An Improved Algorithm for Moving Object Tracking

Qiufen YANG^{1,a}

¹ Department of Computer Science, Hunan Radio &TV University, Changsha, China ^ayqf5569@sohu.com

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 proposes a robust visual tracking method based on sparse representation. To make the observation model more robust and reduce the requirement of storage space, the incremental learning algorithm is used to update the target templates. Experimental results demonstrate that in case of object deformation, partial occlusion and illumination change, the proposed visual tracking algorithm can still maintain better tracking accuracy and robustness

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

Introduction

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

Current object tracking methods mainly include: template matching-based theory-based tracking method, filtering tracking method[2] and classification-based tracking method[3]. The template matching-based method has the advantages of being simple and of high matching accuracy, but is sensitive to shielding and deformation. The filtering theory-based tracking method includes two major types, Kalman Filter[4] and Particle Filter[5]. 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 method is to consider object tracking as a binary classification problem, 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 a good choice if high real-time capability is required. 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. 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 sparse representation to establish the model of object appearance, reducing the number of dimensions of the object template. By 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.

Sparse representation-based object tracking

A. Sparse representation

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

$$\mathbf{y} \approx U \mathbf{a} = \mathbf{a}_1 \mathbf{u}_1 + \mathbf{a}_2 \mathbf{u}_2 + \dots + \mathbf{a}_k \mathbf{u}_k \quad (1)$$

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

 $y = \mathbf{U}\mathbf{a} + \boldsymbol{\varepsilon} \tag{2}$

Where: $\mathcal{E} \in \mathbb{R}^d$ represents the error term introduced by noises and shielding, \mathcal{E} is non-zero elements denoting the noises or shielding of the object. We could use a unit $E = [\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_d]^T \in \mathbb{R}^{d \times d}$ matrix to locate the position of disturbance. So Equation (2) could be rewritten as:

$$y = \begin{bmatrix} U, E \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = Dc$$
(3)

Where, $b = (b_1, b_2, \dots b_d)^T \in \mathbb{R}^d$ represents the noise coefficient, $D = [U_1, E]$ is the super-complete dictionary established in this paper, $c^{T} = [a \ b]$ represents the coefficient vector. The sparse solutions of Equation (3) are obtained through solving the minimization problem of l_{1} :

 $\min \| y - Dc \|_{2}^{2} + \lambda \| c \|_{1}$ (4)

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

$$c * = \arg || y - Dc ||_{2}^{2} + \lambda || c ||_{1}$$
 (5)

Where: $||y - Dc||_2^2 + \lambda ||c||_1$, $|||_1$ and $|||_2$ represent the norms of l_1 and l_2 , respectively. By solving $c^T = [a \ b]$ through the equation above, the reconstruction error between the sample and the object template can be defined as:

$$RE = \parallel y - Ua \parallel_2^2 \tag{6}$$

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

C.The upgrading of the subspace of the incremental learning-based object

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

Calculating the mean value of Matrix $C:\overline{I_c} = \frac{fn}{fn+m}\overline{I_A} + \frac{fn}{fn+m}\overline{I_B}, f$,

the forgetting factor, is a non-negative number no more than 1;Calculating B's augmented central matrix:

$$B^{+} = [(\mathbf{I}_{n+1} - \overline{I}_{B}), \cdots, (\mathbf{I}_{n+m} - \overline{I}_{B}), \sqrt{\frac{nm}{n+m}}(\overline{I}_{B} - \overline{I}_{A})]$$
;Calculating

$$\begin{split} B^{+} &-UU^{T}B^{+} \text{ 's orthogonalized matrix } \tilde{B} \text{ and matrix} \\ R &= \begin{bmatrix} f\Sigma & U^{T}B^{+} \\ 0 & \tilde{B}(B^{+} - UU^{T}B^{+} \end{bmatrix}. \\ \text{Apply SVD to } R \text{ and obtain } U_{R} \text{ and } \Sigma_{R} \text{ , then} \\ U_{C} &= [\mathbf{U}_{A} \tilde{B}]\mathbf{U}_{R}, \Sigma_{C} = \Sigma_{R}. \end{split}$$

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 $p(X_t | y_{1:t})$ probability could be derived in two steps of forecasting and updating:

$$\mathbf{p}(\mathbf{X}_{t} | \mathbf{y}_{1:t}) = \int \mathbf{p}(\mathbf{X}_{t} | \mathbf{X}_{t-1}) \mathbf{p}(\mathbf{X}_{t-1} | \mathbf{y}_{1:t-1}) \, \mathrm{d}\mathbf{x}_{t-1}$$
(7)

$$\mathbf{p}(\mathbf{X}_{t} | \mathbf{y}_{1:t}) = \frac{\mathbf{p}(\mathbf{y}_{t} | \mathbf{X}_{t}) \mathbf{p}(\mathbf{X}_{t-1} | \mathbf{y}_{1:t-1})}{\mathbf{p}(\mathbf{y}_{t} | \mathbf{y}_{1:t-1})}$$
(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 at time t is $\{\mathbf{x}_{t}^{i}\}_{i=1}^{N}$ with the corresponding normalized weight of $\{\boldsymbol{\omega}_{t}^{i}\}_{i=1}^{N}$, i.e., $\sum_{i=1}^{N} \boldsymbol{\omega} = 1$. The Particle Filter would use

 $\{\mathbf{X}_{t}^{i}, \boldsymbol{\omega}_{t}^{i}\}_{i=1}^{N}$ to describe the posterior probability:

$$\mathbf{p}(\mathbf{X}_{t} | \mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} \omega_{t}^{i} \delta(\mathbf{x}_{t} - \mathbf{x}_{t}^{i})$$
(9)

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

$$\omega_t^i = \omega_{t-1}^i \frac{\mathbf{p}(\mathbf{y}_t + \mathbf{x}_t^i) \mathbf{p}(\mathbf{x}_t^i + \omega_{t-1}^i)}{\mathbf{q}(\mathbf{x}_t^i + \mathbf{x}_{t-1}^i, \mathbf{y}_{1:t})}$$
(10)

Where: $\delta(\bullet)$ is a Dirac function, $q(\bullet)$ is an important density function; usually $q(\mathbf{x}_t | \mathbf{x}_{t_{t-1}}, \mathbf{y}_{t_t}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$

is used as the important density function, when the observation $p(y_t | x_t)$ likelihood is the weight value. Then the optimal state of the object could be obtained through a maximum a posteriori estimation:

$$x_{t} = \operatorname{argmaxp}(\mathbf{x}_{t} | \mathbf{y}_{1:t})$$
(11)

Results and analysis

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.1.



(a) Location error of the David Indoor sequence



(b) Location error of the Deer sequence



(d) Location error of the Caviar1 sequence Figure.1 Comparison of tracking errors of the algorithm on video sequences

Summary

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 because of changes of lighting or gesture in the tracking process. However, this paper's algorithm only uses the images' overall characteristics and fails to settle all the shielding issues of the object. Therefore, the focus of further researches would be developing more efficient algorithm to describe the object better combining its overall and local characteristics.

References

[1]Tsagkatakis G, Savakis A, Online distance metric learning for object tracking[J]. IEEE Trans on Circuits and Systems for Video Technology, 2011, 21(12):1810-1821.

[2]Yim J, Jeong S, Gwon K, et al. Improvement of kalman filters for WLAN based indoor tracking [J]. Expert systems with Applications, 2010, 37(1):426-433.

[3]Wang Z, Yang X, Xu Y, et al. Camshift guided particle filter for visual tracking [J]. Pattern Recognition Letters, 2009, 30(4): 407-413.

[4]Xie Y, Zhang W, Li C, et al. Discriminative object tracking via sparse representation and online dictionary learning [J]. IEEE transactions on cybernetics, 2013.

[5]Mei X, Ling H. Robust visual tracking using ℓ 1 minimization[C]// Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 1436-1443.Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, 26(1): 131-137.

[6]Levey A, Lindenbaum M. Sequential Karhunen-Loeve basis extraction and its application to images [J]. Image Processing, IEEE Transactions on, 2000, 9(8): 1371-1374.