Neurally Inspired Object Tracking System

Shu-Fai Wong and Kwan-Yee Kenneth Wong
Department of Computer Science and Information Systems
The University of Hong Kong
Email: sfwong, kykwong@csis.hku.hk

***Abstract*— Object tracking is useful in applications like computer-aided medical diagnosis, video editing, visual surveillance etc. Commonly used approaches usually involve the use of filter (e.g. Kalman filter) to predict the location of the object in next image frame. Such approaches actually borrow ideas from signal theory and are limited to applications where dynamic model is known. In this paper, a flexible and reliable estimation algorithm using wavelet network (or wavenet) is proposed to build an object tracking system. This system simulates the perception of motion that occurs in primates. Neural-based filters will be used for color, shape and motion analysis. Experimental results show that object can be tracked accurately without fixing any dynamic model compare with commonly used Kalman filter.**

I. Introduction

Object tracking has been an active research topic in the field of computer vision. Successful tracking may lead to faster and better development of video compression techniques, medical diagnostic scheme, human-computer interaction, movie production and visual surveillance. Comprehensive surveys on human tracking can be found in [1], [2].

In general, there are two main approaches for tracking, namely the feature-based (e.g. points and contours) approach [3], [4], [5] and the model-based approach [6], [7], [8]. Both approaches are, however, time consuming if the whole image is to be analyzed due to the need of finding correspondence features among images. To shorten the time for such correspondence analysis, window searching may help. By restricting the search region, the search time can be reduced. In order to have a small search region, the prediction of the location of the object in next time frame has to be accurate.

The tracking process can be broken down into three stages, namely prediction, observation, and adjustment. In the prediction stage, the location of the target is predicted with reference to previous observations and the searching window is shifted accordingly. In the observation stage, the target is located in the searching window and its location is recorded. In the adjustment stage, the prediction error is used to adjust the prediction parameters so as to minimize the error.

In computer vision society, Kalman filtering [9] and its variations have been used extensively in recent research (e.g. [10], [11], [12]). Kalman filtering follows the prediction-correction procedure. By encapsulating the motion of the object into internal states, Kalman filtering aims at finding appropriate states that gives best-fit observations. Dynamic equation and measurement equation will be used in Kalman filter for representing the change in internal states and conversion from internal state to observation respectively. Although Kalman filter is fast, it suffers from a few and yet serious problems. For instance, too much prior knowledge such as known dynamic model and predefined noise model (uni-modal Gaussian distribution is usually assumed) is required and clutter background is not allowed.

Recently, biological based tracking have been introduced to solve the the problem. However, the use of neural network is limited to recognizing the object only (e.g.[13]) or performing simple trajectory fitting (e.g.[14]). As most researchers noticed, neural network is a kind of biological inspired techniques for pattern recognition. Thus, when it was first used in tracking problem, it was used as a kind of pattern recognizer and used for locating the target. As stated in second paragraph, the bottleneck of tracking problem is how to limit the searching size such that the whole process can be performed as fast as human does. Therefore, researchers started using multi-layer neural network to predict the trend of the trajectory of the object. This method is simply performing regression analysis on the trajectory. Although no dynamic model is needed, this method is susceptible to overfitting which makes prediction inaccurate.

In this paper, we propose a biologically inspired system for object tracking and describe how does it simulate human vision in order to solve the problem. An hierarchical and recurrent architecture, which simulate the visual pathway in primates, is used in this system. In the system, color and motion filters are used for object detection, and a wavelet network filter is used to smooth and learn the trajectory of the moving object. Experimental results show that the estimator can track moving object accurately compare with other previously used approaches.

II. Theoretical Background

A. Anatomy of the motion pathway

The visual system of primates forms a hierarchical and recurrent architecture. Starting from low-level primary visual cortex (V1), the neurons pass signal toward several high-level processing areas that are responsible for color, form and motion analysis. Once analysis result is obtained, the feedback signal is then transfered back to primary visual cortex.

At high level processing area, there are faculties for different kinds of analysis. Area V3, V4 and V5 (MT) are responsible for color, form and motion respectively. In area V3, the visual information related to color will be analyzed. Grouping of regions with similar color is done in V3 eventually. In V4, the visual information related to contour and form will be analyzed. Layering of contour at similar level is done in V4

eventually. In V5 (MT), motion information will be collected. High-level motion pattern will be generated in V5.

Primates handle tracking problem via a motion pathway. This pathway is a subset of the visual pathway described above. Within area V1, there are neurons selectively responsive to direction of motion [15]. Such neurons are often described as "tuned" for direction. This means that they will give response only when a contour moves through their receptive fields in a particular direction. In engineering terms, such neurons are acting as filters that register the presence of motion in certain direction within small receptive field [16].

The next station in motion pathway is area V5 or middle temporal visual area (MT). V5 receives some of the input directly from V1 while the remaining from V1, via V2 and V3, indirectly. The output at V1 will be integrated in V5 together with some color information provided by V3. Within V5, neurons are selective for the direction and speed of stimulus motion [17]. In addition, neurons here have much larger receptive fields than those in V1. Neurons in V5 will further project to higher visual areas which encode more complex forms and patterns of motion.

In addition to information propagation in forward direction, information is also recurrent back to primary visual cortex (at least from area V5 to V1 [18]). In humans, it is demonstrated that back propagation does induce conscious awareness of visual motion. This means conscious analysis on motion perceived can then be done. High-level thinking such as motion prediction can be performed afterward. In addition, several high-level visual areas recurrent back information toward primary visual area. This means integration of outputs from high-level visual processing might be probably taken place at low-level visual area under conscious awareness. Figure 1 shows the summary of the motion pathway.

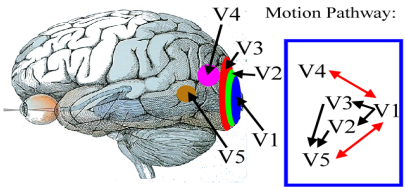

Fig. 1. *The vision-related areas in brain and the motion pathway.*

After reviewing the visual pathway, especially the motion pathway, we can observe that primitive form of motion is collected at V1. This primitive information is then transfered to high-level visual area together with other visual information such as color. Complex motion pattern will then be generated and be recurrent back to V1 for conscious analysis.

B. Wavelet model for signal analysis

As mentioned in previous sub-section, neurons in area V1 are acting as filters for registration of motion direction. Similarly, neurons in area V5 can be also considered as filters for both registration motion direction and speed in column structure. If time domain is added, each neuron is actually responsible for certain type of motion together with a timestamp. Such motion neuron that with motion trend and a timestamp is indeed similar to a wavelet that with defined shape and transition. The idea is illustrated in figure 2(a). When considering spatial-temporal dimension, we can imagine that the responses of the neural filters contribute to the trajectory of the moving object along the time dimension. The idea is also similar to the filtering principle using wavelet analysis. In wavelet theory, signal (or trajectory in spatial-temporal domain under this case) can be broken down into constituent wavelets. As we can observe, each neuron in neural pathway is of similar responsibility as wavelet. In this sub-section, wavelet theory and its application in denoising will be explored. The basic ideas illustrated will then be used in the implementation of the neural network based tracking system.

Wavelet theory was originally developed for signal analysis because it can break down signal into finer components [19]. By decomposing the input trajectory (the input signal) into the constituent wavelets, the major wavelets can be identified. Recombination of these major wavelets forms the smoothened trajectory. The minor wavelets that represent the noise are removed. Combining the constituent wavelets close to the current time frame and calculating the value of superposition will give the estimated location in next time frame. The idea is illustrated by figure 2(b).

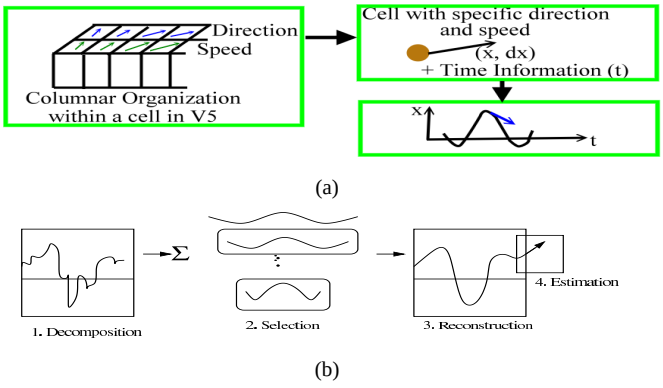

Fig. 2. *In (a), it shows the similarity between the motion information stored in V5 and those stored in wavelet. In (b), it shows the mechanism of wavelet denoising and prediction.*

Wavelet decomposition can be done through filtering. Assuming that we have transition and scale as parameters of the wavelets. For every transition and scale, we have to calculate the corresponding wavelet coefficient. Afterward, the coefficients have to be ordered. Only several wavelets will be selected according to the value of coefficient. Equation (1) and (2) illustrate how an input signal can be represented by superposition of wavelets:

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}}\phi(2^{-j}t-k), \psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t-k) \quad (1)$$

$$f(t) = \sum_{k\in Z} a_k^J \phi_{J,k}(t) + \sum_{j\le J}\sum_{k\in Z} d_k^j \psi_{j,k}(t) \quad (2)$$

where f(t) is the input signal at time t, ϕ is the scaling function, ψ is the mother wavelet, 2^{-j} and k represent the scaling

and transition factor, a_k^J and d_k^j are the scaling and wavelet coefficients respectively.

As described previously, the neurons in area V5 have similar properties as wavelet under the context of filtering. It is possible to integrate the wavelet and neural network together to perform motion filtering. The details will be explained in next section.

III. Architecture of the System

To solve the object tracking problem, we proposed a neural network based tracking system. The proposed system consists of 3 basic components: wavenet estimator, likelihood estimator and active contour fitter. Each component is actually designed to simulate the visual pathway of primate and is supposed to work as good as primate do. The logic flow of the system is described in figure 3.

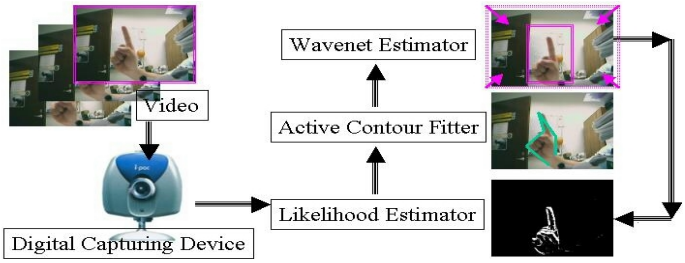

Fig. 3. *The initial searching window is of the size of the whole image. The moving object with skin color is treated as the target. It can be detected by the likelihood estimator within the searching window. Within the region reported above, the object (e.g. hand) can be located by fitting an active contour. Using the active contour fitter, the exact location of the feature points or the object is obtained. This observation is used to refine the parameters of the wavenet estimator, which will resize and shift the searching window accordingly. The whole process then repeats to track the object in the next frame.*

A. Likelihood Estimator

It searches for the most probable region of interest based on color model and optical field. In the proposed system, the likelihood estimator is tuned to report certain range of color (e.g. skin color in human body detection context). Thus, it works like a color detector. The main difference between them is that likelihood estimator does not only report skin color, but also report the area with change in intensity. Such change in intensity is supposed to be introduced from movement of the target. Using the likelihood estimator, the potential moving and proper-colored object will be reported. Similar approach can be found in [20].

The working principle stated above in fact simulates the process done within the input pathway toward area V5 in primate. With reference to figure 1, area V5 receives input from both direction-sensitive neurons in area V1 and color-sensitive neurons in area V3. In the system, color detector simulates the neuron in area V3 which filter out appropriate color, while optical field sensor simulates the motion neurons in area V1 which filter out region with intensity change. Thus, likelihood estimator integrates both information together, and then figure out where is the target and where will it go.

The change in intensity is detected by the reference white adjustment and the background subtraction. Due to the variation of lighting, especially under fluorescent light, reference white adjustment have to be performed to reduce this variation. The change in intensity detected between images may be due to the motion of the object. This region is selected as candidate for further investigation.

The color detector extracts the region with color close to the target. For instance, if human body have to be tracked, the skin color region will be extracted based on the skin color model. This model had been used in face detection, e.g. [21].

The region selected by the estimator represents the region with highest probability of moving and proper-colored region, e.g. moving hand. The resultant images are shown in figure 4.

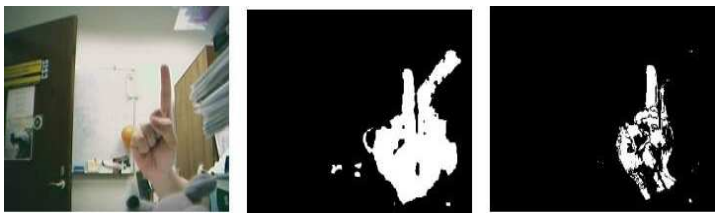

Fig. 4. *The left image is the input image. The middle one is the output of using color detector only. The right image shows the output by using both color and moving region detector.*

B. Active Contour Fitter

It locates the object by model fitting. The other two components can only approximate the location of the object without getting its exact location: by using wavenet estimator, we can approximate the location of the searching window; by using the likelihood estimator, we can approximate the moving skin color region within the window. In contrast, active contour fitter can give better estimation of the location of the object. The active contour is the deformable model for the object. The active contour used is attached to the strong edge and the target color in certain region.

As you may notice, the working principle of active contour fitter is similar to those of form-sensitive neurons in area V4. In primates, neurons in area V4 supposed to account for conscious perception of form and shape. In the system, active contour fitter simulate those neurons which extract the form or shape of the target and make extract measurement of the position of the object. By combining the location information as well as motion information, it is possible to make prediction of target's location in next time frame accurately. This is similar to the case that visual information from high-level processing areas is recurrent back to primary visual area for further processing and adjustment as shown in figure 1.

Active contour [22], [23] had been used in pattern location and tracking for a long time [24], [3]. It is good at attaching to object with strong edge and irregular shape. The snake can be interpreted as parametric curve $v(s) = [x(s), y(s)]$. The snake

is fitted according to the energy function:

$$E^*_{snake} = \int_0^1 \{[E_{int}(v(s))] + [E_{image}(v(s))] + [E_{con}(v(s))]\}ds \quad (3)$$

where E_{int} represents the internal energy of the snake due to bending, E_{image} represents the pixel-based image forces, and E_{con} represents the external constraint forces. The snake is said to be fitted if the E^*_{snake} is minimized.

In the proposed system, the initial position of the active contour is the bounding box of the searching window. It searches for strong edge along the direction toward to centroid of potential region. It stops at the pixel with strong edge characteristic and close to the color in target region (see figure 5). The active contour is also constrained by the curvature and continuity, but with relatively small weighting.

Fig. 5. *The left image is the input image. The right image shows the potential region in white color. The arrow shows the searching path. By searching along the line joining the initial position and the centroid, the pixel with strong edge information is picked as potential contour.*

C. Wavenet Estimator

It makes prediction of the location of the searching window based on the previous observations. The observations can be the observed locations of certain feature points or the whole object. Given a set of observed locations from time 0 to time t, the estimator aims at predicting the location at time t+1. The prediction is based on the smoothened trajectory of the previous observations. In most cases, the video frames captured are noisy and not suitable for analysis. The location of the target reported may be incorrect. This may cause discontinuities and ripples of the trajectory. The estimator will first remove those ripples in the observation curve. Prediction can be made according to the trend of the smoothened curve. The size of the searching window will be resized according to the confidence interval, which is calculated from the error. The larger the error, the wider the confidence interval and the larger the searching window will be.

Wavenet estimator simulates the function of motion- sensitive neurons in area V5. In area V5, neurons encode the direction and speed of the motion. In the proposed system, the movement of object is encoded by movement trajectory. Such trajectory will then be described by a set of wavelets. Each wavelet will then be represented by neuron node within wavenet. As you may notice, such neuron node is similar to neurons in area V5 in the sense that both of them encode motion and time information. By integrating the information given by neuron node in the system, prediction can be made accordingly.

Wavelet and wavenet was initially used in financial forecasting and signal processing [25], [26]. It is widely adopted in these fields because of its simplicity in structure and efficiency in handling curve modeling and smoothing problem.

The wavenet stands for wavelet network. It consists of wavelons and wavelinks. The wavelons are the neurons inside the wavenet. The wavelinks are the links that connect the wavelons. Wavenet's theory can be found in [27], [28]. The architecture of the wavenet is illustrated in figure 6. The input nodes allow signal to be fed into the system, and are connected to the mean node which represents the mean of the input signal. The wavelet-simulating nodes represent the major wavelet constituents. Every wavelet-simulating node has parameters that represent a certain wavelet. In our design, these wavelons have 2 parameters: transition and scale. They are connected to the output nodes which represent the smoothened signal at a certain time frame.

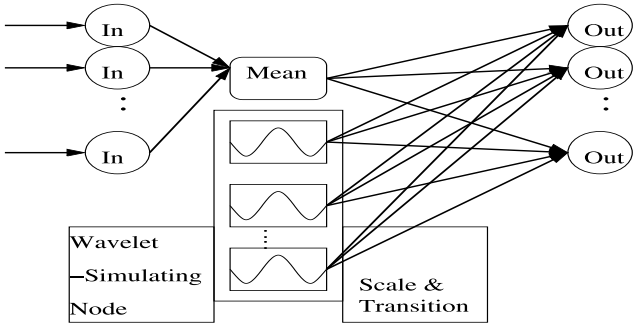

Fig. 6. *The input signal received by input nodes (In) is used to compute the mean signal. The wavelet-simulating nodes can be used to approximate the input signal. The approximated signal is represented by the output nodes (Out).*

The value of the output node is given by:

$$Out(t) = \sum_{i=1}^{N} w_i \psi_i(s_i, k_i, t) + w_0 Mean \quad (4)$$

where Out(t) is the result at output node t, $\psi_i(s_i, k_i)$ is the output at wavelet-simulating node i with scaling s_i and transition k_i, w_i is the wavelink i value, N is the number of wavelet-simulating node.

In terms of physiology, each wavelet-simulating nodes represents the motion state as encoded in V5 neurons. The motion state in neuroscience sense means the direction and speed. Each wavelet is actually a wave with certain position (transition) and spread (scale). Given certain timestamp, the value of the wave and the trend can be evaluated. Thus, the component direction (trend) and the speed (value) can be inferred by the wavelet-simulating nodes. This means each wavelon has the same role as V5 neurons. Besides, several neurons will be activated within the V5 pathway at anytime and the integration of such activations will turn out to be the motion pattern. Similarly, the summation of the output from

wavelon will represent the motion trajectory of the moving object and further prediction can be done accordingly. In primates, the motion state is estimated from observations; In the wavenet, the motion state (or wavelet parameters) is also learnt from observations (i.e. comparison between input and output nodes).

As described in figure 6, the input nodes store the input signal and the output nodes store the approximated signal. The difference or error terms will then be used to refine the parameters of the wavelet-simulating nodes and the corresponding weights. At every timestamp, the wavenet is trained to minimize the error terms at output nodes. After sufficient time of training, the parameters of the wavelet-simulating nodes will represent the major wavelets of the input signal. Those wavelet components ignored in the wavenet can be treated as noise and are removed.

The learning process aims at minimize the difference between the input nodes and the corresponding output nodes. The criteria function of the process is given by:

$$C = \sum_{t=0}^{T} [In(t) - Out(t)]^2 \tag{5}$$

By using gradient descent optimization approach, the refinement can be formulated as:

$$\delta w_i = \sum_{t=0}^{T} \left[2\left[In\left(t\right) - Out\left(t\right)\right]\left[-\psi_i(s_i, k_i, t)\right]\right] \tag{6}$$

$$\delta s_i = \sum_{t=0}^{T} \left[2\left[In\left(t\right) - Out\left(t\right)\right]\left[-w_i \frac{\delta \psi_i(s_i, k_i, t)}{\delta s_i}\right]\right] \tag{7}$$

$$\delta k_i = \sum_{t=0}^{T} \left[2\left[In\left(t\right) - Out\left(t\right)\right]\left[-w_i \frac{\delta \psi_i(s_i, k_i, t)}{\delta k_i}\right]\right] \tag{8}$$

At every timestamp, the weight, scale and transition are updated using the above scheme. The approximation will get closer to the input signal after sufficient time of training. Given a finite number of wavelons, only those relevant and best fit wavelets, which are usually those with large scale value, will be chosen. This means the problem of overfitting can be avoided because the final output depends on the composition of small number of major wavelets. The wavelet-simulating nodes will then represent the major components of the input signal and the trend of the signal can be inferred from these major components. The estimation can be done as:

$$Est(t+1) = \sum_{i=1}^{N} w_i \psi_i(s_i, k_i, t+1) + w_0 Mean \tag{9}$$

It represents the value of superposition of the major wavelets at the prediction time frame. The result is based on the smoothened trend of the input trajectory. The details of the estimation procedure is summarized in following paragraphs.

In the system, the refined centroid of the reported region in the other two estimators is used as the observation for the wavenet estimator. The initial searching window is of the size of whole image. After locating the approximated position of the object and fitting an active contour, the preliminary model is formed. The size of the model forms the base frame of searching window. The size and location of the window will be adjusted according to the result of this estimator.

The wavenet can make prediction based on a set of previous observations after several frames. Each prediction make use of limit number of observations. Each observation point is broken down into x and y direction. These components are fed into two wavenet estimators separately. Prediction from these estimators will form the x and y coordinate of the location of the searching window in next time frame.

The size of the searching window will depend on the prediction error and the curve fitting error. The prediction error is formulated as the difference between the input nodes and output nodes. It corresponds to how well the estimator makes prediction. The curve fitting error corresponds to how well the estimator approximates the previous observations. The larger these errors, the larger the size of the searching window will be. The size of the searching window will be adjusted using these error terms and the base frame size. Searching of target by the other two estimators will then be done again and the whole tracking process repeats.

IV. EXPERIMENTAL RESULT

The proposed system was implemented using Visual C++ under Microsoft Windows. Three experiments was performed to test the system. The experiments were done on a P4 2.26 GHz computer with 512M ram running Microsoft Windows. The object being tracked is the human hand and face.

A. Experiment 1: 1-D signal analysis

In this experiment, the smoothening and prediction performance of the wavelet estimator was tested. The input signal was the y-component of the 2D moving path of the hand. The curve fitting and prediction testing results are shown in figure 7 and figure 8 respectively. The result shows that the wavelet estimator can fit a curve to the noisy signal quite well but without the problem of overfitting. Besides, the prediction made based on the smoothened signal is close to the actual movement except the first half part of the signal which indicates the trajectory learning has not reached equilibrium. Once trajectory is learnt, the prediction result is good and the relative square error is low, which is less than 10%.

B. Experiment 2: tracking moving human body

In this experiment, the face and hand under clutter background was tracked. The result concerning fitting active contour and location estimation are shown in figure 9 and figure 10 respectively. It shows the target can be tracked even when the background consists of skin color. In addition, even when there is locomotion of hand, the trajectory can still be smoothened.

C. Experiment 3: Comparison with other approaches

In this experiment, similar tracking systems implementing Kalman filter and multi-layers neural network were tested. The

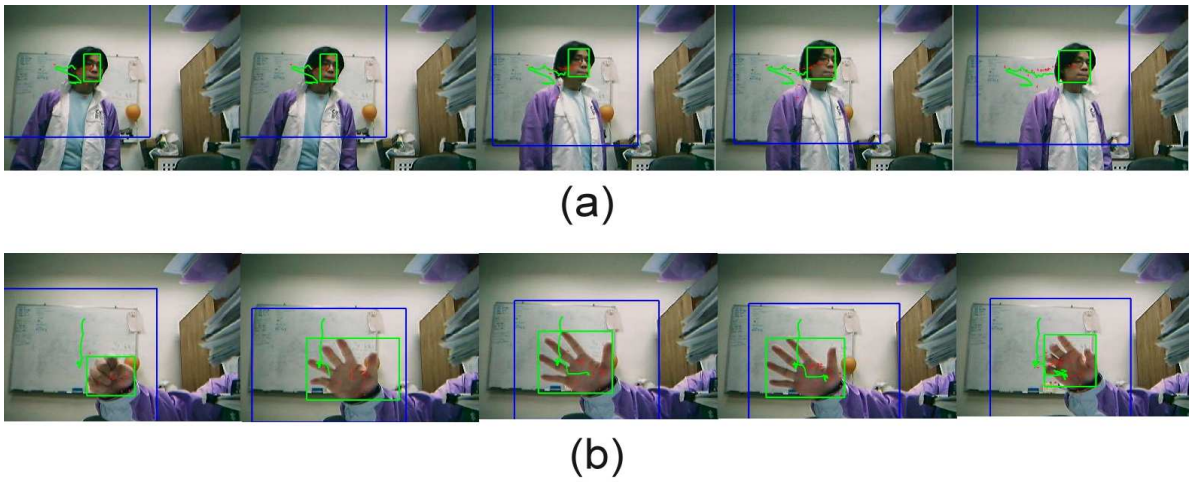

Fig. 10. *In the figures above, the video sequence for face and hand tracking is shown on (a) and (b) respectively. The green (lighter) block is the blob region detected. The blue (darker) block is the maximum bound of the searching window. The red fuzzy dots are the observations (the centroid of blob region). The green continuous line is the predicted trajectory which describes the smoothened trend of the blob region. It shows that both face and hand can be tracked accurately even under clutter background with distracting color (skin color of the balloon, bookshelf etc.).*

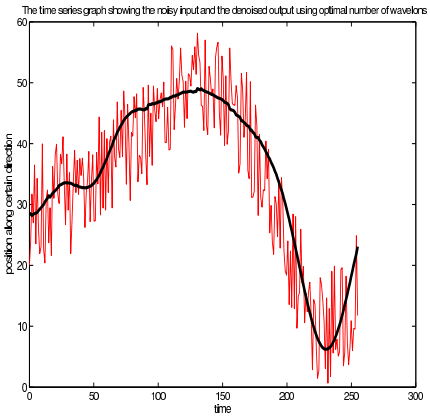

Fig. 7. *In this figure, the red fuzzy line represents the input signal (the y-component of the 2D moving path of the hand). The black solid line represents the smoothened signal using wavenet analysis. Even the input signal contains noise due to unreliability of other two estimators, the trajectory can be still smoothened.*

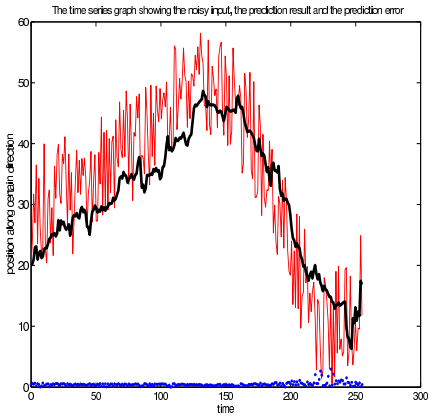

Fig. 8. *In this figure, the red fuzzy line represents the input signal (the y-component of the 2D moving path of the hand). The black solid line represents the prediction at each time frame using wavelet prediction approach. The blue marker line at the bottom shows the relative square error. It shows that prediction using wavelet is reliable even if the signal is very noisy. The predictions are usually close to the observations.*

result is shown in figure 11. From the figure, it shows that both approaches cannot make prediction accurately. For the Kalman filter, its main disadvantage is that we must provide it dynamic model before tracking. If the provided model is not the actual model (e.g. introduction of sudden change), the tracking will fail easily. As shown in the figure, the prediction result is simply the predefined motion with little adjustment. When there is sudden drop or acceleration, the prediction will be far from actual observations. For the multi-layers neural network, its main problem is the occurance of overfitting. As shown in the figure, when the signal is too noisy, the neural network will try to fit all the observations which cause the problem of overfitting. The prediction made will be inaccurate and not general enough. In both cases, the relative prediction error is over 20% in extreme case.

V. CONCLUSION

This paper endeavors to propose an alternate solution in tracking problem which simulate the visual system of primates. Commonly used filtering approach, such as Kalman filter, borrow ideas from signal theory and can be used under the condition that the dynamic model is known. Biological inspired approach, such as multi-layers neural network, have been proposed to learn the dynamic model in real time. However, such approach inherits the problem of overfitting in simple regression analysis, and thus cannot make prediction generally and accurately.

In the proposed system, it simulates the motion pathway in order to solve the tracking problem. As in visual area V1, V3 and V4, it uses color model, optical field, and the

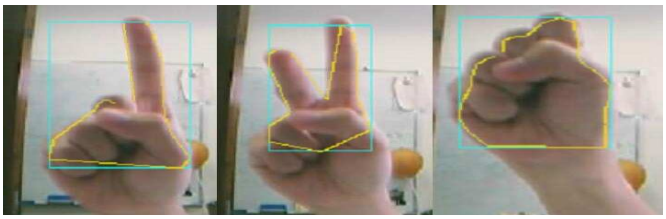

Fig. 9. *The light (blueish green) box is the bounding box for potential moving, skin-color region. The light (yellowish) line is the active contour. Even if the hand is changing its shape during tracking, the active contour can still attach to it quite well in real time.*

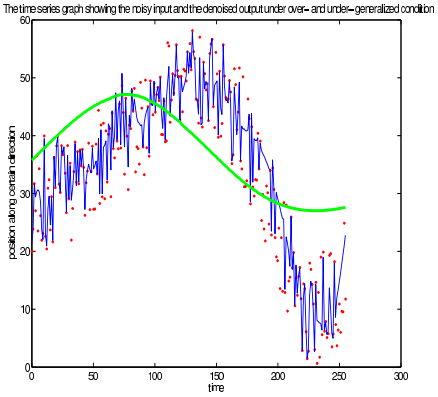

Fig. 11. *In this figure, the red fuzzy dots represents the input signal (the y-component of the 2D moving path of the hand). The green (light) line represents the prediction result using Kalman filter, and the blue (dark) line represents the prediction result using multi-layers neural network. It shows both Kalman filter and multi-layers network cannot make prediction accurately.*

active contour model to locate the position of object within a searching window. The searching window is automatically shifted and resized according to the prediction made from the wavenet estimator. This wavelet estimator indeed simulates the process taken place in visual area V5. Tracking is facilitated due to the reliable position and size of searching window and effective object location algorithm. Besides, the motion states and dynamics are learnt instead of predefined which provides flexibility in system design.

We have demonstrated the performance of the proposed tracking system using several sets of video sequences. The results show that the moving human parts are tracked in real time. The performance is much stable than similar system implementing Kalman filter and multi-layers network.

Although the proposed tracker can give relatively reliable estimation, the learning process is quite long. Furthermore, if we wish to have a better and more exact location of the whole object, the computational time will increase dramatically. Improvement have to be done on the learning algorithm, optimization algorithm, and object location algorithm in order to have more robust and reliable tracking result.

REFERENCES

[1] D. M. Gavrila, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding: CVIU*, vol. 73, no. 1, pp. 82–98, 1999.

[2] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding: CVIU*, vol. 81, no. 3, pp. 231–268, 2001.

[3] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *European Conference on Computer Vision*, 1996, pp. 343–356.

[4] ——, "Condensation – conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[5] ——, "ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework," *Lecture Notes in Computer Science*, vol. 1406, pp. 893–908, 1998.

[6] M. J. Black and A. D. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," in *European Conference on Computer Vision*, 1996, pp. 329–342.

[7] D. DeCarlo and D. N. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *International Journal of Computer Vision*, vol. 38, no. 2, pp. 99–127, 2000.

[8] V. Kruger, A. Happe, and G. Sommer, "Affine real-time face tracking using a wavelet network," in *Proc. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999, pp. 141–148.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[10] H. Breit and G. Rigoll, "Improved person tracking using a combined pseudo-2d-hmm and kalman filter approach with automatic background state adaptation," in *ICIP01*, 2001, pp. II: 53–56.

[11] M. Farmer, R. Hsu, and A. Jain, "Interacting multiple model (imm) kalman filters for robust high speed human motion tracking," in *ICPR02*, 2002, pp. II: 20–23.

[12] D. Jang, S. Jang, and H. Choi, "2d human body tracking with structural kalman filter," *PR*, vol. 35, no. 10, pp. 2041–2049, October 2002.

[13] H. M. Hunke, "Locating and tracking of human faces with neural networks," in *CMU-CS-TR*, 1994.

[14] L. T. Bruton, N. R. Bartley, and Z. Q. Liu, "The classification of motion in image sequences using 3d recursive adaptive filters to obtain neural network input vectors," in *Proc. of the IEEE International Conf. on Neural Networks*, 1995, pp. IV: 1595–1599.

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *Journal of Physiology*, vol. 195, pp. 215–243, 1968.

[16] R. C. Emerson, J. R. Bergen, and E. H. Adelson, "Directionally selective complex cells and the computation of motion energy in cat visual cortex," *Vision Research*, vol. 32, pp. 203–218, 1992.

[17] L. J. Croner and T. D. Albright, "Seeing the big picture: integration of image cues in the primate visual system," *Neuron*, vol. 24, pp. 777–789, 1999.

[18] G. Beckers and V. Homberg, "Cerebral visual motion blindness: transitory akinetopsia induced by transcranial magnetic stimulation of human area v5," in *Proceedings of the Royal Society of London, Series B, 249*, 1992, pp. 173–178.

[19] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[20] S. Khan and M. Shah, "Object based segmentation of video using color, motion and spatial information," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, pp. 746–751.

[21] R. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face detection in color images," in *Proc. IEEE ICIP*, 2001, pp. 1046–1049.

[22] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. Int. Conf. on Computer Vision*, 1987, pp. 259–268.

[23] D. J. Williams and M. Shah, "A fast algorithm for active contours," in *Proc. Int. Conf. on Computer Vision*, 1990, pp. 592–595.

[24] A. Blake and M. Isard, *Active Contours*. Springer, 1998.

[25] C. Chang, W. Fu, and M. Yi, "Short term load forecasting using wavelet networks," *Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 6, pp. 217–223, 1998.

[26] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.

[27] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 889–898, 1992.

[28] Q. Zhang, "Using wavelet network in nonparametric estimation," *IEEE Trans. Neural Networks*, vol. 8, no. 2, pp. 227–236, 1997.