Robust Visual Tracking via Appearance Modeling and Sparse Representation

Ming Li
School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China
Email: lim3076@163.com

Fanglan Ma and Fuzhong Nian
School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China
Email: mfanglancos @163.com, gdnfz@lut.cn

Abstract—**When appearance variation of object, partial occlusion or illumination change in object images occurs, most existing tracking approaches fail to track the target effectively. To deal with the problem, this paper proposed a robust visual tracking method based on appearance modeling and sparse representation. The proposed method exploits two-dimensional principal component analysis (2DPCA) with sparse representation theory for constructing appearance model. Then tracking is achieved by Bayesian inference framework, in which a particle filter is applied to evaluate the target state sequentially over time. In addition, to make the observation model more robust, the incremental learning algorithm is used to update the template set. Both qualitative and quantitative evaluations on four publicly available benchmark video sequences demonstrate that the proposed visual tracking algorithm performs better than several state-of-the-art algorithms**.

Index Terms—**visual tracking, appearance model, sparse representation, 2DPCA**

I. INTRODUCTION

Visual tracking is very important in the field of computer vision. It is widely used in intelligent surveillance, human-computer interface, vehicle navigation, and traffic control, etc. Although many tracking algorithms have been presented in the past decades, the tracking of the non-stationary appearance of objects undergoing significant pose, illumination variations and occlusions still remains a challenge for the community [1].In general, a tracking algorithm consists of three components[2]: an appearance model which estimates the similarity between observed images and the model; a motion model which aims to locate the target between consecutive frames over time; and a search strategy which finds the most likely state in the current frame. In our work, we focus attention on designing a robust appearance model to achieve reliable tracking.

A good appearance model will obtain good tracking results [3].To achieve an effective appearance models, we should consider several factors which could affect the tracking results. In the process of tracking, target appearance will be changed due to non-rigid deformation, target moving, illumination change or occlusion, etc. In order to adapt to the variations of target appearance, we often employ a linear combination of multiple templates to represent the target appearance. Recently, a class of appearance modeling techniques named sparse representation [4-6] has been shown to give state-of-the-art robustness against various disturbances. The methods based on sparse representation attempt to approximate the target state by finding a sparse linear combination over a basis library that contains target templates and trivial templates.

A variety of online tracking algorithms have been put forward to address these challenges. These methods can be formulated in two different ways: generative model and discriminative model [7].Generative algorithms aim at modeling appearance and finding the minimum reconstruction error [8], [9]. On the contrary, discriminative methods focus on searching a decision boundary which can differentiate the target from the background [10], [11]. It has been shown that discriminative models perform better when the training set size is large, while generative models achieve higher generalization when limited data is available[12].To address the difficulties mentioned above, we employ generative appearance model to alleviate tracking drift.

In this paper, a robust visual tracking algorithm based on appearance model and sparse representation is proposed. In our algorithm, 2DPCA and sparse representation were combined to represent the target appearance which can be reconstructed by the eigenbases and noise bases. Then tracking is achieved by Bayesian inference, in which a particle filter is applied to evaluate

Manuscript received October 30, 2013; revised November 4, 2013; accepted November 15, 2013. This work was supported in part by a grant from the National Natural Science Foundation of China (No.6126 3019), the Fundamental Research Funds for the Gansu Universities (No .1114ZTC144), the Natural Science Foundation of Gansu Province (No. 1112RJZA029) and the Doctoral Foundation of LUT.

Ming Li and Fuzhong Nian are with the School of Computer and Communication, Department of Computer Science. Lanzhou University of Technology, Lanzhou 730050, China (e-mail:lim3076@163.com; gdnfz@lut.cn).

Fanglan Ma is with the School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China (e-mail:mfanglancos@163.com).

doi:10.4304/jcp.9.7.1612-1619

the target state sequentially over time. In addition, to make the observation model more robust, we exploit the incremental learning algorithm to update the template set.

The rest of the paper is organized as follows. Section II introduces the proposed appearance modeling and sparse representation algorithm. The combination of our appearance mode and particle filter for visual tracking is presented in Section III. Section IV demonstrates the experimental results. The paper is concluded in Section V.

II. Appearance Modeling and Sparse Representation

In this section, firstly, an appearance modeling method, which is modeled the target appearance with 2DPCA basis vectors is proposed. Then, the reconstruction error is defined with sparse representation to evaluate the similarity between a target candidate and the appearance model.

A. Appearance Modeling

Here, a straightforward image projection technique, called two-dimensional principal component analysis (2DPCA) is developed for image feature extraction [13].2DPCA is based on 2D matrixes rather than PCA which is needed to be transformed two dimensional images into 1D vectors. In 2DPCA method, image features and covariance matrix can be obtained directly using the image matrixes. 2DPCA is easier to evaluate the covariance matrix accurately and require less time to determine the corresponding eigenvectors than PCA.

Let $Y_t = \{y_1, y_2, \cdots y_n\}$ denotes the observed target sample set at the $t-th$, where $y_i \in R^{d \times d}$ represents the ith sample of the observed target sample set. $I_t = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean of the sample set Y_t. We can use 2DPCA to obtain the eigenbases. The 2DPCA algorithm describes as follow:

(1) Computer the scatter (covariance) matrix S_t of the observed target sample set Y_t:

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

(2) Computer the eigenvector values of the scatter matrix S_t:

$$\lambda_1 \ge \lambda_2 \cdots \ge \lambda_d$$

(3) Computer the eigenvector values of S_t corresponding to the identity orthogonal eigenvectors:

$$u_1,\ u_2, \cdots u_d$$

其中： $u_i^T u_i = 1, u_i^T u_j = 0, i, j = 1, 2, \cdots, d, i \ne j$.

(4) Select the first k largest eigenvectors from all of the eigenvectors of S_t .We can get the eigenvectors：

$$u_1, u_2, \cdots u_k$$

which is the eigenbases $U = \{u_1, u_2, \cdots u_k\}$ in this paper.

B. Sparse Representation

In this section, we use sparse representation framework [14-15] to enhance the discrimination of the appearance model and to construct a robust similarity measure for tracking.

Given the target candidate $Y = \{y_j, j = 1, \cdots\cdots d\} \in R^{d \times d}$, where y_j is the $j-th$ column of Y .We define $\bar{Y}$ is the centralized matrix of Y , $\bar{Y} = Y - I = \{\bar{y}_j, j = 1, \cdots d\}$.

Then $\bar{y}_j$ can be sparsely represented by a linear combination of the eigenbases $U = \{u_1, u_2, \cdots u_k\}$:

$$\bar{y}_j = Ua = a_1u_1 + a_2u_2 + \cdots + a_ku_k \quad (1)$$

where $a = (a_1, a_2, \cdots a_k)^T \in R^k$ is the eigenspace coefficient vector. Then, in order to exactly describe the appearance, we take into account image corruptions or sudden occlusion. Equation (1) should be modified to

$$\bar{y}_j = Ua = a_1u_1 + a_2u_2 + \cdots + a_ku_k + \varepsilon \quad (2)$$

where ε is a error term. The nonzero entries of ε denotes image corruption or occlusion. We can use a noise basis set as $E = [e_1, e_2, \cdots\cdots e_d]^T \in R^{d \times d}$ that is an identity matrix to capture the image corruption or occlusion as:

$$\bar{y}_j = [U\ E]\begin{bmatrix} a \\ b \end{bmatrix} = Bc \quad (3)$$

where $b = (b_1, b_2, \cdots\cdots, b_d)^T \in R^d$ is called the noise coefficient vector. By solving the problem as a l_1-regularized least squares problem [16], we can obtain the sparse solutions of the coefficient vector $c^T = [a\ b]$:

$$c^* = \arg\min_c \left\| \bar{y}_j - Bc \right\|_2^2 + \lambda \left\| c \right\|_1 \quad (4)$$

where $\|\ \|_1$ and $\|\ \|_2$ represent the l_1 and l_2 norm respectively, λ is a parameter to balance reconstruction error and sparsity. With the solution $c^T = [a\ b]$, the reconstruction error of a target candidate and the appearance model can be computed as:

$$RE_j = \left\| \overline{y}_j - Bc^* \right\|_2^2 \quad (5)$$

we use RE to evaluate the similarity between a target candidate and the appearance model.

C. Updating the Appearance

In the target tracking process, the target appearance may change due to pose changes, illumination variations and occlusion, etc. A fixed appearance model fails to deal with varieties of appearance changes [20], [21]. So it is necessary to update the appearance model in time for visual tracking.

Suppose we have an existing image set $A=[I_1, I_2, \cdots I_n]$, $\overline{I}_A = \frac{1}{n}\sum_{i=1}^{n} I_i$ denotes the sample mean of the image set A, $\overline{A}=[I_1-\overline{I}_A, \cdots I_n-\overline{I}_A]$ denote the centered data matrix, where each column $I_i - \overline{I}_A$ is an observation image. Then we can computer the unitary matrix U_A and the diagonal matrix $\sum_A$ from the singular value decomposition (SVD) of $\overline{A}$ .A new image matrix $B=[I_{n+1}, I_{n+2}, \cdots I_{n+m}]$, where $\overline{I}_B = \frac{1}{m}\sum_{i=n+1}^{n+m} I_i$ is the sample mean of the new image matrix B . Let $C=[A,B]=[I_1,\cdots I_n, \cdots I_{n+m}]$, the goal is to efficiently computer the SVD of C : $C = U_C \sum_C V_C^T$. The detail of the algorithm [17] is presented as follow:

Given U_A and $\sum_A$, as well as $\overline{I}_A$ and $\overline{I}_B$, computer $\overline{I}_c$ as well as U_C and $\sum_C$.

(1) Computer the sample mean of the matrix C

$$\overline{I}_c = \frac{n}{n+m}\overline{I}_A + \frac{m}{n+m}\overline{I}_B \quad (6)$$

(2) Computer the augmented(centered) matrix of B :

$$B^+ = [(I_{n+1}-\overline{I}_B), \cdots, (I_{n+m}-\overline{I}_B), \sqrt{\frac{nm}{n+m}}(\overline{I}_B - \overline{I}_A)] \quad (7)$$

(3) $\tilde{B}$ and R can be obtained as follows:

$$\tilde{B} = orth(B^+ - UU^T B^+) \quad (8)$$

$$R = \begin{bmatrix} f\Sigma & U^T B^+ \\ 0 & \tilde{B}(B^+ - UU^T B^+) \end{bmatrix} \quad (9)$$

where $orth()$ denotes the orthogonal operation.

(4) Computer the SVD of R :

$$R \overset{SVD}{=} U_R \Sigma_R V_R^T \quad (10)$$

we can get U_R and $\sum_R$. So

$$U_C = [U_A \ \tilde{B}]U_R \quad (11)$$

$$\Sigma_C = \Sigma_R \quad (12)$$

III. TRACKING BY BAYSIAN INFERENCE

We embed the proposed sparse representation appearance model [22], [23] into a Bayesian inference framework to form a robust tracking algorithm. A Markov model with hidden state variable is used to estimate the tracking result [24]. Given a set of observed images $y_{1:t} = (y_1, \cdots y_t)$ at the $t-th$ frame. Applying Bayes' theorem, we estimate the hidden state variable x_t recursively:

$$p(x_t | y_{1:t}) \propto p(y_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} \quad (13)$$

where $p(y_t | x_t)$ represents the observation model and $p(x_t | x_{t-1})$ denotes the dynamic (motion) model. In the particle filter framework [18], the posterior distribution $p(x_t | y_{1:t})$ is approximated by a set of weighted samples $\{(x_{t-1}^i, w_{t-1}^i)\}_{i=1}^N$, where w_{t-1}^i is the weight for particles x_{t-1}^i and N is the total number of particles. We substitute the integration of the Bayes filter (9) with Monte Carlo approximation

$$p(x_t | y_{1:t}) \approx kp(y_t | x_t) \Sigma w_{t-1}^i p(x_t | x_{t-1}^i) \quad (14)$$

where k is a normalization constant. The candidate samples $\{x_{t-1}^i\}_{i=1}^N$ are drawn from an importance distribution $q(x_t | x_{1:t-1}, y_{1:t})$, and then the weights of the samples are updated as

$$w_t^i = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | x_{t-1}^i)}{q(x_t | x_{1:t-1}, y_{1:t})} \quad (15)$$

The samples are resampled to generate an unweighted particle set according to their importance weights to avoid degeneracy. In this paper, we choose $q(x_t | x_{1:t-1}, y_{1:t}) = p(x_t | x_{t-1}^i)$ as the importance density function and the weights become the observation

TABLE I.
THE PROPOSED TRACKING ALGORITHM

Input: Video frames $F_1, F_2, \cdots, F_t$.

Output: Target states $x_1, x_2, \cdots, x_t$ in frames $F_1, F_2, \cdots, F_t$ respectively

for $t = 1 : FrameNumber$ **do**
if $t = 1$
then
Initialization. Select the target in the first frame manually. Initialize the appearance model using the target observation in the first frame.
else
1. In frame F_t, draw particles $\{x_t^i, i = 1, \cdots N\}$ according to the dynamic model $p(x_t | x_{t-1})$;
2. For each particle x_t^i, calculate the likelihood $p(y_t | x_t^i)$;
3. Estimate the target state x_t using the MAP estimation method and store the target observation y_t simultaneously.
4. Update the appearance model with the new target observation y_t.

end if
end for

likelihood $p(y_t | x_t)$. The optimal state of the target can be obtained by the maximum a posterior (MAP) estimation:

$$\hat{x}_t = \arg\max_{x_t} p(x_t | y_{1:t}) \tag{16}$$

A. Dynamic Model

In our tracking framework, we apply an affine image wrap to model target motion between two consecutive frames .The six parameters of the affine transform are used to model $p(x_t | x_{t-1})$ of a tracked target.

Let $x_t = [\hat{x}_t, \hat{y}_t, \alpha_t, \beta_t, \varphi_t, \gamma_t]$, where $\hat{x}_t, \hat{y}_t, \alpha_t, \beta_t, \varphi_t, \gamma_t$ represent x, y translation, scale, aspect ratio, rotation angle and angle of inclination at time t respectively. Each parameter in x_t is governed by a Gaussian distribution around their previous state x_{t-1} .So $p(x_t | x_{t-1})$ takes the form of:

$$p(x_t | x_{t-1}) = N(x_t; | x_{t-1}, \psi) \tag{17}$$

where ψ is a diagonal covariance matrix whose elements are the corresponding variances of the affine transformation parameters, i.e., $\delta_{\hat{x}}^2, \delta_{\hat{y}}^2, \delta_{\alpha}^2, \delta_{\beta}^2, \delta_{\varphi}^2, \delta_{\gamma}^2$.

B. Observation Model

In the Bayesian inference framework, the observation model plays an important role in handling the unpredictable changes. Using the reconstruction error, the observation likelihood can be define as

TABLE II.
IMAGE SEQUENCES USED IN OUR EXPERIMENTS

Image sequences	#Frames	Challenging factors
David Indoor	898	illumination variation scale change, out-plane rotation
Deer	71	abrupt motion, background clutter
Car4	659	illumination variation ,scale change
OneLeaveShopReenter2cor	500	partial occlusion, scale change

$$p(y | x) = \exp(-RE) \tag{18}$$

It is evident that the reconstruction error between a target candidate and the appearance is smaller, the observation likelihood is greater and the candidate is more reliable. TABLE I gives a summary of our tracking algorithm.

IV. EXPERIMENTAL RESULTS AND ANAYSES

To evaluate the performance of the proposed algorithm, we run our tracking algorithm on four publicly available benchmark video sequences, which included illumination changes, significant occlusion, and pose variance, and so on. The main challenging factors of these image sequences are showed in TABLE II. The experiments are implemented in MATLAB R2012a on an Intel 2.5GHz Dual Core PC with 4GB memory. For each sequence, the initial position of the target is selected manually in the first frame. For 2DPCA representation, the size of each image is normalized to 32×32, 16 eigenvectors and 1024 noise bases are used in all experiments. 600 particles are used and the template set is updated in every 5 frames to balance the tracking accuracy and the computational efficiency. Additionally, we compare our tracking algorithm with Incremental Visual Tracking (IVT) tracker [9], L1 tracker (L1) [19] and Multiple Instance Learning (MIL) [10]. Both qualitative and quantitative evaluations are demonstrated as follow.

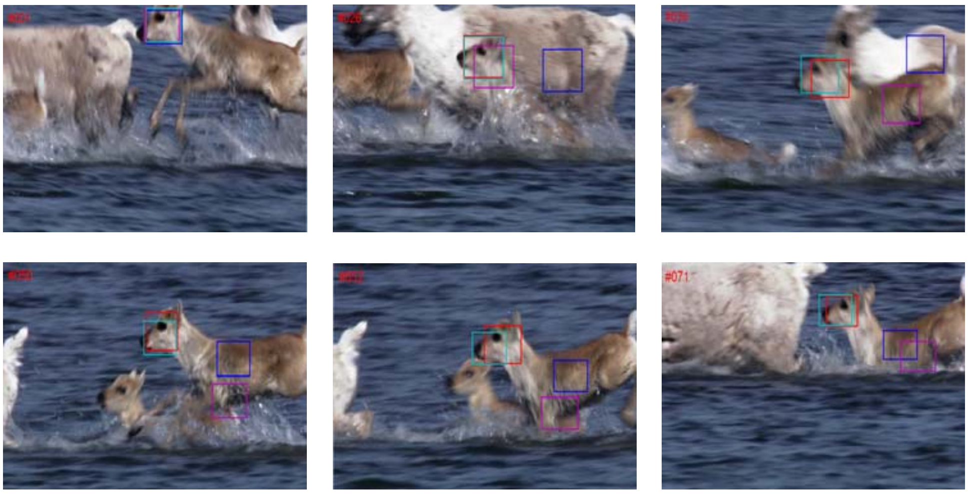
A. Qualitative Evaluation

The David Indoor sequence is used to evaluate the performance of these four tracking algorithms when the target object undergoes significant illumination change and pose variation. Tracking results are show in Fig.1 (a).We noted that the IVT and our tracker perform better than the others. MIL is sensitive to illumination change and gets lost in tracking the target.

Fig.1 (b) shows the tracking results using the Deer sequence which exhibits challenges on abrupt motion and occlusion. The MIL and our tracker perform well in tracking the deer whereas the L1 and IVT methods drift away in early stage.

The Car4 sequence is used to test the performance of our tracking algorithm in drastic illumination changes. The IVT and Our tracking perform better than other trackers. The L1 tracker drifts away when the car goes

(a. Tracking results of the David Indoor sequence)



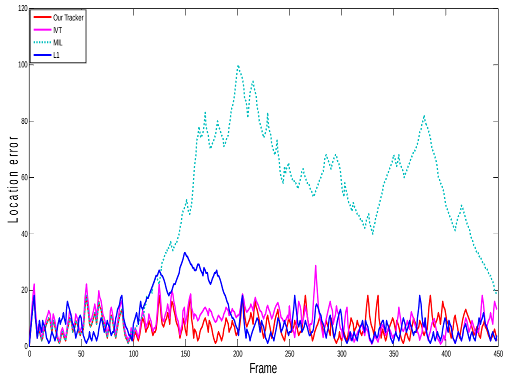
(b. Tracking results of the Deer sequence)

(c. Tracking results of the Car4 sequence)

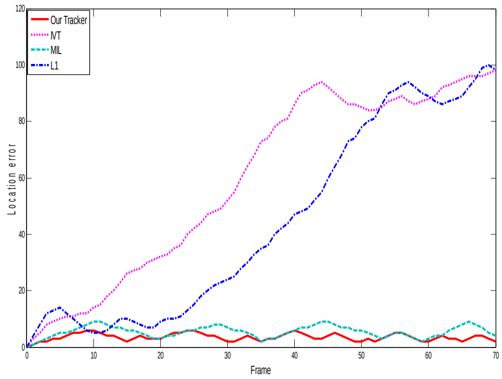
(d. Tracking results of the OneLeaveShopReenter2cor sequence)

IVT L1 MIL Our Tracker

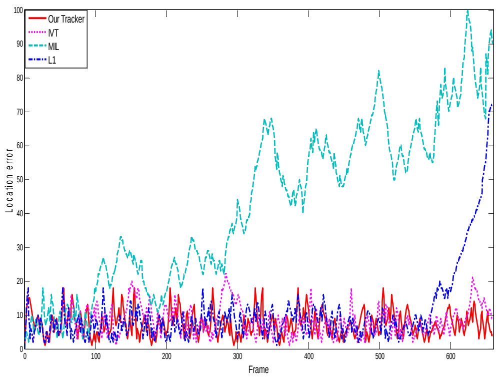
Figure1. Qualitative evaluation



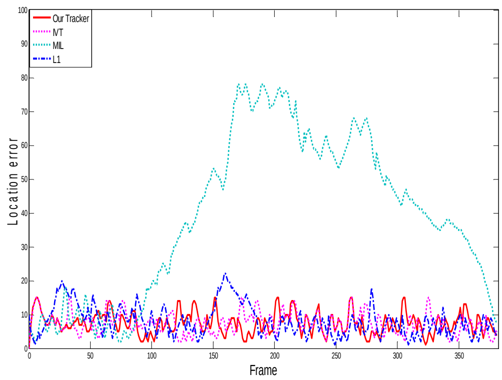
a. The David Indoor sequence



b. The Deer sequence



c. The Car4 sequence



d. The OneLeaveShopReenter2cor sequence

Figure2. Quantitative evaluation

TABLE III.
ANALYSIS OF LOCATION ERRORS

	IVT Tracker			MIL Tracker			L1 Tracker			Our Tracker		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
David Indoor	28.55	9.18	4.45	**98.88**	45.25	27.22	20.02	**6.77**	**3.35**	21.23	7.30	3.64
Deer	102.34	59.10	32.82	9.24	5.46	2.09	97.88	44.01	34.63	**6.35**	**3.57**	**1.28**
Car4	22.36	8.39	4.22	103.21	42.46	24.38	72.44	10.72	10.91	**18.21**	**7.37**	**3.45**
OneLeaveShopReenter2cor	**15.20**	7.50	3.16	78.43	38.61	21.82	22.58	8.15	4.48	15.32	**7.48**	**3.13**

underneath the overpass and the trees. We note that the MIL is sensitive to the effects and misses the target in later stage.

For the OneLeaveShopReenter2cor sequence, it evaluates the robustness of the trackers when scale change occurs, when particle occlusion occurs or when similar objects occurs. The experimental results show that the MIL tracker drifts away when the target is occluded by a similar object. But after occlusion, all of the tracking algorithms track the object accurately again.

B. Quantitative Evaluation

We use the location error that measures the Euclidean distance between the tracking window center and the ground truth to quantify the performance of our tracker and the reference trackers. The location error is defined as follows:

$$error = \sqrt{(x' - x)^2 + (y' - y)^2} \quad (19)$$

where (x', y') is the object position and (x, y) represents the ground truth. The maximum, mean and standard deviation of the location error are given in TBALE III. The values with underline show the best results. From the TBALE III, we can learn that our tracker and L1 tracker obtain the very close results in the David Indoor sequence. For the Deer sequence, our tracker and the MIL tracker can remain stably track the deer when occlusion appears. In the Deer and Car4 sequences, our tracker obtains the best results among four of trackers. In the OneLeaveShopReenter2cor sequence, our tracker has the lowest mean and standard deviation of the location errors. Taking into account overall performance, our tracker has the best effects in tracking process.

V. Conclusion

In this paper, the 2DPCA basis vectors are applied to represent images, and the sparse representation is used to linearly express the target. 2DPCA and sparse representation make the appearance model more robust to appearance and illumination variance. In order to reduce the storage space and improve the accuracy of appearance description, incremental learning algorithm is used for updating the target template set. Experiments on four publicly available benchmark video sequences demonstrate that our algorithm performs better accuracy and robustness than several state-of-the-art algorithms. However, in the matter of solving whole object occlusion problem, our algorithm performance is not very efficient. Our tracking algorithm only employs the global feature and ignores the local cues, so we will integrate multiple features to better describe the objects and explore more efficient algorithms in the future.

Acknowledgment

This work was supported in part by a grant from the National Natural Science Foundation of China (No.6 1263019), the Fundamental Research Funds for the Gans u Universities (No.1114ZTC144), the Natural Science Fo undation of Gansu Province (No.1112RJZA029) and the Doctoral Foundation of LUT.

References

[1] Li W G , Hou Y, Lou H D, Ye G Q. "Robust visual tracking based on Gabor feature and sparse representation",*2012 IEEE International Conference on Robotics and Biomimetics*. pp. 1829-1835, 2012.

[2] Wang D, Lu H C, Yang M H. "Online Object Tracking with Sparse Prototypes", *IEEE transactions on image processing*.vol.22, pp.314-325, 2013.

[3] Nejhum, Shahed, Muhammad Rushdi, and Jeffrey Ho. "Visual tracking using superpixel-based appearance model", *Computer Vision Systems*. *Springer Berlin Heidelberg*. pp. 213-222, 2013.

[4] Ming-Hsuan, and Jeffrey Ho. "Toward robust online visual tracking", *Distributed Video Sensor Networks*. pp. 119-136, 2011.

[5] Parate, Priti. "Fragment-based real-time object tracking: A sparse representation approach." *2012 19th IEEE International Conference on Image Processing (ICIP). IEEE.* pp. 433-436, 2012.

[6] Han Z, Jiao J, Zhang B, et al. "Visual object tracking via sample-based adaptive sparse representation (AdaSR)", *Pattern Recognition*. vol. 44, pp. 2170-2183, 2011.

[7] Yuan Xie, Wensheng Zhang, et al. "discriminative object tracking via sparse representation and online dictionary learning", *IEEE transactions on cybernetics*.pp.2168-2267, 2013.

[8] Sun, Tongfeng, Shifei Ding, and Zihui Ren. "Novel Image Recognition Based on Subspace and SIFT", Journal of Software. vol. 8, pp.1109-1116, 2013.

[9] Ross D A, Lim J, Lin R S, et al. "Incremental learning for robust visual tracking", *International Journal of Computer Vision*. vol. 77, pp.125-141, 2008.

[10] Babenko B, Yang M H, Belongie S. "Visual tracking with online multiple instance learning", *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE*. pp. 983-990, 2009.

[11] Vo, Nam, et al. "Robust visual tracking using randomized forest and online appearance model", *Intelligent Information and Database Systems.* pp. 212-221, 2011.

[12] Salzmann, Mathieu, and Raquel Urtasun. "Combining discriminative and generative methods for 3D deformable surface and articulated pose reconstruction", *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.pp.647-654, 2010.

[13] Jemaa, Yousra Ben, Ahmed Derbel, and Ahmed Ben Jmaa. "2DPCA fractal features and genetic algorithm for efficient face representation and recognition", *EURASIP Journal on Information Security*.vol.1, pp. 1-9, 2011.

[14] Wright J, Ma Y, Mairal J, et al. "Sparse representation for computer vision and pattern recognition", *Proceedings of the IEEE*. vol. 98, pp. 1031-1044, 2010.

[15] Wright J, Yang A Y, Ganesh A, et al. "Robust face recognition via sparse representation ", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. vol. 31, pp. 210-227, 2009.

[16] Mei X, Ling H, Wu Y, et al. "Minimum error bounded efficient ℓ 1 tracker with occlusion detection"，*Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE*. pp. 1257-1264, 2011.

[17] Dagher, Issam. "Incremental PCA-LDA algorithm"，*2010 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*.pp. 97-101, 2010.

[18] Liubai Li. "Efficient Fast Object-Tracking Scheme Based on Motion-vector-located Pattern Match", *Journal of Software*. vol. 7, pp. 998-1005, 2012.

[19] Mei X, Ling H. "Robust visual tracking using ℓ 1 minimization", *Computer Vision, 2009 IEEE 12th International Conference on. IEEE*. pp. 1436-1443, 2009.

[20] Kwon, J., Lee, K. M.. "Visual tracking decomposition", *In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. pp. 1269-1276, June, 2010.

[21] Lee H, Battle A, Raina R, et al. "Efficient sparse coding algorithms", *Advances in neural information processing systems*. pp. 801-808, 2006.
[22] Yang, Ehwa, Moongu Jeon, and Vladimir Shin. "Robust auxiliary particle filter with an adaptive appearance model for visual tracking", *Computer Vision–ACCV 2010. Springer Berlin Heidelberg*. pp. 718-731, 2010.
[23] Dinh T B, Medioni G. "Co-training framework of generative and discriminative trackers with partial occlusion handling", *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. pp. 642-649, 2011.
[24] Wang Y, Wang X, Wan W. "Object Tracking with Sparse Representation and Annealed Particle Filter", *Image and Graphics (ICIG), 2013 Seventh International Conference on. IEEE*, pp. 374-379, 2013.

Ming Li was born in Lanzhou, Gansu Province, China in 1959. He received the bachelor degree in mathematics at Xi'an University of Technology in 1982. Now he is a professor in the School of Computer and Communication at Lanzhou University of Technology, Lanzhou, China. His current research interests are intelligent information processing, pattern recognition, face analysis and object tracking. In recent years, he has authored about 30 papers in international journals, National issues and international conference proceedings.

Prof. Li has served as a reviewer for Journal of Lanzhou University of Technology.

Fanglan Ma was born in Tianshui, Gansu Province, China in 1988. She received the bachelor degree in electronic information engineering at Lanzhou University of Finance and Economics in 2011; she has taken up the subject of Master's Vision Group. Her research interests include intelligent information processing, object tracking and signal processing.

Fuzhong Nian was born in Wuwei, Gansu Province, China in 1974. He received the Ph.D. in electronic information engineering at Dalian University of Technology in 2011. Now he is an Associate professor in the School of Computer and Communication at Lanzhou University of Technology, Lanzhou, China. His current research interests are complex system and complex network, intelligent information processing. In recent years, he has authored about 30 papers in international journals, National issues and international conference proceedings.