SPARSE COLLABORATIVE APPEARANCE MODEL USING ROBUST OBJECT TRACKING

[1]Vaibhav S. Sharma, [2].Minesh A. Deshmukh, [3].Prafulla T. Jagtap, [4].Anil S. Jagtap., Prof. Kavita Kumavat
Bramha Valley College of Engineering and Research Institute Nasik, Maharashtra, India.

ABSTRACT: *Sparse model has been applied to visual tracking to find the target with the minimum reconstruction error from the target templates subspace. An active video surveillance system using a robust object tracking algorithm which is based on a sparse collaborative model that performs both holistic templates and local representations to account for drastic appearance changes. Within the collaborative appearance model, we use a sparse generative model (SGM) and sparse discriminative classifier (SDC) for object tracking. In the SDC module, for separating the foreground object from the background based on holistic we present a classifier. In the SGM module, a histogram-based method proposed by us that takes the spatial information of each local patch into consideration. The update scheme considers both the original templates and most recent observations, thereby enabling the algorithm to deal with appearance changes alleviate and effectively the tracking drift problem. number of experiments on various challenging videos illustrate that the tracker performs fairly against several state-of-the-art algorithms.*

KEYWORDS: *SDC: sparse discriminative classifier, SGM: sparse generative model.*

I. INTRODUCTION

To identify human being object in video surveillance system is our goal but classification between moving and non-moving object and further classification of moving other objects and human being are difficult challenges to us. SDC and SGM are two basic methods to sort the given problem.

Basically video surveillance is passive monitoring system in which user interaction is not involved due to which incidents at various places are being taken place. To avoid such incidence we have proposed active surveillance method in existing system. The basic steps for object tracking are firstly object detection then object classification and then object tracking. Object Detection is to identify objects of interest in the video sequence frames and to cluster pixels of those objects. In object classification, object can be classified as vehicles, birds, sailing clouds, wobbling tree and other moving objects. Object Tracking can be defined as the problem of approximating the path of an object in the image frame plane as it moves around a scene. The approaches to track the objects are point tracking, silhouette and kernel tracking [2].

II. LITERATURE SURVEY

There is a rich literature in robust object tracking with the most related work.

Wei Zhong et al. [1] presents algorithm using a robust object tracking based on a sparse collaborative appearance model. Within the collaborative appearance model, used a sparse discriminative classifier (SDC) and sparse generative model (SGM) for object tracking. the SDC module, a classifier differentiating the foreground object of frame from the background based on holistic. The training image set is composed of negative templates and positive templates. The object which is to be targeted is represented by background, positive templates and images with part of target object are represented by negative templates. This function provides better object localization as sample templates containing only glimpse of the target are treated as the negative sample templates. So system effectively deals with complex background and cluttered.

In the SGM module, a histogram-based method is presented that takes local appearance information of patches and occlusions into consideration. In this module, overlapped sliding windows are used on the normalized images to obtain collection of all patches and each patch is converted to a vector. Then the dictionary is generated with cluster centers of all the collected patches using the k-means algorithm and the sparse coefficient vector of each patch is normalized and concatenated to form a histogram. Histogram segments of occluded patches are not taken into account when computing the similarity between histograms of candidate and template histogram. SGM module effectively estimates and rejects the occluded patches to improve robustness. Since the appearance of an object often changes significantly during the tracking process, the update scheme is important and necessary. An update scheme is developed in which the SDC and SGM modules are updated independently. For the SDC module, the negative templates every several frames from image regions away the current tracking result are updated. The positive templates remain the same in the tracking process. For the SGM module, the dictionary D is fixed during the tracking process. Therefore, the dictionary is not incorrectly updated due to tracking failures or occlusions. Thus the system effectively deals with appearance changes. However, this system is less effective in handling tracking drifts problem as in this system errors are likely to accumulate during update scheme and can cause tracking failure. And trackers based on holistic appearance model are less effective in handling drifts.

David A. Ross et al. [2] present an appearance based object tracker which incrementally learns a low dimensional subspace illustration of target object for robust object tracking while target undergoes pose, appearance changes, and illumination. To assessment the locations of the target objects in consecutive frames, a sampling algorithm with likelihood estimates, which is in variation to other tracking methods that usually solve complex optimization problems using gradient descent is used. Also, it regularly updates the model representation to show appearance variation of target. Although it has been shown to perform well when target objects undergo pose variation and lighting. this method is less effective in handling heavy non-rigid distortion or occlusion as a result of the adopted holistic appearance model.

S. Avidan et al. [3] present an ensemble tracker which provides pixel based binary classification to classify between target and background. Ensemble tracker keeps an implicit representation of background and foreground using classifiers. In this technique, ensemble tracker combines clump of weak classifier into single strong classifier using AdaBoost to show better result than any of the weak classifier. The tracker regularly updates collection of weak classifier to isolate the foreground and the strong classifier is used to label pixels in next frame. The strong classifier which is used to label every pixels in the next frame as either belonging to the object or the background provides a confidence map. The spike of the map, and hence the new position of the object, is found using mean shift. Although this method is able to separate between target and background, the pixel-based illustration is rather limited and thereby constrains its ability to handle heavy occlusion and clutter.

Adam et al. [4] focused on the fragments-based method to handle occlusions. In this method, histograms are derived for each template patch and then these histograms are analyzed with those derived from multiple regions in target image. The template object is illustrated by multiple image fragments or patches. The patches are arbitrary and are not based on an object model. each patch votes on the achievable positions and scales of the object in the current frame, by comparing its histogram with the related image patch histogram. It minimizes a robust statistic in order to combine the vote maps of the multiple patches. The target object is selected by a voting map formed by comparing histograms of the candidate patches and the related templates. However, the template is not updated and thus this approach is sensitive to large appearance variations.

III. EXISTING SYSTEM

Visual tracking has been mainly formulated within the Bayesian filtering framework in which the goal is to determine a posteriori probability, $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, of the target state by,

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1},$$

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1}),$$

Where $\mathbf{x}_t$ is the object state, and $\mathbf{z}_t$ is the observation at time *t*. Let $\mathbf{x}_t = [l_x, l_y, \theta, s, \alpha, \phi]^\top$, where lx , ly ,θ, s, α, φ denote x, y translations, rotation angle, scale, aspect ratio, and skew respectively. We assume that the affine parameters are independent and modeled by six scalar Gaussian distributions. The motion model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ predicts the state at t based on the immediate previous state, and the observation model $p(\mathbf{z}_t|\mathbf{x}_t)$ describes the likelihood of observing $\mathbf{z}_t$ at state $\mathbf{x}_t$ .The particle filter is an effective realization of Bayesian filtering, which anticipates the state regardless of the underlying distribution. The optimal state is obtained by the maximum a posteriori estimation (MAP) over a set of N samples,

$$\hat{\mathbf{x}}_t = \arg_{\mathbf{x}_t^i} \max p(\mathbf{z}_t|\mathbf{x}_t^i)p(\mathbf{x}_t^i|\mathbf{x}_{t-1})$$

Where $\mathbf{x}_t^i$ is the *i* -th sample at frame *t*. In the next two sections, we present a tracking algorithm within the particle filter framework. We improve the motion model $p(\mathbf{x}_t^i|\mathbf{x}_{t-1})$ as an efficient two-step particle filter, and we present an effective and robust observation model $p(\mathbf{z}_t|\mathbf{x}_t^i)$ based on the collaboration of discriminative and generative models.

IV. ALGORITHM

A) Sparse Discriminative Classifier (SDC)

Motivated by the demonstrated success of sparse representation for vision tasks [5]–[6], [7], [8], [9], we propose

a sparse discriminative classifier for object tracking. In the following, we use the vector x to represent intensity values of a raster scanned image.

Confidence Measure: The SDC method is developed based on the assumption that a target image region

can be better represented by the sparse combination of positive templates w1hile a background patch can be better represented by the span of negative templates. Given a candidate region x, it is represented by the training template set with the coefficients α computed by

$$\min_{\boldsymbol{\alpha}} \|\mathbf{x}' - \mathbf{A}'\boldsymbol{\alpha}\|_2^2 + \lambda_3\|\boldsymbol{\alpha}\|_1$$

Where x' is the projected vector of x and λ3 is a weight parameter. A candidate region with smaller reconstruction error using the foreground template set shows it is more likely to be a target object, and vice versa. Thus, we formulate the confidence value Hc of the candidate x by

$$H_c = \frac{1}{1+\exp\left(-\left(\varepsilon_b - \varepsilon_f\right)/\sigma\right)}$$

Where $\varepsilon_f = \left\|\mathbf{x}' - \mathbf{A}'_+ \alpha'_+\right\|_2^2$ is the reconstruction error of the foreground template set α+ with the candidate x , and A+ is the corresponding sparse coefficient vector. Similarly, $\varepsilon_b = \left\|\mathbf{x}' - \mathbf{A}'_- \alpha'_-\right\|_2^2$ is the reconstruction error of the candidate x using the background template set A−, and α−is the corresponding sparse coefficient vector. The variable σ is fixed to be a small constant that balances the weight of the discriminative classifier and the generative model presented in Section V-B.

In [10], the reconstruction error is computed based on the target (positive) templates, which is less effective for tracking since both the negative and indistinguishable samples(e.g., patches covering some part of a target object) have large reconstruction errors when computed with the target (positive) template set. Thus, it introduces ambiguities in differentiating whether such patches are from the foreground or background. In contrast, our confidence measure exploits the distinct properties of the foreground and the background in computing the reconstruction errors to better distinguish patches from the positive and negative classes.

B) Sparse Generative Model (SGM)

Motivated by recent advances of sparse coding for image classification [11]–[12] as well as object tracking [6], commenced a generative model for object representation that takes local appearance information of patches and occlusions into consideration.

(1) Histogram Generation: For simplicity, use the grayscale features to represent the local appearance information of a target object where each image is normalized to 32 × 32 pixels. use overlapped sliding windows on the regulate images to obtain M patches and each patch is converted to a vector , where G denotes the size of the patch. The sparse coefficient vector β of each patch is computed by

$$\min_{\beta_i} \left\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\beta}_i\right\|_2^2 + \lambda_4 \left\|\boldsymbol{\beta}_i\right\|_1, \quad \text{s.t. } \boldsymbol{\beta}_i \geq 0,$$

Where the dictionary $\mathbf{D} \in \mathbb{R}^{G \times J}$ is generated from *J* cluster centers using the *k*-means algorithm on the *M* patches from the first frame (which consists of the most representative patterns of the target object), and $\lambda 4$ is a weight parameter. In this work, the sparse coefficient vector $\boldsymbol{\beta}_i \in \mathbb{R}^{J \times 1}$ of each patch is normalized and concatenated to form a histogram by

$$\boldsymbol{\rho} = \left[\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \ldots, \boldsymbol{\beta}_M^\top\right]^\top$$

Where $\boldsymbol{\rho} \in \mathbb{R}^{(J \times M) \times 1}$ of every patch is normalized and integrated to form a histogram by is the histogram for one candidate region.

(2) Occlusion Handling: to deal with occlusions, modify the produced histogram to exclude the occluded patches when describing the target object. A patch with large reconstruction error is observed as occluding part and the corresponding sparse coefficient vector is assign to be zero. Thus, a histogram is generated by $\boldsymbol{\varphi} = \boldsymbol{\rho} \odot \mathbf{o}$.

Where $\odot$ denotes the element-wise multiplication. Each element of o is an indicator of occlusion at the corresponding patch and is obtained by

$$o_i = \begin{cases} 1 & \varepsilon_i < \varepsilon_0 \\ 0 & \text{otherwise} \end{cases}$$

Where $\varepsilon_i = \left\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\beta}_i\right\|_2^2$ the reconstruction error of patch yi, and ε0 is is a predefined threshold which determines whether the patch is occluded or not. We thus have a sparse histogram ϕ for each candidate region. The proposed representation scheme takes spatial information of local patches and occlusion into account, thereby making it more effective and robust.

V. EXPECTED RESULT

It will classify human being object in video surveillance system and also classify between moving and non-moving object. It should classify of moving other objects and human being. If human being object detected then it should generate alarm and send SMS.

VI. CONCLUSION

In this paper, demonstrate an effective and robust tracking method based on the collaboration of generative and discriminative modules. In the tracking algorithm, holistic templates are incorporated to construct a discriminative classifier that can efficiently deal with cluttered and complex background. Local representations are adopted to form a robust histogram that considers the spatial information among patches with an occlusion handling module, which enables our tracker to better handle heavy occlusions. The contributions of these holistic discriminative and local generative modules are integrated in a unified manner. Furthermore, the online update scheme reduces drifts and enhances the proposed method to adaptively account for appearance changes in dynamic scenes. Quantitative and qualitative comparisons with nine state-of-the-art algorithms on sixteen challenging image sequences demons.

REFERENCES

[1] Wei Zhong, Huchuan Lu, Ming-Hsuan Yang, "Robust Object Tracking via Sparse Collaborative Appearance Model" IEEE Transaction on image processing, Vol. 23, No. 5, May 2014.

[2] J. Lim, D. Ross, R.-S. Lin, and M.-H. Yang, "Incremental learning for visual tracking," in Advances in Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2005, pp. 793–800.

[3] S. Avidan, "Ensemble tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 494–501.

[4] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2006, pp. 798–805.

[5] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 1822–1829.

[6] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, "Robust tracking using local sparse appearance model and k-selection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2011, pp. 1313–1320.

[7] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Online discriminative object tracking with local sparse representation," in Proc. IEEE Workshop Appl. Comput. Vis., Jan. 2012, pp. 425–432.

[8] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2042–2049.

[9] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Low-rank sparse learning for robust visual tracking," in Proc. ECCV, 2012, pp. 2042–2049.

[10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[11] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in Proc. IEEE Conf. CVPR, Jun. 2009, pp. 1794–1801.

[12] S. Gao, I. W.-H. Tsang, L.-T. Chia, and P. Zhao, "Local features are not lonely—Laplacian sparse coding for image classification," in Proc. IEEE Conf. CVPR, Jun. 2010, pp. 3555–3561.