Acta Technica *62* (2017), No. 5A, 757–766

Visual object tracking based on random compression of multi-channel haar-like feature

Jiashu Dai[2,4], Nan Yan[2], Tao Liu[2], Jun Wang[2], Tingquan Deng[3]

Abstract. The compressed sensing algorithm has achieved good application effect in the field of visual tracking. In this paper, a moving object tracking algorithm based on the random compression of multi-channel Haar-like features is proposed. Firstly, a new multi-channel Haar-like feature is introduced by considered the color distribution of the object, and extract it by random compression projection. Then, the characteristics of the feature distribution in the samples are studied, and a weighted naive Bayesian classifier is constructed. Finally, in order to adapt to the change of the object appearance, the classifier parameters are updated in real time. The effectiveness of the algorithm is verified by contrast experiments which carried on the open and challenging video sequences.

Key words. Visual tracking, bayesian classifier, random compression, multi-channel, haar-like.

1. Introduction

Visual object tracking is an important research area in computer vision and pattern recognition which aims to tracking an object accurately in real surveillance scene. The main challenge of visual object tracking is how to adaptive to the changes of object appearance. To address this problem, recent state-of-art approaches focus

[1] Acknowledgement - This research was financially supported by National Natural Science Foundation of China (No.61672386), Anhui Province Higher Education Enhance Plan Natural Science Research Project (No.TSKJ2016B04), the Program of Educational Commission of Anhui Province(No.KJ2017A104), Science and Technology Projects of Production (No.2015cxy03), Scientific Research Starting Foundation of Anhui Polytechnic University (No. 2015YQQ011), and Foundation of Key Laboratory of Computer Application Technology (No. JSJKF201603).

[2] Workshop 1 - School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China

[3] Workshop 2 - Laboratory of Fuzzy Information Analysis and Intelligent Recognition, Harbin Engineering University, Harbin, 150001, China

[4] Corresponding author: Jiashu Dai; e-mail: daijiashudai@ahpu.edu.cn

on different appearance models which are generative methods and discriminative methods. The VOT (Workshop on Visual Object Tracking Challenge) program providing a platform for discussing visual tracking problems [1].

The generative methods obtain the state of the tracking object by searching the best matching region of the object's appearance model in the sequence image. In the generative tracking algorithms, the object's appearance model is only learned from the object region in the image which will ignore the background information in the scene. In this kind of algorithm, most of the appearance models of the object are based on the template and subspace.Template matching tracking algorithm is a kind of generative tracking algorithm. In order to adapt to the change of the object appearance, Fussenegger et al. [2] proposed a level set based tracking algorithm that builds the shape model incrementally from new aspects obtained by segmentation or tracking. Chaudhry et al. [3] model the temporal evolution of the object's appearance/motion using a linear dynamical system. Then, this model is learned from sample videos and is used as dynamic templates for visual tracking. Hu et al. [4] proposed an incremental tensor subspace tracking algorithm which models appearance changes by incrementally learning a tensor subspace representation. The moving object tracking problem is considered as sparse approximation problem in [5]. The appearance of the object is linearly composed of the object template and the noise template. The sparse representation is obtained by solving the L1 regularization problem, and finally the minimum reconstruction error area is the object region.

The discriminant methods view the moving object tracking as a two classification problem. It searches for the optimal discrimination function between the object and the background, and tries to distinguish the object accurately from the background. Unlike the generative tracking method, the discriminant tracking method learns both the information of the foreground and background, and constructs a classifier that can separate the object from the background. At the same time, the discriminant tracking method also attempts to construct the most discriminative object features.Collins et al. [6] proposed an online selection discriminant tracking feature which assuming that the best distinguishing feature between object and background is the best feature for tracking. Guo et al. [7] proposed a maximum trust boosting algorithm for moving object tracking. Yang et al. [8] proposed a multiple kernel boosting tracking framework. In order to alleviate the "offset" problem, Babenko et al. [9] proposed a stable multi-instance moving object tracking algorithm (MIL algorithm) which instead the supervised boosting learning by online multi instance learning. In the MIL algorithm, the training samples in the sample instance bag have the same weight, while in the actual situation, the importance of the different samples should be different. To solve this problem, Zhang et al. [10] propose a weighted multiple instance tracking algorithm (WMIL algorithm). Feng et al. [11] construct a sparse tracking algorithm which combined the context information. Hu et al. regard image blocks as the two-order tensors, graph embedding is used to keep graphic structure.

With the rapid development of deep learning theory, great progress has been made in the field of visual tracking.Excellent tracking performance is achieved by

using the features which are extracted automatically from the multi-layer nonlinear transformation. The convolution neural network is able to simulate the appearance model in the tracking task because of its excellent ability in feature extraction and image classification.Recurrent neural networks have achieved certain effects when dealing with partial occlusion.

How to extract the object features and how to construct the classifier are the most core problems of the moving object tracking under the discriminant framework.Dr. Zhang proposed a compressed sensing tracking algorithm which creatively randomly compressed the high-dimensional features into low-dimensional subspace. Because of its simplicity and efficiency,it has gained the attention of many scholars. In order to solve the uncertainty of random projection in the compression tracking algorithm, Gao proposed an tracking algorithm based on the maximum stable extremum region (MSER).

In this paper, a new visual object tracking algorithm is proposed based on the random compression of Multi-channel Features. The algorithm is designed from two aspects: the structure of the feature and the construction of the classifier.By the contrast experiments, this algorithm is more effective than the original compressed sensing tracking algorithm (CT).

2. Random Compression Projection of Multi-channel Features

Because of the Haar wavelet function can extract the difference of the average gray level in different regions, the structure information of the object can be extracted by using structure which is similar to the Haar wavelet. There are many methods to construct the Haar-like features, such as follows.

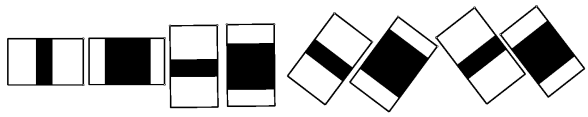

Fig. 1. TheHaar-like features

According to a series of rectangle templates that have been constructed, the Haar-like feature value f is calculated in the image by following formula:

$$f = \left| a \sum_{(x,y)\in S_B} I(x,y) - b \sum_{(x,y)\in S_W} I(x,y) \right| \tag{1}$$

Where S_B, S_W are respectively represents the black and white areas in the rectangular template, a, b are the weighted coefficient of rectangle box. $I(x,y)$is the gray value of the pixel point (x,y).

In the multi-channel video sequences, the color information of object or background is very rich, and there is a certain difference between the color distributions of different objects. At the same time, in most of the scenes, the color information

between the object and the background is also very different. As shown in Figure 2, the main color distribution of "Tiger" object is "white", "orange" and "black", while the green plants in the background information are more prominent. The multi-channel Haar-like is constructed, where the rectangle box information is constructed from the same rectangle box information in each color channel image. The new multi-channel rectangle box information is fused by the weight of the object color distribution, then the multi-channel Haar-like feature is calculated by random compression projection.

Fig. 2. The multi-channel Haar-like feature

According to the multi-channel histogram of the appearance multi-channel image of the object which is to be tracked, the number of color values in each color channel histogram which is larger than the average is counted asn_R, n_G, n_B. The pixel value in the rectangle box x_i in the multi-channel Haar-like feature is

$$x_i = \lambda_R x_{Ri} + \lambda_G x_{Gi} + \lambda_B x_{Bi} \tag{2}$$

Where x_{Ri}, x_{Gi}, x_{Bi} are the sum of pixel value between the rectangle boxes in the R, G, B color channel and $\lambda_R, \lambda_G, \lambda_B$ are the correlation coefficient of the R, G, B color channel, and the value is

$$\lambda_R = \frac{n_R}{n_R + n_G + n_B}, \lambda_G = \frac{n_G}{n_R + n_G + n_B}, \lambda_B = \frac{n_B}{n_R + n_G + n_B} \tag{3}$$

As an efficient method of dimensionality reduction, random projection can compress high-dimensional data into low-dimensional subspace and approximate maintain the structural relationship between data. In classical compression tracking algorithm, the author Dr. Zhang has been successfully applied it to gray image feature compression.

By Johnson and Lindenstrauss lemma, we can see that any group of data in high-dimensional subspace can be compressed into $O\left(\varepsilon^{-2} \log n\right)$-dimension subspace by map f, and the distance between data points in compressed subspace can be approximately unchanged.

The number of Multi-channel Haar-like features is large and can reach to m = (w $\times$ h)^2 dimensions. In order to reduce the computational complexity and ensure the validity of the selected classifier features, a small number of discriminative features are selected by random projection. Through the random compression projection, the m-dimensional feature $(x_1, x_2 \cdots x_m)^T$ can be compressed into a n-dimensional (n << m) feature $v = (v_1, v_2 \cdots v_n)^T$ by an n $\times$ m random projection matrix R,

that is

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \tag{4}$$

The random projection matrix R should satisfy the condition of restricted isometry property (RIP) which is subject to a Gaussian distribution with a mean of 0 and variance of 1. From [10], the sparse random matrix R can satisfy the JL lemma when r_{ij} selected from $\left\{-\sqrt{3}, 0, \sqrt{3}\right\}$ with the probability of 1/6, 2/3, 1/6 respectively. In this case, the random projection result can be approximated to the original random projection result based on the Gaussian distribution.

3. Moving Object Tracking Based on Random Compression of Multi-channel Haar-like Feature

In order to construct a classifier that can distinguish the objet and background, it is necessary to extract the positive and negative samples of the object to be tracked. The initial object location and size information is given by the red rectangle box in Figure 3. The positive sample can be select in a tiny range near the ground-truth. The upper left corner of the positive sample is randomly selected in the green circular area R_1, then the image blocks of the same size are intercepted in the original image as a positive sample. The negative samples are selected in a certain range away from the ground-truth. The upper left corner of the negative sample is randomly selected in the rings between the blue circle R_2 and R_3, then the image blocks of the same size are intercepted in the original image as a negative sample.

Fig. 3. Positive and negative samples selection

In the next frame, in order to obtain the object state, we randomly select the candidate object blocks in a certain range from the previous state, and then identify

the best object block by the classifier which is designed in the next.

When the positive and negative samples are collected, the eigenvector $v = (v_1, v_2 \cdots v_n)^T$ corresponding to each sample is composed of a set of multi-channel Haar-like features, v_i is the i-thHaar-like feature through random compression projection. Since there are no correlations among the features when constructing the multi-channel Haar-like features. That is the Haar-like features are independently distributed. Therefore, the samples can be classified by naive Bayesian classifier.

$$H(v) = \log \frac{(\prod_{i=1}^{n} p(v_i|\, y=1)p(y=1))}{(\prod_{i=1}^{n} p(v_i|\, y=0)p(y=0))} \tag{5}$$

Assuming the prior of positive and negative sample is consistent, that is$p(y = 1) = p(y = 0)$. Then,

$$H(v) = \sum_{i=1}^{n} log\left(\frac{p(v_i|\, y=1)}{p(v_i|\, y=0)}\right) = \sum_{i=1}^{n} H(v_i) \tag{6}$$

Where, $H(v_i)$ can be regarded as the weak classifier corresponding to the multi-channel Haar-like featurev_i.

The eigenvectors after random projection often obey Gauss distribution. It can be assumed that the conditional distribution $p(v_i|\, y = 1)$and $p(v_i|\, y = 0)$ of weak classifiers $H(v_i)$obey the Gauss distribution with mean $\mu_i^1, \mu_i^0$and variance σ_i^1, σ_i^0 respectively.

The contribution of each element in the extracted object feature vector should be different and need to consider the weight influence of different features. In view of the Gauss distribution model, we know that if the variance of Gauss distribution is greater, the data is more scattered and unstable. Therefore, when the variance of Gaussian distribution of a Haar-like feature, the reliability of it is lower. That is to say, this feature has a very small effect on discriminating the object and background, it is considered that the weak classifier corresponding to it should given a smaller weight. Similarly, if the variance of a Gaussian distribution is smaller, it means that the data is more centered. In this case, the data is more reliable and therefore has a greater weight.

In view of the above analysis, the classifier to determine whether a newly acquired image block is an object can be designed as follows.

$$H'(v) = \sum_{i=1}^{n} log\left(\frac{w_i^1 p(v_i|\, y=1)}{w_i^0 p(v_i|\, y=0)}\right) \tag{7}$$

Where, w_1^i, w_0^i are the weights of the multi-channel Haar-like features belong to the positive or negative samples. And can be calculated by the following formula.

$$w_i^1 = \frac{1}{\sqrt{\sigma_i^1}\sum_i \left(\sqrt{1/\sigma_i^1}\right)} \tag{8}$$

$$w_i^0 = \frac{1}{\sqrt{\sigma_i^0} \sum_i \left(\sqrt{1/\sigma_i^0}\right)} \tag{9}$$

In the next frame of the tracking process, the multi-channel Haar-like feature vectors of candidate image blocks are extracted, the candidate image block with the largest classification score is the position in the next frame.

In order to alleviate the offset problem, the classifier parameters need to be updated in real time according to the next formulas.

$$\mu_i^1 \leftarrow \lambda \mu_i^1 + (1-\lambda)\,\mu^1 \tag{10}$$

$$\sigma_i^1 \leftarrow \sqrt{\lambda\left(\sigma_i^1\right)^2 + (1-\lambda)\left(\sigma^1\right)^2 + \lambda(1-\lambda)\left(\mu_i^1 - \mu^1\right)^2} \tag{11}$$

Where, the$\lambda$is the update parameter,
$\mu^1 = \frac{1}{n}\sum_{k=0|y=1}^{n-1} v_i(k)$,$\sigma^1 = \sqrt{\frac{1}{n}\sum_{k=0|y=1}^{n-1}\left(v_i(k) - \mu^1\right)^2}$.

4. Comparative Experiments and Analysis

In order to verify the validity of the proposed algorithm, we use the benchmark test datasets [1]. The dataset is labeled with the tracking difficulty which can provide a complete experimental dataset to the researchers. The comparative experiments are CT tracking algorithm, MIL tracking algorithm, WMIL tracking algorithm, IVT tracking algorithm and L_1-APG tracking algorithm. The qualitative and quantitative comparison shows the stability and validity of the proposed algorithm.

The object state information in the initial frame needs to use the ground-truth. According to the initial object state, the radius of $R_1$is set to 4 pixels, the radius of $R_2$is set to 8 pixels, the radius of $R_3$is set to 45 pixels, the radius of $R_4$is set to 30 pixels, the number of the collected image blocks is 50, and the classifier's update parameter is 0.85. In MIL and WMIL tracking algorithms, the selection of feature pools is 250, the number of weak classifiers is 50, and the learning rate of classifier updating is 0.85. The object Gauss variance selected by the IVT algorithm is selected by (2,2,0.5,0.5), the number of particles is 200 and the object region is interpolated to 32 x 32. The template size of the L_1-APG tracking algorithm is.

Experimental 1- The multi-channel video sequence "*Tiger1*" have many challenges, such as the illumination change, occlusion, non-rigid deformation, motion blur, object suddenly move and the object disappear in the image. The tracking result of these algorithms is shown in Figure 4. As you can see from the figure, the CT algorithm tracking failure while the proposed algorithm tracked successfully. Because it considers the information of the color distribution in the initial tracking stage which can improve the performance of the multi-channel Haar-like feature based on the main color. Meanwhile, the influence of the feature on classifier according to the distribution characteristics of the feature. Under the illumination change scene, it has little influence on the feature.

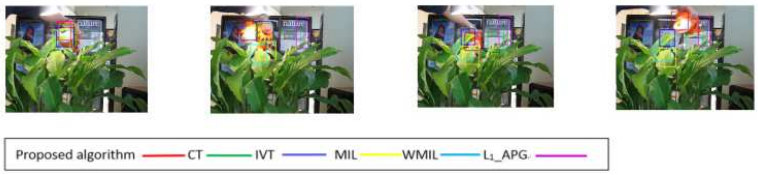

Fig. 4. The experimental results of "*Tiger1*"

Experimental 2- The multi-channel video sequence *"MountainBike"* have many challenges, such as the object is rotated in the image plane, and the object is similar to the background information. The tracking result of these algorithms is shown in Figure 5. As you can see from the figure, the CT algorithm tracking failure while the proposed algorithm tracked successfully. The main consideration of the proposed algorithm is the information of the object's appearance color distribution. Actually, according to the main color information in the background, the tracker moves to a certain extent when the object rotates in space.

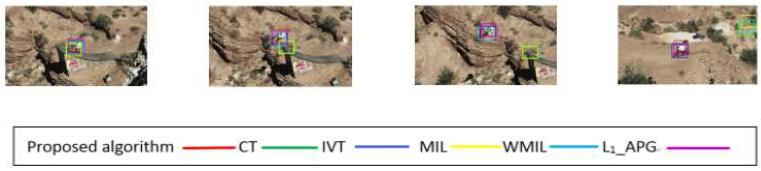

Fig. 5. The experimental results of "*MountainBike*"

The center point error of tracking frame in the tracking result is defined as formula 12.

$$errors = \sqrt{(x_e - x_g)^2 + (y_e - y_g)^2} \tag{12}$$

Where, x_e, y_e are the center coordinates of the experiment result, x_g, y_g are the center coordinates of the ground-truth result. The center point error in the two experiments is shown in Figure 6.

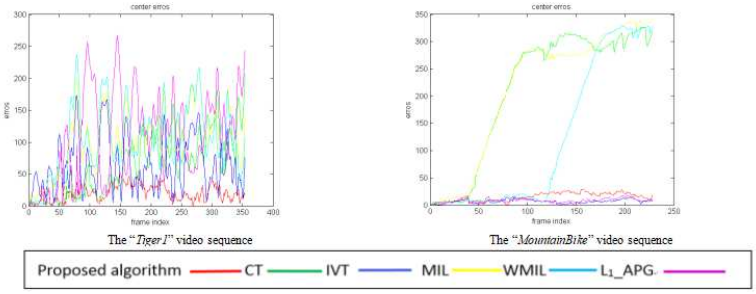

Fig. 6. The center point errors of tracking bounding box

As can be seen from the above figure, the center point error of CT algorithm in the" *Tiger1* " video sequence is more than 20 pixels, while the proposed algorithm

is less than 10 pixels. The center point of CT algorithm in the "*MountainBike*" video sequence is far away from the object position, while the proposed algorithm is around the ground-truth.

5. Conclusion

In this study, we propose a novel tracking algorithm that tracking the object robustly by extract the multi-channel Haar-like feature through random compression projection. Considering the influence of multi-channel Haar-like feature distribution on discriminating object categories, a weighted naive Bayesian classifier is proposed. Finally, the classifier parameters are updated in real time. Through comparison experiments on the challenging video sequences illustrates the effectiveness of the proposed algorithm. In the scenarios which are tracking failure by CT algorithm, the proposed algorithm can tracking it because of the object's appearance information and feature distribution characteristics.

References

[1] M. Kristan, R. Pflugfelder, J. Matas, A. Leonardis, M. Felsberg: *The Visual Object Tracking VOT2015 Challenge Results.* IEEE International Conference on Computer Visio *6* (2015), No. 3, 564–586.

[2] Fussenegger. M, Roth. P, Bischof. H: *A level set framework using a new incremental, robust Active Shape Model for object segmentation and tracking.* Image and Vision Computing *27* (2009), No. 8, 1157–1162.

[3] Hu. W, Li. X, Zhang. X: *Incremental tensor subspace learning and its applications to foreground segmentation and tracking.* International Journal of Computer Vision *91* (2011), No. 3, 171–180.

[4] Chaudhry. R, Hager. G, Vidal. R: *Dynamic template tracking and recognition.* International journal of computer vision *105* (2013), No. 1, 41–52.

[5] Mei. X, Ling. H: *Robust visual tracking and vehicle classification via sparse representation.* IEEE transactions on pattern analysis and machine intelligence *33* (2011), No. 11, 797–804.

[6] Y. Wu, B. Ma, M. Yang, J. Zhang Y. Jia: *Metric Learning Based Structural Appearance Model for Robust Visual Tracking.* IEEE Trans. Circuits Sys *24* (2014), No. 5, 521–528.

[7] W. Guo, L. Cao, T. X. Han, S. Yan, C. Xu: *Max-confidence boosting with uncertainty for visual tracking.* Transactions on Image Processing *24* (2015), No. 5, 555–563.

[8] Yang. F, Lu. H, Yang. M. H: *Robust visual tracking via multiple kernel boosting with affinity constraints.* IEEE Transactions on Circuits and Systems for Video Technology *24* (2014), No. 2, 198–210.

[9] B. Babenko, M. H. Yang, S. Belongie: *Robust Object Tracking with Online Multiple Instance Learning.* IEEE Transactions on Pattern Analysis & Machin *34* (2003), No. 4, 587–606.

[10] Zhang. K, Song. H: *Real-time visual tracking via online weighted multiple instance learning.* Pattern Recognition *46*, (2013), No. 1, 264–275.

[11] Feng. P, Xu. C, Zhao. Z: *Sparse representation combined with context information for visual tracking.* Neurocomputing *225* (2017) 589–597.

[12] Hu. W, Gao. J, Xing. J: *Semi-supervised tensor-based graph embedding learning and*

its application to visual discriminant tracking. IEEE transactions on pattern analysis and machine intelligence *39* (2017), No. 1, 585–590.

Received November 16, 2017