavelet transform (CWT) that address the aforementioned problems. It extends the local energy densities based ME [5] method with a bandpass velocity Kalman Filter to extract faster moving target accurately in multiframes. Particularly, a self-adaptive rotation of coordinates is created in the EM based Gaussian density estimation processing to effectively resolve the measurement pairing uncertainty when estimating the motion of multiple objects. Inspired by our previous work in [8], we construct adjustable spatio-temporal processing blocks to conduct the efficiency multiple target tracking. Moreover, a practicalfunctional relation is established to reasonably obtain the initial value of scale parameter and to help the further camera zoom operation. The proposed CWT-based algorithm can estimate both linear and nonlinear motion successfully in the simulations with synthetical data and real data taken in air surveillance.

The structure of this paper is as follows: Section II introduces the mathematical model of the spatio-temporal filtering method, reviewing the joint model properties of the energy-based moving target representation. In Section III the improved CWT-base algorithm for multi-target tracking is introduced and the factors that influence the performance of the proposed algorithm are analyzed. The experimental results both in the simulation and in the real video data are shown in Section IV. Conclusions are given in Section V.

II. CONSTRUCT A FILTER FOR TARGET MOTION ESTIMATION

Inspired by the motion tracking technology illustrated with CWT [5]-[9] which rely on signal transformations that model motion and object deformations, we provide a filter to estimate motion parameters of targets as a motion-selective sub-band decomposition for video signals. Here, the principle of each module for the proposed filter is elaborated.

A. Principle of Spatio-Temporal Filter

For a time-varying signal $s(x,y,t)$, i.e. the object (2+1)D luminance signal is a function of spatial variables x and y and time variable t and in the case of a linear motion with a constant velocity (v_x,v_y) its motion can be characterized as:

$$s(x,y,t)=s(x-v_xt,y-v_yt,0) \tag{1}$$

The spatio-temporal CWT of $s(x,y,t)$ is defined as an inner product between the signal of interest $s(\vec{x},t)=s(x,y,t)$ and the wavelet basis $\psi_{\vec{g}}$ parameterized by vector $\vec{g}$, that is:

$$S_\psi(\vec{g})==\frac{1}{\sqrt{c_\psi}}\iint \psi_{\vec{g}}^*(\vec{x},t)s(\vec{x},t)d^2\vec{x}dt \tag{2}$$

The vector $\vec{g}$, directly associated with motion characteristics, is defined as $\vec{g}=\{a,c,\theta,\vec{b},\tau\}$ where a is the spatio-temporal dilation, c and θ which can be represented by $c=\sqrt{v_x^2+v_y^2}$, $\theta=arc\tan(v_y/v_x)$ reach the velocity; the spatio-temporal is given by $\vec{b}$ and τ. The asterisk * represents the complex conjugate operator, and the constant c_ψ is associated with the admissibility condition and depends on the wavelet family. Alternatively, the spatio-temporal CWT can be expressed in the wave number-frequency domain as:

$$S_\psi(\vec{g})==\frac{1}{\sqrt{c_\psi}}\iint \psi_{\vec{g}}^*(\vec{k},\omega)s(\vec{k},\omega)d^2\vec{k}d\omega \tag{3}$$

As it is suggested in [6] a Morlet wavelet is chosen as mother wavelet ψ because of its separability of transformations and its tails taper off smoothly, which ensures a balanced tradeoff between frequency and time resolution. The Morlet wavelet can be defined in the spatio-temporal domain and the wavenumber-frequency domain respectively, by:

$$\psi(\vec{x},t)=(e^{i\vec{k}_0\cdot\vec{x}}e^{-\frac{1}{2}|\vec{x}|^2}-e^{-\frac{1}{2}|\vec{x}|^2}e^{-\frac{1}{2}|\vec{k}_0|^2})\times(e^{i\omega_0t}e^{-\frac{1}{2}t^2}-e^{-\frac{1}{2}t^2}e^{-\frac{1}{2}\omega_0^2}) \tag{4}$$

$$\hat{\psi}(\vec{k},\omega)=(e^{-\frac{1}{2}|\vec{k}-\vec{k}_0|^2}-e^{-\frac{1}{2}\left(|\vec{k}|^2+|\vec{k}_0|^2\right)})(e^{-\frac{1}{2}(\omega-\omega_0)^2}-e^{-\frac{1}{2}(\omega^2+\omega_0^2)}) \tag{5}$$

The construction of the above-mentioned filter given by (4) and (5) is known as CWT used to map the input signal space to a physically meaningful parameter space. It can also be seen as a tool for motion based filtering, which has an excellent orientation and scale selectivity, thus will be applied to facilitate the ME algorithm development.To extract motion estimates we define a set of local energy densities.

In cases where the target undergoes the nonlinear motion, the motion process can be partitioned into small segments and the target motion within each segment will be assumed linear. Then the aforementioned model could be extended to the case of nonlinear motion by windowing the motion in time in a video.

B. Energy-Based Representation for the CWT

The CWT described above is perfectly reconstructing and preserving the energy provided by the mother wavelet ψ and a CWT energy density which can be regarded as evaluation function of the parameter vector $\vec{g}$ is defined as (6).

$$\varepsilon(\vec{g})\equiv\left|S(a,c,\theta,\vec{b},\tau)\right|^2\equiv\left|S_\psi(\vec{g})\right|^2 \tag{6}$$

1.Velocity energy density: By fixing the scale to be the one evaluated in previous frame and partial integration on position parameters we can get the velocity (speed-orientation) energy density, which can be interpreted as an estimator of *local* velocity:

$$\varepsilon^1_{a_0,\tau_0}(c,\theta)=\iint_{\vec{b}}|\left\langle\psi_{(a_0,c,\theta,\vec{b},\tau_0)}\,|\,s\right\rangle|^2d^2\vec{b} \tag{7}$$

where τ_0 represents the temporal translation in current frame.

2.Spatial energy density:By fixing the scale and the velocity vector (including of speed and orientation angle) to be the previous ones, we can get the spatial energy density:

$$\varepsilon^2_{a_0,\tau_0,c_0,\theta_0}(\vec{b})=\frac{1}{(a_0)^4}|\left\langle\psi_{(a_0,c_0,\theta_0,\vec{b},\tau_0)}\,|\,s\right\rangle|^2 \tag{8}$$

3.Scale energy density: By fixing the velocity vector (including of speed and orientation angle) to be the previous ones and operating partial integration on position parameters, we can get the scale energy density:

$$\varepsilon^3_{c_0,\theta_0,\tau_0}(a_0)=\frac{1}{a_0^4}\iint_{\vec{b}}\left|\psi_{(a_0,c_0,\theta_0,\vec{b},\tau_0)}\left|s\right|\right|^2d^2\vec{b} \tag{9}$$

Then locally optimization operations can be done to the above three energy densities to get the motion parameters. We can also say that the CWT based filter is derived into three filters, separately called velocity filter, position filter and scale filter. Thus our CWT-based ME algorithm is constructed completely via the defined three energy densities and consisted of a *frame-by-frame* optimization of the motion parameters associated with the tracking target in the image sequence.

III. MOTION ESTIMATION STRATEGY

With the introduction in section II, a spatio-temporal filter based target tracking algorithm can be developed. Our ME algorithm,which can determine target coordinates frame-by-frame, can be viewed as a target tracking algorithm after the initial detection, such as optical flow operation, has been performed. A flow diagram of the state updating process for either one target or multi-target performed by the CWT-based ME algorithm is depicted in Fig. 1.

A Nelder-Mead[5][14]simplex search algorithm was employed for energy density optimization, which appeared to work well in the situation that the targets are far apart. Unfortunately, it can't handle the energy densities of multiple close-by or crossing targets. Note that if $\left|\vec{k}_0\right|$ and ω_0 are large enough, (4) and (5) approximate the expression of a modulated Gaussian filter in spatio-temporal and wavenumber-frequency domains respectively. Motivated by the mathematical convenience of Gaussian functions and inspired by the tracking strategy in [6], we adopted an EM based Gaussian mixture estimation [10][11] as a joint multiple density processing technique in order to deal with target interferences.The proposed approach, which handles the energy densities via Gaussian mixture estimation and uses EM to calculate the parameters, is robust to such errors, because of the simultaneous processing of the adjustable block formulated with several sequential frames and the novel technology of self-adaptive rotation of coordinates. At last, one of the data association algorithm [12] is designed to handle target interference which causes track bias and realize a high performance multi-target tracking algorithm.The improved strategy and the implementation based on the existing CWT tracking algorithm are presented here, which is elaborated in following.

A. Adjustable Spatio-Temporal Processing Block

Often there is the case that ME task is not tracking all of the moving targets in the video but only several of them. If all pixels are processed, the added uninteresting targets will bring not only the interferences to the tracking but also the computation cost. To solve this problem we propose an adjustable processing window strategy which puts the close-by targets in the same window, and only the data of pixels within the windows are processed. These windows guarantee that the square-integrable signals over time and space belong to processing blocks symbolically represented by s_{w_p}, where p is the index of processing blocks. Thus the time-varying motion parameter values are obtained by processing the blocks once.The detail is described in the following way.

First the locations of the tracking targets in current frame can be roughly evaluated according to their locations and velocities in the previous frame, and one step further their location coordinate distances dx and dy. These target relationships can be defined as either "directly related", which would met condition represented by (10), or "indirectly related" respectively.

$$dx_{ij} \le d_0 \ \ and \ \ dy_{ij} \le d_0 \tag{10}$$

where d_0 is determined artificially according to the speed of targets and the temporal length of blocks.All of the directly related and indirectly related targets are collected together into one block if they are included in its spatial boundaries.

B. Estimation of velocity

Being different from the Nelder-Mead maximum likelihood (ML) searching algorithm, the EM algorithm is a stochastic approximation procedure for ML. The energy density values of all the sampling points shall be firstly calculated as independent observations, each with underlying probability density function (PDF). Thus the high computational cost will associate with EM-based CWT for obtaining the motion parameter ML estimates, especially for that of the velocity. To improve the computational efficiency of the EM-based CWT, a Kalman filter is adopted here to estimate the target velocity, while other parameters of targets are estimated by EM-based CWT in equation (8) and (9). The Morlet mother wavelet of (4) and (5) with $\vec{k}_0 = [-2, 0]$ and $\omega_0 = 2$ are used in equation(9) and (10). Usually, Kalman filters are characterized by two equations: a state equation and an observation equation. The state equation is an adaptive predictor that updates the $U(k)$ of the filter $U(k) = A(k-1,k)U(k-1) + W(k)$,where $U(k) = (x_k, y_k, v_{x_k}, v_{y_k})$ is the state prediction at frame k and $W(k)$ is the prediction error. $A(k-1,k)$ is the transition matrix. For the sake of combining a more reasonable kinematic model with the tracking ability for a moving target in a higher speed, the velocity in the state vector of the Kalman filter is adopted as the final parameters, and we still derive final location measurements from target energy densities with a Gaussian mixture.

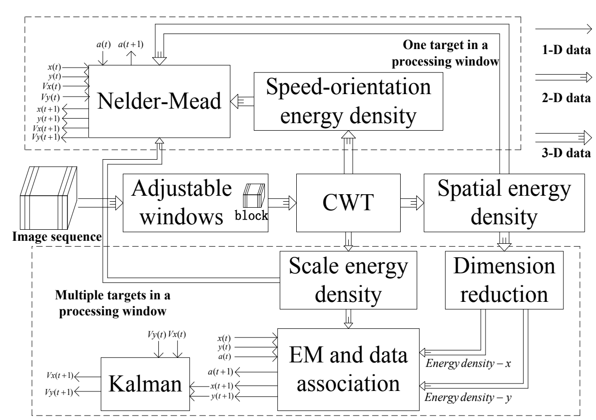

Fig. 1. Flow diagram of tracking algorithm

C. Self-adaptive Rotation of Coordinates

Though an EM algorithm can be employed directly for 2D density function estimation, the computational complexity is much higher than 1D density function estimation. Therefore, a novel extraction operation with the self-adaptive rotation of coordinates is presented for the 2D spatial energy density, $\varepsilon^2_{a_0,\tau_0,c_0,\theta_0}(\vec{b})$, and a similar operation can be applied to 1D scale energy density, $\varepsilon^3_{c_0,\theta_0,\tau_0}(a)$.

To reduce the dimension of spatial energy density, it is a common way to simply decompose it by projection to x axis and to y axis respectively in the coordinates $x-o-y$ shown in Fig. 2. However, massive experiments show a large error between the mean value obtained by EM Gaussian mixture estimation and the real location of the target when some one-dimensional coordinates of multiple targets are close to each other. We address the problem by turning the original image coordinates system anticlockwise around the origin o with an angle θ $(0 \le \theta \le \pi/2)$ to form a new coordinate system $x'-o-y'$, so as to increase the coordinate differences between the targets.

For example, as shown in Fig. 2, the coordinate difference between target 3 and target 4 in the original coordinates $x-o-y$ is small, but their coordinate differences are enlarged in the new coordinates $x'-o-y'$. If there are n_p targets whose locations are $(x_i(k), y_i(k))$, $(i=1,2,...,n_p)$ in the processing window w_p in $x-o-y$, then its correspond location in the coordinates $x'-o-y'$ can be calculated by:

$$\begin{bmatrix} x'_i(k) \\ y'_i(k) \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} x_i(k) \\ y_i(k) \end{bmatrix} \tag{11}$$

Furthermore, the location coordinate differences between the targets can be obtained by:

$$\begin{cases} dx'_{ij}(k) = \left| x'_i(k) - x'_j(k) \right| \\ dy'_{ij}(k) = \left| y'_i(k) - y'_j(k) \right| \end{cases}, (i, j = 1,2,...,n_p, i \ne j) \tag{12}$$

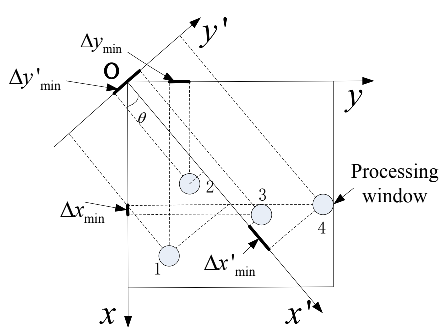

Fig. 2. Schematic diagram of coordinate system changing

Obviously, the optimal coordinate rotating angle θ should be the value that can make all of the coordinate differences as large as possible, and the majorization can be realized by the proposed self-adaptive coordinate rotation traversal method illustrated by equation (13). Consequently, the location energy density ε^2 is reducing dimensions into $\varepsilon^2_{x_i'}$ and $\varepsilon^2_{y_j'}$ shown in formular(14).

$$\theta = \max_{\theta_h} \left\{ \min\left\{ dx'_{ij}(k), dy'_{ij}(k) \right\} \right\}, \theta_h = 0, \Delta\theta, ...\pi/2 \tag{13}$$

$$\begin{cases} \varepsilon^2_{x_i'} = \sum\limits_{(x_i', y_j') \in w_p} \varepsilon^2(x_i', y_j') \bullet \Delta y_j', i = 1,2,...,n_x \\ \varepsilon^2_{y_j'} = \sum\limits_{(x_i', y_j') \in w_p} \varepsilon^2(x_i', y_j') \bullet \Delta x_i', j = 1,2,...,n_y \end{cases} \tag{14}$$

The Gaussian Mixture Model (GMM) [12]with the known number of components is chosen to model real data of the reduced one-dimensional spatial energy densities. Extracting the parameters of each single Gauss is a problem of curve fitting. The GMM shall be normalized before fitted:

$$h(x_i) = \frac{\varepsilon^2_{x_i'}}{\sum\limits_i \varepsilon^2_{x_i'} \bullet \Delta x_i'}, h(y_j) = \frac{\varepsilon^2_{y_j'}}{\sum\limits_j \varepsilon^2_{y_j'} \bullet \Delta y_j'} \tag{15}$$

We successfully translate the fitting issue into a density estimation problem, and thus the powerful iteration algorithm EM under the Maximum Likelihood frame is employed.

D. Functional Relationship of Speed, Scale and Radius

Assuming the motion parameter of the tracked target in frame $k-1$ of s_{w_p} is given by: $g_i^{k-1} = (x_i^{k-1}, y_i^{k-1}, c_i^{k-1}, \theta_i^{k-1}, a_i^{k-1}, k-1)$, where a_i^{k-1} represents the kinematic scale of the targets relative to the block and $(x_i^{k-1}, y_i^{k-1}, c_i^{k-1})$ in pixels. Particularly, to determine a_i^1 more reasonably, we analysed the relation between the target radius r, the speed c and the kinematic scale a, and then established the polynomial fitting for scale a and radius r respectively. The cubic polynomial whose fitting parameters got from a great number of experimental data with variation range of the target radius r from 1 to 12 pixels and that of the velocity magnitude c from 1 to10 pixel/frame is finally obtained in an empirical manner as the (16) and (17). Formula (16) provides a reasonable initial value a_i^1 to the benefit of the more accurate ME processing, while (17) could provide a control parameter r for automatic zoom of the camera or visual guidance based on the target tracking.

$$\begin{aligned} a = &-0.0029r^3 + 0.0042c^3 + 0.0069r^2c - 0.0018rc^2 + 0.0435r^2 \\ &-0.0188c^2 - 0.1390rc + 0.5702r + 0.3846c + 0.1526 \end{aligned} \tag{16}$$

$$\begin{aligned} r = &0.0539a^3 - 0.0082c^3 - 0.2153a^2c + 0.0656ac^2 - 0.1332a^2 \\ &-0.2674c^2 + 1.5504ac - 0.0943a - 1.5498c + 2.3668 \end{aligned} \tag{17}$$

IV. Experimental Results

In order to test the proposed approach, we perform a number of experiments with synthetic and real videos. The purpose of using synthetic videos is to control the precise parameters of the target displacements, and thus to get ground truth available, for the verification of the location estimation results.

A. Simulation Experiments

TEST1: The video includes a continuously varied gray grade background and three moving targets, whose trajectories are formulated by (18), (19) and (20) respectively, moving with the corresponding linear, accelerated rectilinear and circular. The targets, whose radiuses are given in pixels: $r_1 = 3$, $r_2 = 3$ and $r_3 = 5$, are of a circular shape which means the intensity value of the pixel on the target is the identical.

$$\begin{cases} x_1(k) = 200 + 2\sqrt{2}(k-1) \\ y_1(k) = 205 + 2\sqrt{2}(k-1) \end{cases} \tag{18}$$

$$\begin{cases} x_2(k) = 360 + (-1.41)(k-1) - 0.5\times 0.1(k-1)^2 \\ y_2(k) = 240 + (-1.41)(k-1) + 0.5\times 0.1(k-1)^2 \end{cases} \tag{19}$$

$$\begin{cases} x_3(k) = 370 + 100\cos(3.725 - 0.021(k-1)) \\ y_3(k) = 295 + 100\sin(3.725 - 0.021(k-1)) \end{cases} \tag{20}$$

We compare our tracking results (method 1) with the Nelder-Mead searching algorithm based CWT approach in [5] (method 2) and the EM-based CWT strategy in [6] (method 3) in TEST 1. Part of estimated location values are listed in Table I . The Space-time representations of the trajectories using three different methods are shown in Fig. 3-1, to 3-3.

The results show that all of the three methods can provide a good indication of the small target motion, without interference from the background luminance variation.As shown in the figures, at and after the three crossing point, the ME algorithm proposed in this article can successfully stay in track, however, the tracking trajectories are mistakenly merged or disorientated by the other two methods. What's more, as the distance between targets from far to near or from near to far there is no false tracking which reflects from the side that the strategy of processing window selection is appropriate.

B. Realistic Video Experiments

TEST2: The video is photographed on the roof about 60m high. Five moving cars are our tracking targets in the video with 720×576 pixels^2 of each frame. The initial speeds of them are between 0.8 pixels per frame and 2.5 pixels per frame. Fig. 4-1 to 4-3 refer to the tracking results of 14th, 110th and 149th frame respectively. The centroid distance, which is about 10 pixels, of the two targets tracked by the blue and green window is the minimum of all target distances throughout the entire process, and the radiuses of the two targets are all about 3 pixels.

This result shows that our tracking strategy is not interfered by the static background and has strong anti-shielding capability and adaptation to small targets. In addition, the tracking result of the closely spaced three cars in the upper part of the image shows that the problem of target mutual interferences can be solved elegantly.

TABLE II
POSITION DATA FOR DIFFERENT TARGETS IN A PROCESSING WINDOW

Methods of Location estimation	Locations of The Targets	Frame Number						
		37	38	39	40	41	42	43
Real Locations	(x_1, y_1)	(301.8,306.8)	(304.7,309.7)	(307.5,312.5)	(310.3,315.3)	(313.1,318.1)	(316.0,321.0)	(318.8,323.8)
Method1	(x_1, y_1)	(304.6,309.1)	(307.7,312.7)	(309.3,315.0)	(312.0,318.1)	(314.7,319.6)	(317.3,322.2)	(319.7,324.6)
Method2	(x_1, y_1)	(273.5,315.8)	(273.5,317.0)	(273.6,318.4)	(273.9,319.9)	(274.2,321.5)	(274.8,323.3)	(275.4,325.4)
Method3	(x_1, y_1)	(302.9,306.5)	(306.2,309.4)	(309.1,312.0)	(311.9,316.8)	(314.7,319.6)	(317.3,322.2)	(319.8,324.7)
Real Locations	(x_2, y_2)	(269.7, 330.3)	(265.4,334.6)	(260.9,339.1)	(256.4,343.6)	(251.7,348.3)	(247.0,353.0)	(242.1,357.9)
Method1	(x_2, y_2)	(271.2,334.0)	(265.9,338.3)	(260.2,341.2)	(255.6,345.2)	(250.2,350.8)	(247.4,355.2)	(242.6,359.9)
Method2	(x_2, y_2)	(274.1,318.4)	(270.9,319.3)	(267.7,320.1)	(264.6,321.0)	(261.4,321.8)	(258.2,322.7)	(255.0,323.5)
Method3	(x_2, y_2)	(273.9,330.2)	(273.0,335.0)	(273.0,340.5)	(266.5,345.7)	(274.1,349.9)	(274.9,354.6)	(275.7,359.3)
Real Locations	(x_3, y_3)	(271.5, 312.2)	(271.9,314.3)	(272.3,316.3)	(272.8,318.4)	(273.3,320.4)	(273.8,322.4)	(274.4,324.4)
Method1	(x_3, y_3)	(273.8,316.4)	(273.4,319.4)	(273.2,322.3)	(272.7,321.8)	(271.2,322.2)	(274.8,323.8)	(275.3,326.0)
Method2	(x_3, y_3)	(273.5,315.8)	(273.5,317.0)	(273.6,318.4)	(273.9,319.9)	(274.2,321.5)	(274.8,323.3)	(275.4,325.4)
Method3	(x_3, y_3)	(268.9,314.4)	(265.2,317.9)	(260.0,320.5)	(263.6,319.2)	(252.0,321.0)	(247.1,322.8)	(242.7,325.2)

Method 1: The Method Detailed in This Article
Method 2: Use Nelder-Mead to Extract Parameters of Targets From Energy Densities
Method 3: Use EM to Extract Parameters of Targets From Energy Densities But without Coordinate Rotation

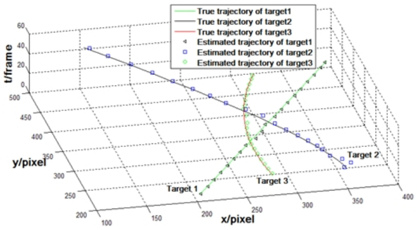

Fig. 3-1 Trajectories Using Proposed Method

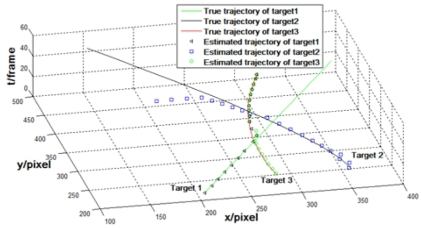

Fig. 3-2 Trajectories Using Nelder-Mead-Based CWT Method

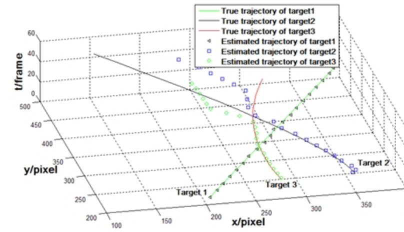

Fig. 3-3Trajectories Using EM-Based CWT Strategy Without Coordinate Rotation

Fig. 4-1 Frame 14

Fig. 4-2 Frame 110

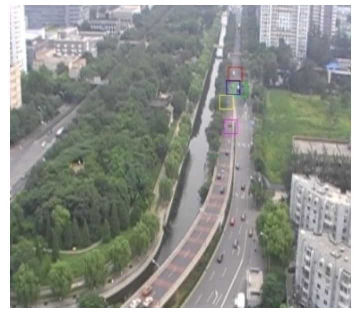

Fig. 4-3 Frame 149

V. CONCLUSION

We have presented a novel strategy for the extraction of motion parameters from video sequences using the CWT. The proposed approach has the advantages of being robust to local spatio-temporal illumination variations, local measurement noise and object occlusions, by processing sequential multi frames in a video simultaneously. Instead of three energy density functions, two of them are approximated by Gaussian mixtures and estimated by a joint EM estimation with self-adaptive rotation coordinate technique, meanwhile Kalman filter as a tool to analyze the velocity of the spatio-temporal signals. All of the ME processing is carried through in an adjustable spatio-temporal processing block. This makes it more computational efficient than the pre-existing CWT methods in order to provide robust results in such cases. Experiments with both synthetic and real video sequences lead to more accurate location estimation, even in the presence of local occlusion.

ACKNOWLEDGMENT

The research leading to these results has received funding from the National Natural Science Foundation of China (Grant No. 60974108 and Grant No. 61673039). In addition, thanks are due to Yuming Liu for his assistance with the experiments.

REFERENCES

[1] T. S. Huang, Ed., *Image Sequence Analysis*, CA: Springer—Verlag, 1981.

[2] A. Briassouli, D. Matsiki and I. Kompatsiaris, “Continuous wavelet transform for time-varying motion extraction,” *IET Image Process*, Vol. 4, no. 4 pp. 271—282, 2010.

[3] S. Mallat, “Multi frequency channel decompositions of images and wavelet models,” *Speech Signal Processing*, vol. 7, pp. 2091—2110, December 1989.

[4] Luersen, Ma, Le Riche and R, “Golbalized Nelder-Mead method for engineering optimization,” *Computers & Structures*, vol. 82, no. 23—26, pp. 2251-2260, September 2004.

[5] Mujica F.A., Leduc J.P., Murenzi R. and ET AL, “A new motion parameter estimation algorithm based on the continuous wavelet transform,” *IEEE Transaction on Image Processing*, vol. 9, no. 5, pp. 873—888, 2000.

[6] L. Hong, Y. Ruan, W. Li, D. Wicker, and J. Layne, “Energy-based video tracking using joint target density processing with an application to unmanned aerial vehicle surveillance,” IET Computer Vision, vol. 2, no. 1, pp. 1—12, March 2008.

[7] Wang Rui，Zhao Yan，Liu Yuming and Yuan Yan. “A target tracking algorithm based on the continuous wavelet transform of spatio-temporal processing block,” *International Conference on Electronics and Optoelectronics*, vol. 2, pp. 95—98，July 2011.

[8] Wang Rui，Zhao Yan，Tang Yujun and Yuan Yan. “A spatio-temporal filtering method for motion estimation,” *International Conference on Computer Science & Education*，vol. 8, pp. 830—834，2011.

[9] Patrice Bra Lilt. “A new scheme for object-oriented video compression and scene analysis based on motion tuned spatio-temporal wavelet family and trajectory identification,” *Signal Processing and Information Technology*, pp. 250—254, 2003.

[10] Laird N.M. Dempster A.P., Rubin D.B., “Maximum likelihood from incomplete data via the EM algorithm,” J. R. Soc. Stat. B, vol. 39, no. 1, pp. 138, 1977.

[11] Vlassis, N, Likas, A, “A greedy EM algorithm for Gaussian mixture learning,” Neural Processing Letters, vol. 15, no. 1, pp. 77—87.

[12] Pavlina Konstantinova, Alexander Udvarev,“A Study of a Target Tracking Algorithm Using Global Nearest Neighbor Approach,” *International Conference on Computer Systems and Technologies*, 2003.

[13] Pavlina Konstantinova, Alexander Udvarev, Tzvetan Semerdjiev, “A Study of a Target Tracking Algorithm Using Global Nearest Neighbor Approach,” International Conference on Computer Systems and Technologies, 2003.