Moving target-tracking algorithm based on sparse representation and particle filter

Qiu-fen Yang[1*], Can-jun Li

Department of Computer Science, Hunan Radio &TV University, Changsha, 410004, China

**Corresponding atuhor's e-mail: yqf5569@sohu.com*

Received 12 January 2013, www.cmnt.lv

Abstract

This paper proposes a target tracking algorithm based on 2-dimensional PCA (principal component analysis), which can solve the difficulty of current target tracking algorithm to adapt to the appearance change of target caused by the illumination, shield and position change. First of all, the 2-dimensional PCA method A and sparse representation are used to build the target appearance model, which can reduce the dimension of target; then, by introducing the update method of increment subspace to conduct online update of the target template, it can reduce the algorithm's requirement of memory space and increase the accuracy of target appearance description; finally, the simulation experiment is conducted. The simulation result shows that compared to other tracking algorithm for moving target, the algorithm proposed in this paper can more accurately track the moving target in the video image, which also shows great robustness to the illumination and position change, and it has significant advantages for the target tracking with serious shield.

Keywords: Object tacking, Sparse representation, Incremental learning, Appearance change

1 Introduction

Visual object tracking is a hot spot in current computer vision studies, covering video monitoring, human-machine interface, robot perception, behaviour understanding and motion recognition, etc. As the tracking process would be affected by such factors as the rotation, deformation, shielding and the changes of lighting, visual tracking technology has always been worth further researches [1].

Current object tracking methods mainly include: template matching-based tracking methods [2], filtering theory-based tracking methods [3] and classification-based tracking methods [4]. The template matching-based methods have the advantages of being simple and high matching accuracy, but are sensitive to shielding and deformation. Filtering theory-based tracking methods include two major types, Kalman Filter [5-6] and Particle Filter [7-8]. Kalman Filter can only deal with linear, Gaussian and single mode conditions. Particle Filter is suitable for non-linear and non-Gaussian object tracking. The classification-based methods is to consider object tracking as a binary classification problem [9], applying classification treatment to the foreground and background, using classifier to classify the tracking area and realize accurate positioning of the object. However, constructing a classifier needs massive positive and negative samples and it is not suitable for the requirement on high real-time capability. In recent years, the sparse representation theory effectively solves the object recognition issue under changes of lighting and gestures and under shielding, and is gradually applied in object tracking. Mei, et al [10] were the first to introduce sparse representation in the frame of Particle Filter and in object tracking. They used L1 minimization to solve the sparse representation coefficient, settling the issue of object shielding in a very satisfactory manner. However, when the appearance of the objects has major changes, this algorithm could not track the objects stably, entailing high complicity and massive computing amount.

Based on the analysis above, under the framework of Particle Filter, this paper uses 2DPCA method and sparse representation to establish the model of object appearance, reduces the number of dimensions of the object template. Through introducing the incremental subspace updating method to conduct online update of the object template, reducing the requirement of the algorithm on storage space and increasing the accuracy of the description of the object appearance.

2 Sparse representation-based object tracking

Two-dimension principal component analysis (2D-PCA) is a forward image processing techniques of image characteristics extraction, which directly use two-dimension image data as the analysis object to construct covariance matrix. The calculation of 2D-PCA is easy, and feature extraction time is short, which especially fits the image feature extraction of facial image. This method has successful application in the fields of facial identification and so on. Compared with traditional PCA method, 2D-PCA directly use two dimension image matrix as the analysis object, it doesn't need to transfer image into one dimension vector quantity first and calculate then as traditional PCA, so this method preferably stores the two dimension space information of image. At the same time, 2D-PCA directly use original image matrix to construct it while constructing image covariance matrix, compared with traditional PCA covariance matrix, dimension number of image covariance matrix in 2D-PCA is much smaller, which not only reduces calculation quantity, but also avoids matrix singular problem of PCA from happening when the training sample number are relatively few.

2.1 2DPCA EIGENSPACE

Assuming that $Y_t = \{y_1, y_2, \ldots y_n\}$ represents the observation sample set of the object on t frames, where $y_i \in R^{d*d}$

denotes No. i Observation sample, $I_t = \frac{1}{n}\sum_{i=1}^{n} y_i$ denotes the mean image of the observation sample set to Y_t. To lower down the storage cost and the computing time of the computer, the object image is applied with 2DPCA. Compared with PCA, 2DPCA is a method based on images and there is no need for a vector transformation. The calculation of the co-variance matrix is direct and simple and needs less time to calculate the eigenvector. 2DPCA algorithm is described below:

(1) Calculating the total scatter of the observation sample set Y_t

$$S_t = \frac{1}{n}\sum_{i=1}^{n}(y_i - I_t)(y_i - I_t)^T = \frac{1}{n}Y_t Y^T$$

(2) Calculating the eigenvalues of the scatter matrix $\lambda_1 \ge \lambda_2 ... \ge \lambda_n$.

(3) Calculating the unit orthogonal eigenvectors corresponding to the eigenvalues $u_1, u_2, ..., u_n$

(4) Reserve the eigenvectors corresponding to the first k eigenvalues to establish the eigenspace $\mathrm{U} = \{\mathrm{u}_i\}|_{i=1}^{k}$.

2.2 SPARSE REPRESENTATION

Given that the object template set and its eigenspace $U = \{\mathrm{u}_1, \mathrm{u}_2, \cdots \mathrm{u}_k\} \in \mathrm{R}^{d\times k} (d \gg k)$ contain k object templates, then the tracking result $y \in R^d$ could be approximately denoted with the eigenspace $\mathrm{U} = \{\mathrm{u}_i\}|_{i=1}^{k}$ below:

$$y \approx Ua = a_1 u_1 + a_2 u_2 + \cdots + a_k u_k , \tag{1}$$

where: $a = (a_1, a_2, ..., a_k)^T \in R^k$ denotes the coefficient vector of the object. As in the tracking process, the object is frequently shielded or interfered by noises and errors are often produced, an error term is introduced below:

$$Y = Ua + \varepsilon , \tag{2}$$

where: $\varepsilon \in R^d$ represents the error term introduced by noises and shielding, ε 's non-zero elements denote the noises or shielding of the object. We could use a unit matrix $E = [e_1, e_2, ..., e_d]^T \in R^{d\times d}$ to locate the position of disturbance. So Equation (2) could be rewritten as:

$$y = [U, E]\begin{bmatrix} a \\ b \end{bmatrix} = Dc , \tag{3}$$

where, $b = (b_1, b_2, ... b_d)^T \in R^d$ represents the noise coefficient, $D = [U, E]$ is the super-complete dictionary established in this paper, $c^T = [a \quad b]$ represents the coefficient vector. The sparse solutions of Equation (3) are obtained through solving the minimization problem of l_1:

$$\min \|y - Dc\|_2^2 + \lambda\|c\|_1 . \tag{4}$$

We could obtain the sparse solutions of the coefficient vector $c^T = [a \quad b]$:

$$c^* = \arg\min_c \|y - Dc\|_2^2 + \lambda\|c\|_1 , \tag{5}$$

where: $\| \ \|_1$ and $\| \ \|_2$ represent the norms l_1 of l_2 and, respectively. By solving $c^T = [a \quad b]$ through the equation above, the reconstruction error between the sample and the object template can be defined as:

$$RE = \|y - Ua\|_2^2 . \tag{6}$$

We use RE to evaluate the similarity between the sample and the object template.

2.3 THE UPGRADING OF THE SUBSPACE OF THE INCREMENTAL LEARNING-BASED OBJECT

Assuming that the image of the first n frames $A = [I_1, I_2, ..., I_n]$ with a mean value of $\overline{I_A} = \frac{1}{n}\sum_{i=1}^{n} I_i$ and a centralized matrix of $\overline{A} = [I_1 - \overline{I_A}, ..., I_n - \overline{I_A}]$. Apply SVD (Singular Value Decomposition) to $\overline{A}$ to obtain a unitary matrix U_A and a diagonal matrix Σ_A. Each array of Matrix $\overline{A}$ would be the basic vector of its subspace. Make $B = [I_{n+1}, I_{n+2}, ..., I_{n+m}]$ to be the new image of m frames with a corresponding mean value of $\overline{I_B} = \frac{1}{m}\sum_{i=n+1}^{n+m} I_i$ and $C = [A, B] = [I_1, ... I_n, ... I_{n+m}]$, then what needs to be solved is Matrix C 's unitary matrix U_C and diagonal matrix Σ_C. See below for concrete algorithms:

(1) Calculating the mean value of Matrix C : $\overline{I_c} = \frac{fn}{fn+m}\overline{I_A} + \frac{m}{fn+m}\overline{I_B}$. f , the forgetting factor, is a non-negative number no more than 1;

(2) Calculating B 's augmented central matrix:

$$B^+ = \left[(I_{n+1} - \overline{I_B}), ..., (I_{n+m} - \overline{I_B}), \sqrt{\frac{nm}{n+m}}(\overline{I_B} - \overline{I_A})\right]$$

(3) Calculating $(B^+ - UU^T B^+)$'s orthogonalized matrix B and matrix

$$R = \begin{bmatrix} f\Sigma & U^T B^+ \\ 0 & \tilde{B}(B^+ - UU^T B^+) \end{bmatrix}$$

(4) Apply SVD to R and obtain U_R and Σ_R, then $U_C = \begin{bmatrix} U_A & \tilde{B} \end{bmatrix} U_R$, $\Sigma_C = \Sigma_R$.

3 Particle filter frame

Suppose the status of the object at time t is x_t, the object observation is y_t. According to the state transition probability $p(x_t | x_{t-1})$ and the observation probability $p(y_t | x_t)$, the posterior probability $p(x_t | y_{1:t})$ could be derived in two steps of forecasting and updating:

$$p(x_t | y_{1:t}) = \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} , \tag{7}$$

$$p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t)p(x_{t-1} \mid y_{1:t-1})}{p(x_{t-1} \mid y_{1:t-1})}. \quad (8)$$

Equations (7) and (8) constitute the optimal Bayesian estimation and represent the Particle Filter algorithm required posterior probability density through and the weighted sum of a series of random samples.

Assuming that the particle on-time is $\{x_t^i\}_{i=1}^N$ with the corresponding normalized weight of $\{w_t^i\}_{i=1}^N$, i.e., $\sum_{i=1}^N w_t^i = 1$. The Particle Filter would use to describe the posterior probability:

$$p(x_t \mid y_{1:t}) \approx \sum_{i=1}^N w_t^i \delta(x_t - x_t^i). \quad (9)$$

The updating method of the particle's weight value w_t^i is:

$$w_t^i = w_{t-1}^i \frac{p(y_t \mid x_t^i)p(x_t^i \mid x_{t-1}^i)}{q(x_t^i \mid x_{t-1}^i, y_{1:t})}, \quad (10)$$

where: $\delta(\bullet)$ is a Dirac function, $q(\bullet)$ is an important density function; usually $q(x_t \mid x_{1:t-1}, y_{1:t}) = p(x_t \mid x_{t-1})$ is used as the important density function, when the observation likelihood $p(y_t \mid x_t)$ is the weight value. Then the optimal state of the object could be obtained through a maximum a posteriori estimation:

$$\hat{x}_t = \arg\max p(x_t \mid y_{1:t}). \quad (11)$$

3.1 MOTION MODEL

In the Particle Filter-based tracking frame, the motion model is used to forecast the possible state of the object in two adjacent frames. Usually the object is located by a rectangular box and its rotation, shift and scaling and other moving changes are described with the affine transformation of the rectangular box. This paper defines the state vector of the object as:

$$X_t = [x_t, y_t, \alpha_t, \beta_t, \varphi_t, \gamma_t], \quad (12)$$

where, the 6 parameters are corresponding to 6 affine transformation parameters of the corresponding rectangular in turn, namely the shift on x and y directions, the dimensional changes, the width/height ratio, the rotation angle and the gradient. Supposing that the probability mode of the object's state transition follows Gaussian distribution, i.e.:

$$p(x_t \mid x_{t-1}) = N(X_t; X_{t-1}, \psi), \quad (13)$$

where: $N(\bullet)$ is Gaussian distribution; ψ is the co-variance matrix. In normal status, assuming that various parameters are mutually independent of each other, and ψ is a diagonal matrix whose elements are the variances of each affine parameter $\delta_x^2, \delta_y^2, \delta_\alpha^2, \delta_\beta^2, \delta_\varphi^2, \delta_\gamma^2$.

3.2 THE OBSERVATION MODEL

The observation model describes the similarity between the candidate region of the object and the object model. In Bayesian inference frame, the observation model plays an important role on disposing of unknown status. Using the reconstruction error, the observation likelihood function could be defined as

$$p(y \mid x) = \exp(-RE). \quad (14)$$

From Equation (14) we can see that the smaller the reconstruction error between the sample and the object template is the more the sample's corresponding weight value would be, and the sample would be more reliable.

3.3 WORK PROCEDURES OF THIS PAPER'S TRACKING ALGORITHM

Import: video image sequence $\{F_t\}(t = 1,...,T)$, with the quantity of particles of N.

Export: the position of the object obtained through tracking each frame $\{x_t\}(t = 1,...,T)$.

(1) Initialization: manually select the initial object template and disturb several pixels to generate n templates; use affine transformation to transform the n images with centralized templates into 32×32 window images and apply 2DPCA algorithm to generate the template set's 2DPCA eigenspace U, i.e., the object template set.

(2) Constructing the super-complete dictionary. This paper's super-complete dictionary $D = [U, E]$ is constructed with eigenspaces U and E.

(3) Particle generation. Use the affine transformation matrix obtained through initialization to generate N particles (the affine transformation parameter of the candidate object template) as per Gaussian distribution.

(4) Calculate the sparse representation coefficient. Use minimized l_1 to solve the sparse representation coefficient of the template space corresponding to each particle. Calculate the affine coordinates of the tracking object of the current frame with the coefficient.

(5) Re-sampling of particles. Apply re-sampling to the particle set according to the size of weight values and generate N particles of tracked in the next frame.

(6) Update the object template. Use maximum a posteriori estimation to obtain particle $\hat{x}_t$ with the maximum weight value and reserve the observation sample corresponding to $\hat{x}_t$. For every 5 frames, use the updating algorithm to update the object template.

(7) Export the results. Show the tracking results of the current frame and return to Procedure 3.

4 Simulation test

4.1 SIMULATION ENVIRONMENT

On a PC with Intel double cores and a memory of 2.50GHz and 4G, use MATLAB R2012a to complete the simulation. Select Incremental Visual Tracking (IVT), L1tracker (L1) and Multiple Instance Learning (MIL) methods for comparison with this paper's method. During the concrete tracking test process, the position of the object image in the first frame video is manually decided. The sampling number of particles is 600, with a dimensional number of 12 for the object subspace. The updating frequency of the subspace is set to be 5 frames. The observation image used by the tracking algorithm is window images of 32×32.

4.2 RESULTS AND ANALYSIS

4.2.1 Qualitative analysis

Fig. 1 uses David Indoor video sequence to evaluate the tracking algorithm's performance under lighting and gesture changes. In the David Indoor sequence, the object would experience two times of significant changes of lighting intensity and the changes of gestures caused by removing glasses and wearing glasses. In the whole tracking process, this paper's algorithm is not very sensitive to the changes of lighting and gestures, while L1, IVT and MIL algorithms show different extent of deviations. Of the three algorithms, MIL is the most sensitive to these influences.

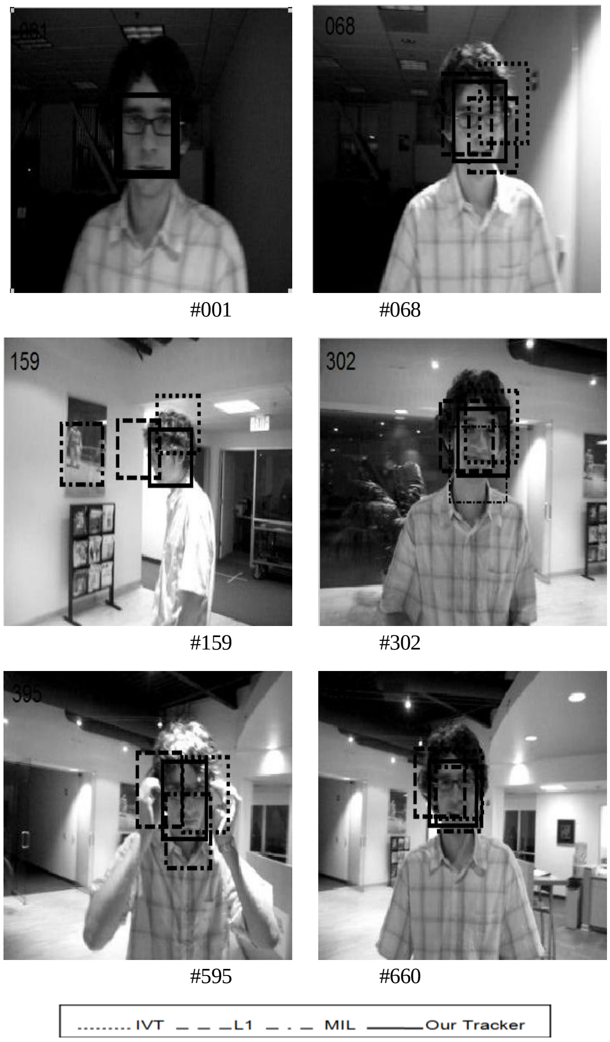


FIGURE 1 Tracking results of each algorithm on the David Indoor sequence

In the Deer video sequence in Fig. 2, this paper's tracking algorithm and the MIL algorithm both show very good performance. Although the MIL algorithm gave a wrong tracking in #052 frame under shielding, correct tracking was soon resumed. LI and IVT algorithms showed poor tracking performance on these fast-moving objects.

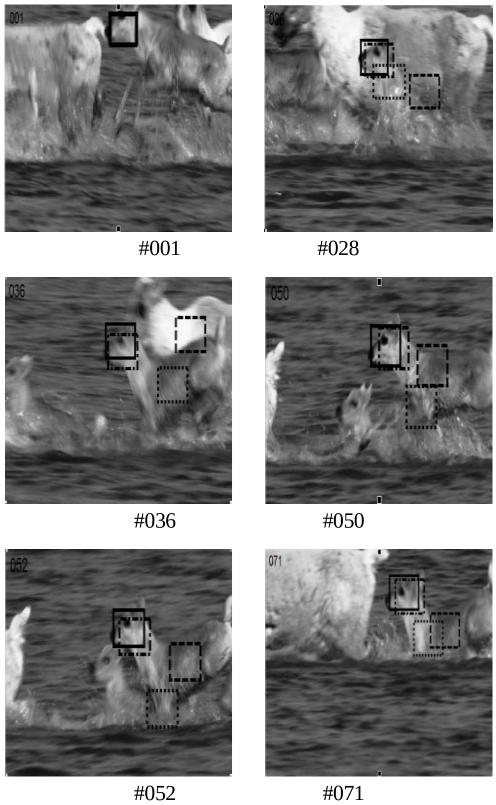


FIGURE 2 Tracking results of each algorithm on Deer sequence

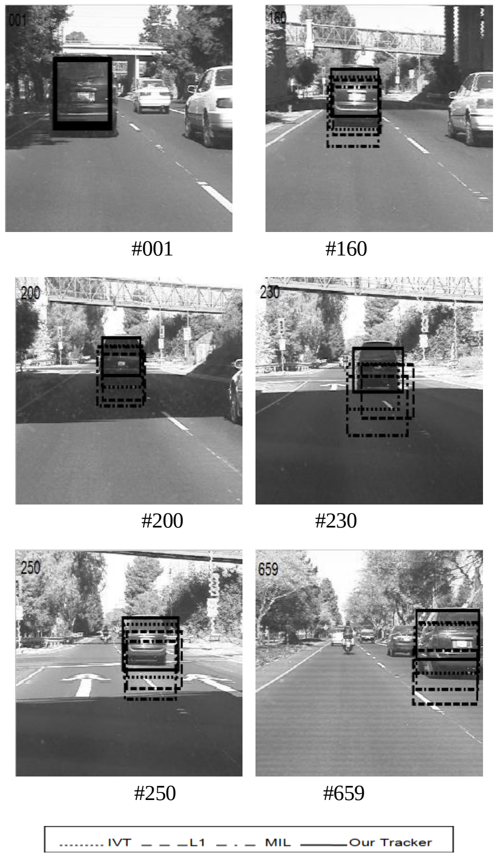


FIGURE 3 Tracking results of each algorithm on Car 4 sequence

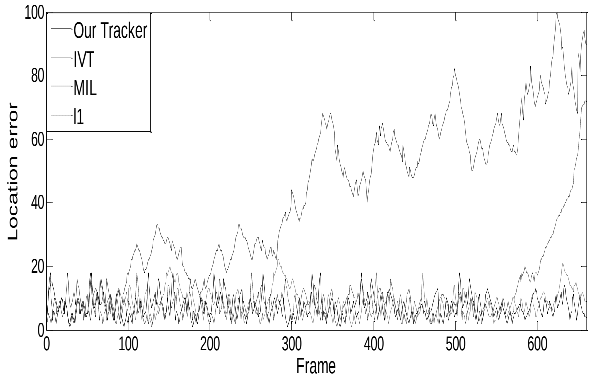
FIGURE 4 Tracking results of different algorithms on Caviar1 sequence

For Car 4 video sequence in Fig. 3, the IVT algorithm and this paper's algorithm showed good performance, while L1 algorithm showed some deviation when the car passed through the bridge. MIL algorithm showed so big deviation that the object was lost.

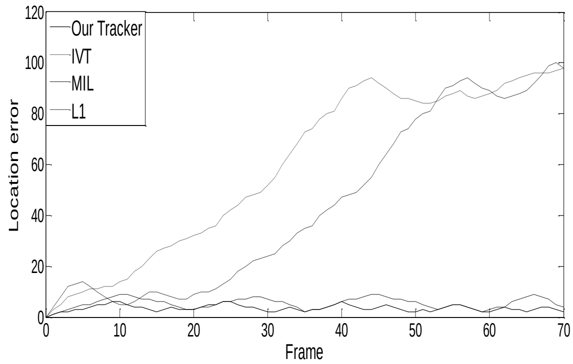
Fig. 4 is the Caviar1 video sequence, showing the women in the monitoring video walking through the corridor. When the shielding of a similar object happened, the MIL algorithm performed poorly, and the L1 and IVT algorithms also presented deviations to different extent, while this paper's algorithm showed quite good performance against partial shielding and the disturbance of similar objects.

4.2.2 Quantitative analysis

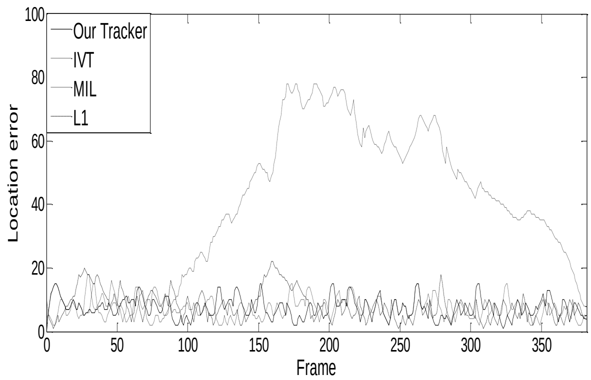
Besides, we used location error - the Euclidean distance between the central location of the tracking result and the test video sequence for the quantitative analysis of the performance of this paper's tracker and the referential tracker, as shown in Fig. 5. The maximum value, mean value and the standard deviation of each tracker are shown in Table 2. Table 2 shows that in the David Indoor video sequence, this paper's tracker and L1 tracker obtained similar optimal results; while for the Deer sequence, the MIL tracker and this paper's tracker showed stable tracking performance upon shielding; compared with other three trackers, this paper's tracker showed the best tracker for Deer sequence and Car4 sequence; in the Caviar1 sequence, this paper's tracker obtained the minimum mean value and standard deviation. Considering the overall performance, this paper's tracker showed the best performance in the tracking process.



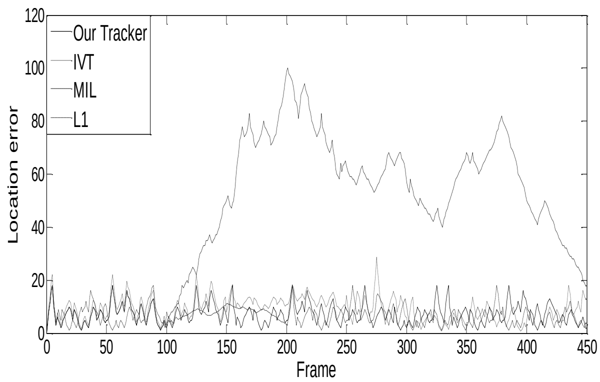
(a) Location error of the David Indoor sequence



(b) Location error of the Deer sequence



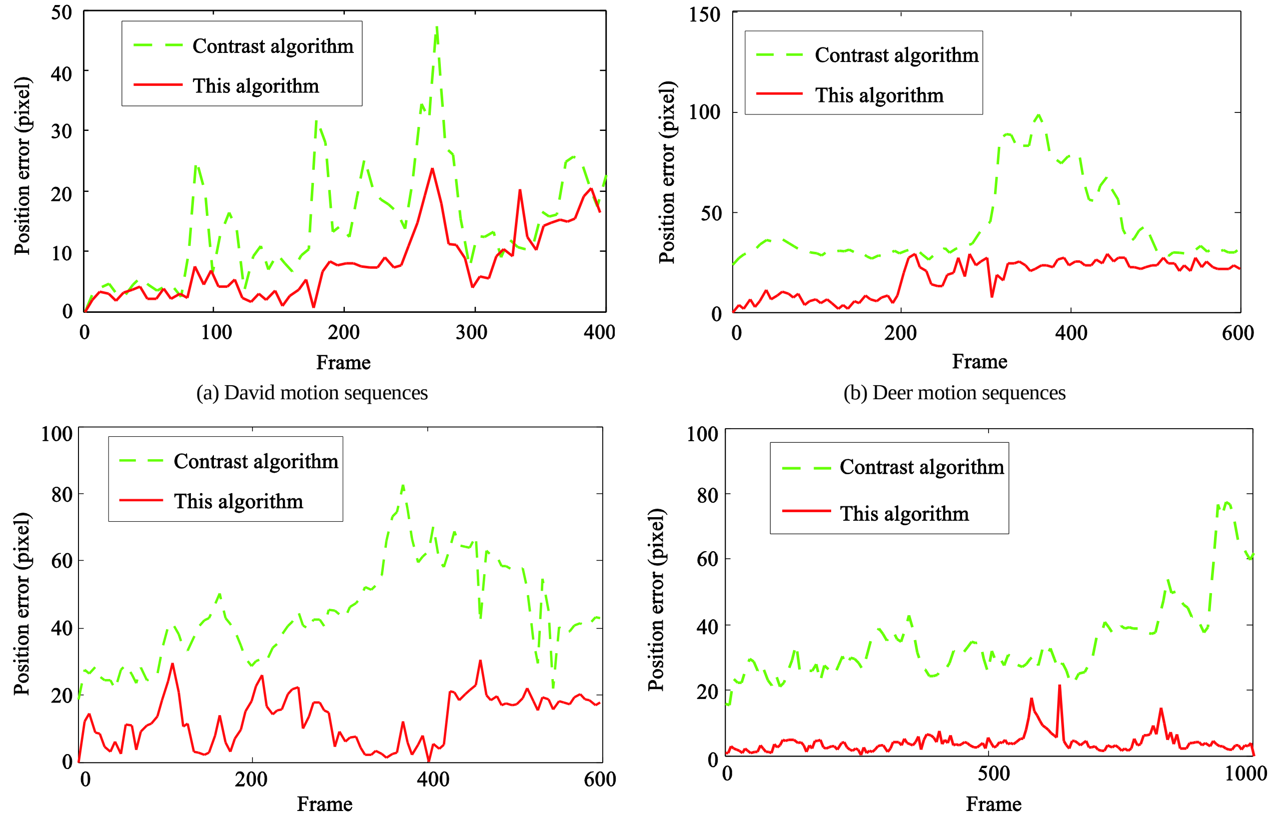
(c) Location error of the Car4 sequence



(d) Location error of the Caviar1 sequence

FIGURE 5 Comparison of tracking errors of the algorithm on video sequences

From Figure 6, we can find that the algorithm in this paper has a better tracking performance compared with the moving target tracking algorithm in the literature [15], especially that it maintains a stable tracking performance when there are occlusions.



(a) David motion sequences (b) Deer motion sequences

(c) Car4 motion sequences (d) Faceocc motion sequences

FIGURE 6 Comparison of tracking errors between this algorithm and the contrast algorithm

TABLE 1 Numerical analysis of location error

		David Indoor	Deer	Car4	Caviar1
	Max	28.55	102.34	22.36	15.20
IVT Tracker	Mean	9.18	59.10	8.39	7.50
	Std	4.45	32.82	4.22	3.16
	Max	98.88	9.24	103.21	78.43
MIL Tracker	Mean	45.25	5.46	42.46	38.61
	Std	27.22	2.09	24.38	21.82
	Max	20.02	97.88	72.44	22.58
L1 Tracker	Mean	6.77	44.01	10.72	8.15
	Std	3.35	34.63	10.91	4.48
	Max	21.23	6.35	18.21	15.32
Our Tracker	Mean	7.30	3.57	7.37	7.48
	Std	3.64	1.28	3.45	3.13

5 Conclusion

This paper proposes the object tracking algorithm of 2DPCA and sparse representation and uses 2DPCA and sparse representation to establish the object appearance model, avoiding the computation of high-dimensional data. Using the incremental subspace learning algorithm, this paper updates the object appearance model in a self-adaptive manner, reducing the algorithm's requirement on the storage space and improving the accuracy of appearance description. Experiment results showed that, compared with IVT, MIL and L1 algorithms, this paper's algorithm could track the moving objects in sequence images and show good robustness to the appearance changes of the object out of lighting or gesture changes in the tracking process. However, this paper's algorithm only uses the images' overall characteristics and fails to settle all of the shielding issues of the object. Therefore, the focus of further studies would be developing more efficient algorithm to describe the object better combining its overall and local characteristics.

Acknowledgement

This work was supported by the Scientific Research Fund of Hunan Provincial Education Department (NO.13C614).

References

[1] Sopian K, Wan Daud W 2006 Challenges and future developments in proton exchange membrane fuel cells *Renewable Energy* **31**(5) 719-27

[2] Yang Han Xuan, Zheng Feng, Wang Liang, et al. 2011 Recent advances and trends in visual tracking: A review *Neurocomputing* **74**(18) 3823-31

[3] Ning Ji Feng, Zhang Lei, Zhang David, et al. 2012 Robust Mean Shift Tracking with Corrected Background-Weighted Histogram *IET Computer Vision* **6**(1) 62-9

[4] Song Kangkang, Chen ken, Guo Yunyan 2013 Mean Shift object tracking algorithm assisted by depth cues *Computer Engineering and*

Applications **49**(23) 177-80
[5] Babenko B, Yang M, Belongie S 2011 Robust Object Tracking with Online Multiple Instance Learning *IEEE Transaction on Pattern Analysis and Machine Intelligence* **33**(8) 1619-32
[6] Yao Zhijun 2013 A New Spatiogram Similarity Measure Method and Its Application to Object Tracking *Journal of Electronics & Information Technology* **35**(7) 1644-9
[7] Leichter I, Lindenbaum M, Rivlin E 2010 Mean shift tracking with multiple references color histograms *Computer Vision and Image Understanding* **114**(3) 400-8
[8] Laura S L, Erik L M 2012 Distribution fields for tracking *IEEE Conference on Computer Vision and Pattern Recognition* 1910-7
[9] Zhang Kai Ua, Zhang Lei, Yang Ming Hsuan 2012 Real-time Compressive Tracking *Proc of European Conference on Computer Vision* 864-77
[10] Zhang Kai Hua, Song Hui Hui 2013 Real-time Visual Tracking via Online Weighted Multiple Instance Learning *Pattern Recognition* **46**(1) 397-411
[11] Li X, Zhang T, Shen X, et al. 2010 Object tracking using an adaptive Kalman filter combined with mean shift *Optical Engineering* **49**(2) 20503-6
[12] Brox T, Rousson M, Deriche R.Colour, texture and motion in level set based segmentation and tracking *Image and Vision Computing* **28**(3) 76-390
[13] Yao Zhi-jun, Liu Jun-tao, Lai Zhong-yuan, et al. 2011 An improved Jensen-Shannon divergence based spatiogram similarity measure for object tracking *Acta Automatica Sinica* **37**(12) 1464-73
[14] Jeyakar J, Venkatesh Babu R, Ramakrishnan K R 2008 Robust object tracking with background-weighted local kernels *Computer Vision and Image Understanding* **112**(3) 296-309
[15] Liu Chao, Hui Jing 2014 Object tracking algorithm based on improved CAMShift *Computer Engineering and Applications* **50**(11) 149-53

Author

Qiu-fen Yang, born in 1973, Hunan Province of China

Current position, grades: associate professor
University studies: received her B.Eng. and M.Eng. degrees degree from Nanjing University, Jiangsu, China and National Defense Science and Technology University, ChangSha, Hunan,China in 1996 and 2004,respectively. She received her Ph.D.degree in the School of software, Central South University, Changsha, Hunan, China in 2014.
Scientific interest: Her current research interests are in the areas of data fusion and computer vision.
Publications: more than twenty papers published in various journals.
Experience: She has teaching experience of 18 years, has completed three scientific research projects.She is now an associate professor at Hunan Radio & TV University of science and technology.

Can-jun Li,born in 1976,Hunan Province of China

Current position, grades: associate professor at Hunan Radio & TV University of science and technology
University studies: received the degree of the B.Eng. (Industrial Automation) and the M.Eng. (Control Science and Engineering) from Zhengzhou University, Henan, China and Central South University, Changsha, China in 2000 and 2007, respectively.
Scientific interest: Her current research interests are in the areas of modelling of complex system.