Real-Time Image Tracking with An Adaptive Complementary Filter

Zuocai Wang, Bin Chen, Jin Wu and Tao Yan

***Abstract*—Image tracking is a key technique in many applications. Various algorithms have been proposed to conduct image tracking in different environments. However, many of them does not achieve good performances in terms of drastic motion changes of the image. To enhance the tracking ability in such circumstances, we present a novel algorithm using the complementary filter. An adaptive law is designed in this case to make the filter much more accurate when facing limiting conditions. Experiments are carried out showing the performances of the proposed method and other representative methods. Comparison results reflect that the proposed filter is more efficient than representative ones.**

***Index Terms*—Image Tracking, Complementary Filter, Computer Vision, Industrial Electronics.**

I. Introduction

CONSUMER electronics have extensively changed our daily life [1]. Due to this, camera, as an expensive tool in ancient times, has become very popular nowadays. Using a camera, the user can take arbitrary amount of photos and videos [2], [3]. In fact, throughout biological research, imaging is proved to be the most important signal source of the human body. This indirectly shows that one image may contain numorous useful information [4], [5].

In computer engineering, camera is usually adopted to track a certain object automatically [6]. This needs precision data acquisition, system model and etc. to make sure the estimated area is correct [7], [8]. System modelling is usually formed empirically with some known parameters [9], [10]. However, with the change of object's motion, the system model varies at the same time. One option is to use inertial measurements to observe the attitude and translation. Since this operation requires integration of angular rate and acceleration, the error diverges as time increases [11]. Moreover, the addition of the sensors will directly add the cost of the whole tracking device. To overcome this, many algorithms uses features of the images to track certain areas [12], [13]. For instance, the scale-invariant feature transform (SIFT, [14]) and speeded up robust features (SURF, [15]) are mostly adopted. These methods can obtain efficient feature extraction but may meanwhile result in outliers. Random sample consensus (RANSAC, [16]) is then utilized to remove outliers with correspondence analysis. However, such sequential operations would make the whole process time-costly [17], [18].

In some platforms with low hardware configuration e.g. embedded chips, smart wearables, the computational resources are highly restricted [19], [20]. This generates another demand on vision-based object tracking. Many existing methods suffer from their high time consumption in real applications. This is because many operations like eigen value decomposition and singular value decomposition (SVD, [21]) are included. Recently, a spatio-temporal context (STC, [22]) learning method is introduced by Zhang et al [23]. which updates the estimates of the tracked area with spatio-temporal information. This method is proved to be very fast with the processing speed at 350FPS on an i7-CPU computer. Although the algorithm is proved by experiments to be very effective, it still has some disadvantages. Since STC is composed by the spatio-temporal context and spatial context together with a weight, the tracking performance is significantly restricted by this weight. In presented literature, the weight is an empirically given fixed constant which denotes basically one certain tracking speed. However, when the image's motion becomes drastic, the fixed weight can not adapt the new motion very well, which, would then induce larger errors.

In this paper, the associated image tracking task is modelled with a complementary filter. Adaptive law is then designed based on existing works on control theory and applications. We name the method as the Complementary-Filter STC (CF-STC). Besides, some critical issues around the proposed method are discussed as well which ensures the correctness and robustness of the proposed filter. Experiments are conducted which give comparisons between the CF-STC and other representative methods.

This paper is briefly structured as follows: Section II gives the brief introduction to complementary filter. Section III includes the theory of image tracking. The proposed adaptive law is given in Section IV. Section V contains the experiments, results and comparisons. Concluding remarks are given in Section VI.

II. Complementary Filter

A. Basic Knowledge

For a dynamical system, it can be modelled with

$$\left\{ \begin{array}{l} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{array} \right. \tag{1}$$

where $\mathbf{x}$ and $\mathbf{y}$ denote the state variable and measurement respectively. t stands for the time while $\mathbf{u}$ denotes the external control input. In most estimation cases, the control

Z. Wang is with Chengdu Institute of Computer Applications, Chinese Academy of Sciences and University of Chinese Academy of Sciences, Chengdu, 610041, China. (e-mail: wzuocai@163.com)

Correpondent: B. Chen is with Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, 610041, China. (e-mail: chenbin306@sohu.com)

J. Wu is with School of Automation, University of Electronic Science and Technology of China (UESTC), Chengdu, China. (e-mail: jin_wu_uestc@hotmail.com)

T. Yan is with Chengdu Institute of Computer Applications, Chinese Academy of Sciences and University of Chinese Academy of Sciences, Chengdu, 610041, China. (e-mail: yantao@cbpm-kexin.com)

(Advance online publication: 10 February 2018)

input is not existed hence the equation is rewritten as

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \end{cases} \tag{2}$$

For discrete time, we have

$$\begin{cases} \mathbf{x}_k = \mathbf{\Phi}_{k,k-1}\mathbf{x}_{k-1} \\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k \end{cases} \tag{3}$$

where k denotes the kth time epoch with $t = k\Delta t$ and Δt is the sampling time interval. $\mathbf{\Phi}_{k,k-1}$ is called the transition matrix which transforms the state in $k-1$th epoch to the kth epoch. Now we are going to conduct state estimation based on the state model and measurement model given above. Luenberger observer can achieve this by making a copy of $\mathbf{y}_k$, such that [24]

$$\mathbf{y}_k^- = \mathbf{C}\mathbf{x}_k^- \tag{4}$$

where

$$\mathbf{x}_k^- = \mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} \tag{5}$$

where $\hat{\mathbf{x}}$ denotes the estimation of $\mathbf{x}$. By associating the state and measurement models with a gain matrix $\mathbf{L}$ (also known as the Luenberger gain), we have

$$\hat{\mathbf{x}}_k = \mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{L}(\mathbf{y}_k - \mathbf{y}_k^-) \tag{6}$$

Inserting (4) into (6) gives

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{L}(\mathbf{y}_k - \mathbf{y}_k^-) \\ &= \mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{L}\mathbf{y}_k - \mathbf{L}\mathbf{C}\mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} \\ &= (\mathbf{I} - \mathbf{L}\mathbf{C})\mathbf{\Phi}_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{L}\mathbf{y}_k \end{aligned} \tag{7}$$

We can see from this equation that the prediction $\mathbf{\Phi}\hat{\mathbf{x}}_{k-1}$ and the measurement $\mathbf{y}_k$ compensate for each other for the final state estimation. This is the definition of generalized complementary filter. In common applications, measurements will be given in the same form with the state variable. In this case, $\mathbf{C} = \mathbf{I}$ and (7) becomes

$$\begin{aligned} \hat{\mathbf{x}}_k &= (\mathbf{I} - \mathbf{L})\Phi_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{L}\mathbf{y}_k \\ &= (\mathbf{I} - \mathbf{L})\mathbf{x}_k^- + \mathbf{L}\mathbf{y}_k \end{aligned} \tag{8}$$

Here, the Luenberger gain becomes the complementary gain [1]. The gain degenerate to a constant when $\mathbf{L} = \zeta\mathbf{I}, \zeta > 0$, which produces a more simplified form

$$\hat{\mathbf{x}}_k = (\mathbf{I} - \mathbf{L})\mathbf{x}_k^- + \mathbf{L}\mathbf{y}_k = (1-\zeta)\mathbf{x}_k^- + \zeta\mathbf{y}_k \tag{9}$$

B. Adaptive Law

When the state estimation is given with (9), one gain ζ_0 can only ensure one type of filter convergence. In the worst case, the filter would even be divergent with time increasing. Adaptive law is studied in many related applications to make sure the filter have strong-tracking ability so that the estimated state is well fit the real conditions. Note that the system may be drastic when the derivative of the state is too large. To detect this, we can empirically set a threshold, such that

$$\left\| \frac{d\mathbf{x}(t)}{dt} \right\| > J \tag{10}$$

where J denotes the threshold, $\|\cdot\|$ stands for the norm 2. For discrete system, the criterion can be transformed into

$$\|\mathbf{x}_k - \mathbf{x}_{k-1}\| > J\Delta t \tag{11}$$

Then we can use the results in last sub-section, namely

$$\mathbf{x}_k - \mathbf{x}_{k-1} \approx \mathbf{x}_k^- - \hat{\mathbf{x}}_{k-1} = (\mathbf{\Phi}_{k,k-1} - \mathbf{I})\hat{\mathbf{x}}_{k-1} \tag{12}$$

Once the system is detected to be drastic, we can adjust the complementary gain dynamically. For common linear system, the violence of the drastic status is linear with the derivative of the state. Hence we can design the following adaptive law

$$\begin{cases} \zeta = \zeta_0, \|\mathbf{x}_k - \mathbf{x}_{k-1}\| \le J\Delta t \\ \zeta = \gamma\zeta_0, \|\mathbf{x}_k - \mathbf{x}_{k-1}\| > J\Delta t \end{cases} \tag{13}$$

where the scale factor γ is given by

$$\gamma = K\frac{\|(\Phi_{k,k-1} - \mathbf{I})\hat{\mathbf{x}}_{k-1}\|}{\Delta t} \tag{14}$$

where K denotes a zoom ratio.

III. Image Tracking

Biologically, each image has the focus of attention, which can be easily extracted. This information can be combined with the confidence map of the tracked area to make prediction of the spatio-temporal context (STC). Mathematically, the tracking problem is equivalent to finding a confidence map of the maximum likelihood of the object location [3]

$$c(\mathbf{r}) = P(\mathbf{r}|\mathbf{S}) \tag{15}$$

where $\mathbf{r}$ is the object location and $\mathbf{S}$ stands for the object in the scene. Using the image intensity, the context feature set is defined as

$$\mathbf{F}^c = \{\mathbf{c}(\mathbf{z}) = (Intensity(\mathbf{z}), \mathbf{z}) | \mathbf{z} \in \Omega_c(\mathbf{r}^*)\} \tag{16}$$

where $\mathbf{c}(\mathbf{z})$ denotes the context feature at location $\mathbf{z}$ while Ω_c is the neibourghood of the tracked location $\mathbf{r}^*$. Note that the tracked location is defined by

$$\mathbf{r}_t^* = \underset{\mathbf{r}\in\Omega_c(\mathbf{r}_{t-1}^*)}{\arg\max}\ c_t(\mathbf{r}) \tag{17}$$

Then the confidence map can be further given by

$$\begin{aligned} & c(\mathbf{r}) = P(\mathbf{r}|\mathbf{S}) \\ & = \textstyle\sum_{\mathbf{c}(\mathbf{z})\in\mathbf{F}^c} P(\mathbf{r}, \mathbf{c}(\mathbf{z})|\mathbf{S}) \\ & = \textstyle\sum_{\mathbf{c}(\mathbf{z})\in\mathbf{F}^c} P(\mathbf{r}|\mathbf{c}(\mathbf{z}), \mathbf{S})P(\mathbf{c}(\mathbf{z})|\mathbf{S}) \end{aligned} \tag{18}$$

where $P(\mathbf{r}|\mathbf{c}(\mathbf{z}), \mathbf{S})$ is called the spatial context model while $P(\mathbf{c}(\mathbf{z}), \mathbf{S})$ is the context prior model. The confidence map can be given by

$$c(\mathbf{r}) = P(\mathbf{r}|\mathbf{S}) = be^{-\left|\frac{\mathbf{r}-\mathbf{r}^*}{\alpha}\right|^\beta} \tag{19}$$

where α, b, β are parameters, while the context prior model is given by

$$P(\mathbf{c}(\mathbf{z})|\mathbf{S}) = Intensity(\mathbf{z})ae^{-\frac{|\mathbf{z}-\mathbf{r}^*|^2}{\sigma^2}} \tag{20}$$

where σ denotes the value of the variance of the tracked location. And the spatial context model is given by

$$h^{sc}(\mathbf{x} - \mathbf{z}) = P(\mathbf{r}|\mathbf{c}(\mathbf{z}), \mathbf{S}) \tag{21}$$

Hence we can obtain the spatial context model with fourier transformation, since

$$\begin{aligned} & c(\mathbf{r}) = \textstyle\sum_{\mathbf{c}(\mathbf{z})\in\mathbf{F}^c} h^{sc}(\mathbf{x} - \mathbf{z})P(\mathbf{c}(\mathbf{z})|\mathbf{S}) \\ & = h^{sc}(\mathbf{r} - \mathbf{z}) \otimes P(\mathbf{c}(\mathbf{z})|\mathbf{S}) \end{aligned} \tag{22}$$

where $\otimes$ stands for the convolutional operator. Then the spatial context model is obtained by

$$h^{sc}(\mathbf{x}-\mathbf{z}) = F^{-1}\left\{\frac{F\left(be^{-\left|\frac{\mathbf{r}-\mathbf{r}^*}{\alpha}\right|^{\beta}}\right)}{F\left(P(\mathbf{c}(\mathbf{z})|\mathbf{S})\right)}\right\} \tag{23}$$

where F means the fourier transform and F^{-1} is the inverse fourier transform. When the spatial context is obtained, it can be associated with the last obtained one in the weighted form, which generates the STC

$$H_k^{stc} = (1-\beta)H_{k-1} + \beta h_{k-1}^{sc} \tag{24}$$

The scale and variance are given by

$$\begin{cases} s'_t = \sqrt{\frac{c(\mathbf{r}_t^*)}{c(\mathbf{r}_{t-1}^*)}} \\ \bar{s}_t = \frac{1}{n}\sum_{i=1}^{n} s'_{t-i} \\ s_{t+1} = (1-\lambda)s_t + \lambda\bar{s}_t \\ \sigma_{t+1} = s_t\sigma_t \end{cases} \tag{25}$$

Algorithm 1 Improved image tracking algorithm using adaptive complementary filter.

Initialize:
Initial gain: ζ_0,
Zoom ratio: K,
Initial image position: $\mathbf{r_0}$, Time epoch: k=0.
Output: $\mathbf{r}^*$.
while image obtained **do**

1) $k = k + 1$.
2) Extract the tracked position from last spatio-contemporal context.
3) Get context prior model from (20).
4) Get the spatial context model from (21).
5) **IF** $t = 1$, then the spatio-contemporal model is initialized as $H_{k=1}^{stc} = h_{k=1}^{sc}$.
6) **ELSE**
7) Aadptively calculate the weight β with $\beta = \frac{K\left\|H_k^{stc} - H_{k-1}^{stc}\right\|}{\Delta t}$.
8) Update the spatio-contempotal model with (24).

end while

We can see from (24) and (9) have very similar mathematical form. This indicates that the spatio-contemporal context model can be basically described by a complementary filter. The internal reason is that based on such model, the state variable is chosen as the spatio-contemporal context model. Here, H_k^{stc} denotes the estimated state, H_{k-1} constitutes the state that inherited from last time epoch, namely the estimated state. Using h_{k-1}^{sc}, the state is compensated with measurements. It should be noted that here the state variable no longer be vectors. Yet, throughout the adaptive law discussion given above, we can design an improved image tracking algorithm, which is given in the Algorithm 1. The adaptive law is modified to make (24) adaptively updated with the gain β. The commitment is that we use the inf-norm to replace the former norm-2 so that the difference information can be represented more effectively.

IV. Experiments and Results

In this section, we conduct experiments to show the advantage of the proposed algorithm.

A. Experiment 1: Face Tracking with Normal Motion

Fig. 2: Data set for validation of the proposed algorithm.

The data set is from the bitmap (BMP) Image Sequences for Elliptical Head Tracking [25]. The overall information of the data set is shown in Fig. 2. From the data set, we can see that it contains many sophisticated motions, which in a degree adds the complexity to the face tracking task. To validate the performance, the parameters are initialized as $\zeta_0 = 0.05$, $K = 0.01$ while the initial face position is set to $\mathbf{r}_0 = (52, 32)$ with the window size of 40 pixels. The initial gain is set to 0.055. By running the test program with MATLAB r2015b on a PC with configuration of an i7-4core CPU, 8GB RAM and 512G solid state disk, the results are obtained from two different algorithms. Fig. 2 shows the tracking performances of the two estimators. The upper ones are from the proposed filter while the lower one are from Zhang's algorithm. We can easily see that our method is better than Zhang's method in this case. This is because the gain in Zhang's method is fixed while the gain is chosen adaptively in the proposed filter. The information of the adaptive gains is depicted in Fig. 3. As shown, we can see that the gain changes with the general motion of the face, which enhances the tracking performance.

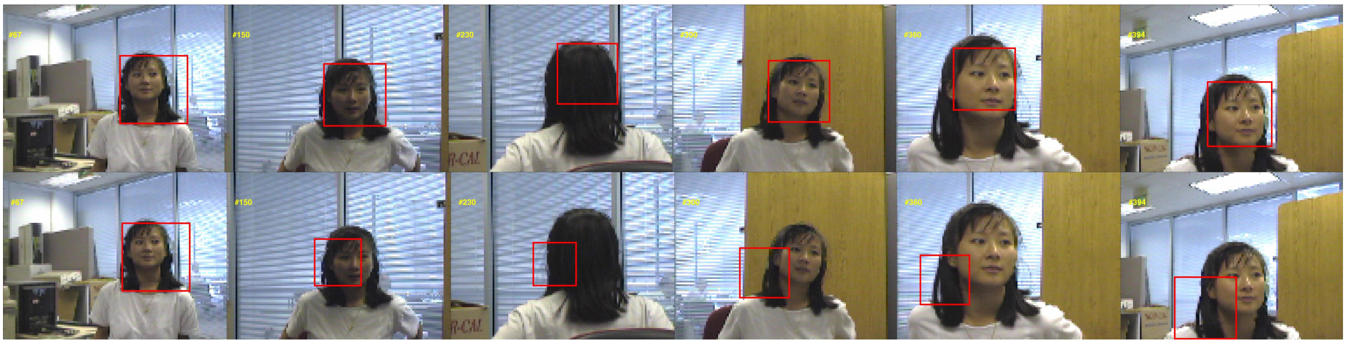

Fig. 1: The tracking performance of two different algorithms. The upper ones are from the proposed filter while the lower ones are from Zhang's method.

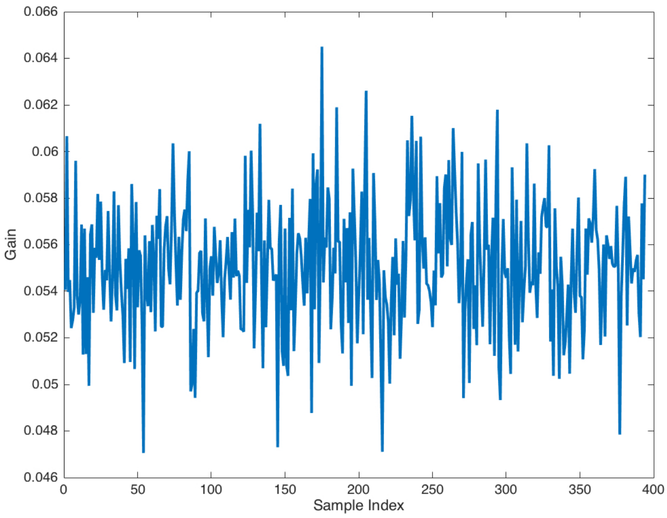

Fig. 3: The adaptive gain calculated by the proposed filter.

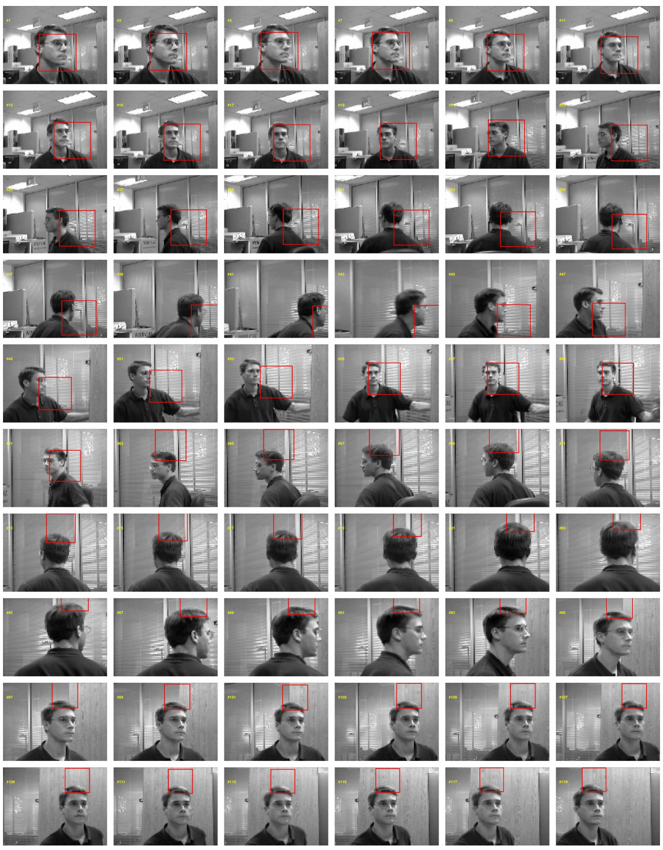

Fig. 4: Image tracking results of the conventional algorithm.

B. Experiment 2: Image Tracking with Fast Motion

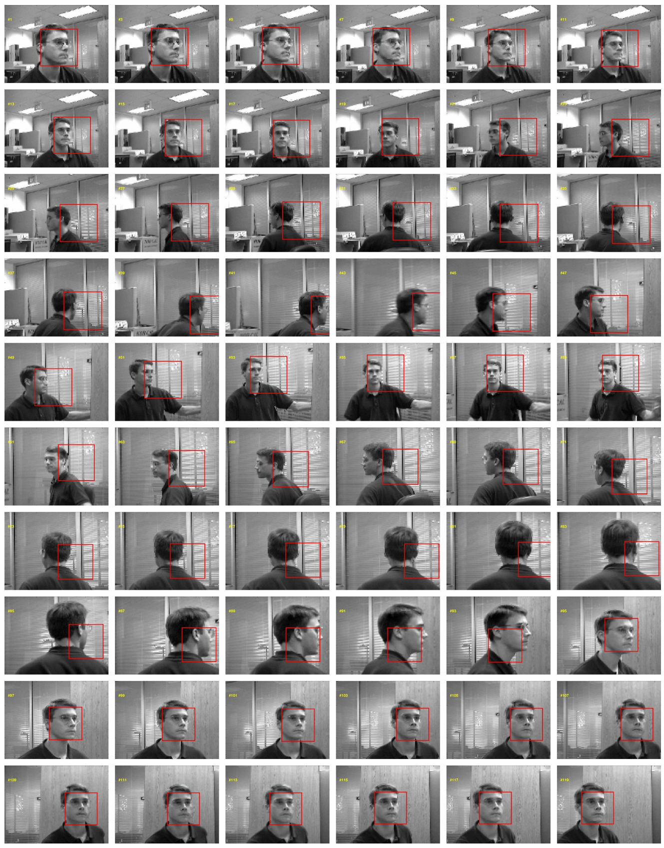

Fig. 5: Image tracking results of the proposed algorithm.

In this sub-section, the image tracking results are evaluated by a public image data set where the face motion is mostly drastic. The initial gains are set to 0.0325 jointly for the two compared algorithms. In this way, the two methods are evaluated which generates the Fig. 4 and 5. It is obvious that the conventional one cannot reject the drastic motion well so that the final results seem to be very far way from expected positions. However, the proposed algorithm adopts adaptive gain for image tracking. It automatically detects the drastic motion hence the estimated results are much more better than the conventional one. The corresponding gains during the tracking computations are plotted in Fig. 6.

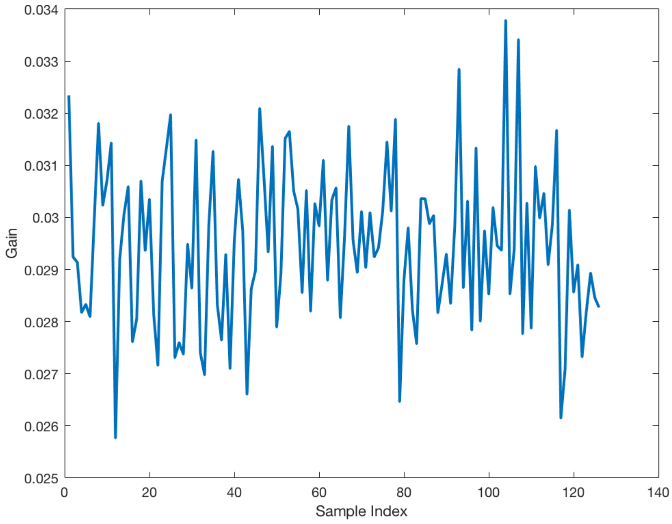

Fig. 6: The adaptive gain calculated by the proposed filter.

Fig. 7: Sample receipts to be identified and tracked in the experiment.

C. *Experiment 3: Receipt Serial Number Tracking*

In modern living, there are so many receipts that describes the history of numerous trading activities. For batch processing of such information, we may expect to find an effective way to track the serial numbers of the receipts. Using the proposed method, the task can be accomplished. The picked-up samples are shown in the Fig. 7. We may see from these figures that the brightness changes. Also, there are some fuzzy areas inside the tracked region. With our developed algorithm, the result is shown in Fig. 8. The general parameters are the same with that in last sub-section while the initial position of the area is set to $\mathbf{r}_0 = (302, 243)$.

Fig. 8: Tracked result using the sample receipts.

We can see that the proposed algorithm can efficiently track the serial numbers of the receipts no matter how the local features changes. This is because of the proposed algorithm adopts the spatio-contemporal context and later identify the state estimates with adaptive complementary filter. The context information enables the algorithm to be smooth and robust that can reject outer disturbances induced by images noises, drastic motion and etc.

V. Experiment 4: Pedestrian Tracking

The proposed algorithm can also be applied for object tracking for autonomous detecting. In real applications, the pedestrian tracking leads to automatic detecting ability of robots. In this experiment, the raw data is from a famous representative paper [26]. The parameters are the same with aforementioned experimental parts. The results are shown in Fig. 9 and 10. The raw image sequences are in fact formed by moving pedestrians on the road where one is often covered by another. In such occasion, the tracking ability is significantly challenged. With fixed-constant gain given in conventional literature, the tracker is no longer available just few frames after the initialization. The proposed method, however, can adaptively compute the gain and dynamically detect the position of the chosen pedestrian. Obviously, this again verifies the robustness of the proposed method.

VI. Conclusion

Based on control theory, this paper solved the image tracking problem when the obtained images contains drastic motion. The fixed-constant gain of the complementary filter is studied. Adaptive law is then designed to enhance the strong-tracking ability of the filter. Using the spatio-contemporal model of describing the motion of images, we improved the estimation performance with the adaptive

Fig. 9: Pedestrian tracking results using proposed method.

Fig. 10: Pedestrian tracking results using conventional method.

complementary filter. Experiments and analysis are given which shows that the proposed algorithm can track the area more effectively when large motion takes place. In other words, the adaptive ability is much better than the previous one. We hope that this method would be of benefit to related applications with image tracking.

ACKNOWLEDGMENT

This research was supported by the West Light Foundation of the Chinese Academy of Sciences (No. 2011180) and National Natural Science Foundation of China under grant number of No. 61450010.

REFERENCES

[1] J. Wu, Z. Zhou, J. Chen, H. Fourati, and R. Li, "Fast Complementary Filter for Attitude Estimation Using Low-Cost MARG Sensors," *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6997–7007, 2016. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7508437

[2] J. Cruz-Mota, M. Bierlaire, and J. P. Thiran, "Sample and pixel weighting strategies for robust incremental visual tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 5, pp. 898–911, 2013.

[3] X. Zhang, W. Hu, S. Chen, and S. Maybank, "Graph-embedding-based learning for robust object tracking," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 2, pp. 1072–1084, 2014.

[4] I.-C. Chang and S.-Y. Lin, "3D human motion tracking based on a progressive particle filter," *Pattern Recognition*,

vol. 43, no. 10, pp. 3621–3635, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320310002049
[5] T. Darrell, B. Moghaddam, and A. P. Pentland, “Active face tracking and pose estimation in an interactive room,” in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR’96, 1996 IEEE Computer Society Conference on*. IEEE, 1996, pp. 67–72.
[6] J. Tu, H. Tao, and T. Huang, “Face as mouse through visual face tracking,” *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 35–40, 2007.
[7] K. Lu, Z. Ding, and S. Ge, “Sparse-representation-based graph embedding for traffic sign recognition,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1515–1524, 2012.
[8] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu, “Unsupervised Object Class Discovery via Saliency-Guided Multiple Class Learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 862–75, 2015. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6888473
[9] A. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.
[10] A. K. Jain and S. Z. Li, *Handbook of face recognition*. Springer, 2011.
[11] C.-h. Chen, L.-F. Pau, and P. S.-p. Wang, *Handbook of pattern recognition and computer vision*. World Scientific, 1993, vol. 2.
[12] F. Yang and M. Paindavoine, “Implementation of an rbf neural network on embedded systems: real-time face tracking and identity verification,” *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1162–1175, 2003.
[13] X. Li, A. Dick, C. Shen, A. Van Den Hengel, and H. Wang, “Incremental learning of 3D-DCT compact representations for robust visual tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 863–881, 2013.
[14] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” *International journal of computer vision*, vol. 74, no. 1, pp. 59–73, 2007.
[15] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
[16] R. Raguram, J.-M. Frahm, and M. Pollefeys, “A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus,” in *European Conference on Computer Vision*. Springer, 2008, pp. 500–513.
[17] K. Schwerdt and J. L. Crowley, “Robust face tracking using color,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 90–95.
[18] H. Liu, S. Li, and L. Fang, “Robust Object Tracking based on Principal Component Analysis and Local Sparse Representation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 61172161, pp. 2863–2875, 2015.
[19] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.
[20] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
[21] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-Squares Fitting of Two 3-D Point Sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.
[22] J. Sun, X. Wu, S. Yan, L.-F. Cheong, T.-S. Chua, and J. Li, “Hierarchical spatio-temporal context modeling for action recognition,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2004–2011.
[23] K. Zhang, L. Zhang, M.-H. Yang, and D. Zhang, “Fast Tracking via Spatio-Temporal Context Learning,” *arXiv preprint arXiv:1311.1939*, pp. 1–16, 2013. [Online]. Available: http://arxiv.org/abs/1311.1939
[24] M. Zeitz, “The extended luenberger observer for nonlinear systems,” *Systems & Control Letters*, vol. 9, no. 2, pp. 149–156, 1987.
[25] (1998, August). [Online]. Available: http://cecas.clemson.edu/ stb/research/headtracker/seq/
[26] A. Ess, B. Leibe, and L. Van Gool, “Depth and appearance for mobile scene analysis,” in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

Zuocai Wang received his B.S. Degree in 2012. He is now with Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China, where he working as a PhD candidate. His research interests include computer vision, automation science and electronics.

Bin Chen received his B.S. Degree and M.S. Degree in 2012 and 2013 respectively. He is now a Professor with Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China. His research interests include intelligent systems, machine vision and pattern recognition.

Jin Wu received his B.S. Degree in 2016. He is now with School of Automation, University of Electronic Science and Technology of China, Chengdu, China, as a research assistant. His research interests include multi-sensor fusion, automatic control and systems engineering.

Tao Yan received his B.S. Degree in 2012. He is now with Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China, where he working as a PhD candidate. His research interests include computer vision, instrumentation and measurement.