Object Tracking with Sparse Representation based on HOG and LBP Features

Abhijeet Boragule, JungYeon Yeo, GueeSang Lee*
Department of Electronics and Computer Engineering
Chonnam National University, Gwangju, 500-757, Korea

ABSTRACT

Visual object tracking is a fundamental problem in the field of computer vision, as it needs a proper model to account for drastic appearance changes that are caused by shape, textural, and illumination variations. In this paper, we propose a feature-based visual-object-tracking method with a sparse representation. Generally, most appearance-based models use the gray-scale pixel values of the input image, but this might be insufficient for a description of the target object under a variety of conditions. To obtain the proper information regarding the target object, the following combination of features has been exploited as a corresponding representation: First, the features of the target templates are extracted by using the HOG (histogram of gradient) and LBP (local binary patterns); secondly, a feature-based sparsity is attained by solving the minimization problems, whereby the target object is represented by the selection of the minimum reconstruction error. The strengths of both features are exploited to enhance the overall performance of the tracker; furthermore, the proposed method is integrated with the particle-filter framework and achieves a promising result in terms of challenging tracking videos.

Key words: Feature Extraction, Sparse Representation, Visual Object Tracking.

1. INTRODUCTION

In visual object tracking, the main aim is to estimate the states of a moving target object in video. Visual object tracking is important topic and it has been applied in many real-world applications, such as video surveillance, smart traffic, etc. Despite numerous object tracking methods have been proposed in recent years, developing robust object tracker is still critical task because of many issues in image sequences and appearance change of the object. Generally, internal factors (shape, pose variations) and external factors (varying illumination, camera motion and occlusions) cause tracking failure (See Fig. 1).

The tracking algorithms can be categorized into appearance models and motion models. The appearance model calculates likelihood of an observed image patch belonging to the object and the motion model describes the states of an object using Kalman filter [1] or Particle filter [2]. In our method, a robust appearance model is proposed. Our appearance model considers features of the target templates and it helps to deal with illumination and texture change problem. In order to make an appearance model, we need to consider numerous factors including how objects were represented in earlier works.

* Corresponding author, Email: gslee@chonnam.ac.kr
Manuscript received Jun. 22, 2015; revised Aug. 03, 2015; accepted Aug. 13, 2015

The object representation scheme can be classified based on adopted features that are color [2], texture [3], intensity [4], Haar feature [5] and pixel based feature [6]. Instead of using intensity value and low-level features, the high-level features with sparse representation provide better object tracking results. The representation models can be described as a generative or discriminative model. In a generative model, the target object is represented in particular feature space and then a search is performed for the best matching scores [9]. In a discriminative model, the visual object tracking is treated as a binary classification problem to determine the boundary between foreground and background using classification algorithms [4], [6], [7]. A discriminative model performs better if the training data is big enough; if limited training data is available, in that case, the generative model achieves higher generalization. Several algorithms exploit both models that have been proposed [8], [10], [11]. In incremental subspace learning, the online classifier method includes a template update method [6], [7] and it has been demonstrated to be effective for object tracking.
The straightforward update of appearance models gives slightly less performance due to update and learning state. However, in recent works, an appearance model is proposed, which uses grayscale-based trivial templates and it is still not perfect for appearance change. In this work, our aim is to develop a robust algorithm using a feature-based appearance model that addresses drastic texture change and illumination. The gray-scale information is not enough to represent the target object under various conditions. The gradual changes form ambiguity to the tracker to update proper templates.

The proposed method uses HOG and LBP algorithms. The most important properties of LBP are it is tolerance regarding monotonic illumination changes and its computational simplicity. The LBP describes the texture and HOG describes the edge and corner information which can be fed to classification task. The combine model is account for shape and appearance; in some cases of motion like blurring the edges are still preserved. In some case of partial occlusion or shape change, the texture information is similar. As compared to other feature descriptors, the HOG and LBP takes very less computation time and it is very effective to tracking.

The feature-based sparse representation facilitates tracking algorithm by adjusting appearance change of the target and the background. Numerous challenging image sequences are performed and the proposed algorithm is efficient and effective for tracking object robustly.

2. RELATED WORKS

Sparse representation has been applied recently in many computer vision problems such as object recognition [14], image enhancement [15], and object tracking [10], [12], [16]. Ling et al. [11] exploits sparse representation for the visual tracking to deal with occlusion by using trivial templates. Despite the success of the state of the art methods, there are still several issues to be addressed. First, the method is able to deal with the occlusion problem by using trivial templates. The trivial templates can be used to model any kind background and foreground of the image. The occlusion condition and background reconstruction error may be small from images. In the end, the minimum reconstruction error of the sample with generative formulation forms the ambiguity to the tracker.

In [12], a set of the sparse and discriminative features is used to enhance the robustness of the tracker. The main problem with this method is that the number of discriminative features is fixed which may cause the tracking failure in dynamic and complex image sequences. B. Liu et al. [19] proposes a tracking algorithm based on histogram of local sparse representation. The target object is located using a mean-shift algorithm. The voting maps are constructed based on reconstruction errors. As a result, the histogram generation scheme does not distinguish foreground and background patches in [19]. In our work, the LBP patches are used and weighting method distinguishes occlusion of the target object. Xi Li et al. [17] used an online tensor decomposition method to capture spatial layout information about appearances for tracking object.

There are many previous works that have been using feature-based tracking. Bing et al. [18] proposed feature-based robust object tracking. In this paper, they have proposed algorithm for feature tracking based on SSD criterion and Tomasi-Kanade’s algorithm. To eliminate features of interest, supportive segmentation method is used. In general, the feature point tracking is the association of points and representation of object using feature points. As a result, this approach gives a bunch of points near at the object and it is very difficult to decide the exact object boundary. Tissinaygma et al. [19] proposed point feature-based tracking algorithm. The edge map and corner points are considered and incorporated with Bayesian hypothesis. The proposed algorithm tracks the object using an edge map and corner points. Overall, it gives feature points as output. The feature correspondence is particularly challenging since a feature point in one image may have many similar points in another image, resulting in ambiguity in feature correspondence. In spite of successes in many real circumstances, these point tracking algorithms are not robust to challenges from image occlusions and background. Guo et al. [20] used LBP and color information to form a new feature. The mean shift similarity criterion is integrated with generated feature. Huiyu Zhou et al. [21] proposed mean shift algorithm with integration of SIFT (Scale Invariant Feature Transform). An expectation maximization algorithm is proposed to optimize better similarity search. Later, the SIFT feature is extended to particle filter by Saied Fazil et al. [22]. For tracking object, they used the color-based particle filter and image is converted into a large collection of feature vectors.

3. MOTIVATION

Generally, in color based tracking, mean shift [1] and camshift [27] algorithm are used and it is very sensitive to illumination. To overcome this problem, in most of studies an appearance model has been used [10], [11], [28]. Generally, an appearance model uses a set of gray-scale trivial templates. Due to the lack of information in gray-scale templates, still the proposed algorithms are not perfect for appearance change. In sparse representation, there are several methods that use features for object tracking. However, the gray-scale templates are not enough to describe the target object. The internal and external factors causes degradation in grayscale templates. In image templates, it can include pixel values, textures, and edges. The object features are more important to represent target object. However, in image sequences, the shape of object is retained, but texture changes gradually after several frames. The texture and shape form ambiguity to update the templates of the tracker and it causes tracking.
failure in [10], [11], [28]. The features are more robust to illumination, internal factors and external factors. Generally, the negative template update forms ambiguity in occlusion with cluttered background situation. In our experiment, the negative templates update method is eliminated. The features are used to represent the target object, even if internal and external factor change. The sparse representation approach with features helps to locate and search object with less training data set.

4. PROPOSED METHOD

In this paper, we propose a novel feature-based object tracking using sparse representation. Firstly, the templates are extracted in the particle filter framework. Further, from the set templates, features are extracted using LBP and HOG algorithm. The feature consistent vector is used as input for the sparse representation and sparsity is achieved by solving $l_1$ minimization problem. The minimum construction error is used to represent the target object. The joint feature model is proposed to enhance the performance of the tracker.

4.1 Feature Based Sparse Representation

In this work, the LBP and HOG algorithms use features from a set of templates. A set of templates is acquired in the particle filter framework. In the process of tracking, the particle filter is used for estimating the target state sequentially. LBP and HOG feature of target templates are formed as shown in fig. 2.

![Feature Extraction Method](image)

**Fig. 2. Result of Feature Extraction Method**

4.1.1 Sparse representation with HOG

The training image set is composed of 32x32 manually selected target location images. Each down sampled image is stacked to the corresponding HOG feature vector. In sparse representation, the equation 1 represents $l_1$ minimization problem. The coefficient computed by equation is used to represent target object using candidates and templates. The candidate is represented by training templates set with the coefficients $\alpha$ computed by

$$
\min_{\alpha} \left\| x' - A' \alpha \right\|^2 + \lambda \left\| \epsilon \right\|^2
$$

(1)

where $A'$ is training HOG feature vector and $x'$ is a candidate HOG feature vector. A candidate with smaller reconstruction error is more likely to be a target object, and vice versa. In here, the confidence value of candidate feature vector is formulated by

$$
\epsilon_i = \left\| x' - A' \alpha \right\|^2
$$

$$
H_c = \exp(-\epsilon_i / \sigma)
$$

(2)

(3)

where $\epsilon_i$ is the reconstruction error of the candidate x with HOG feature set $A$, $\alpha$ is the corresponding sparse coefficient vector and $H_c$ is the confidence value. The $\lambda$ is the variable for sparse representation which controls the coefficient $\alpha$. The constant $\sigma$ is small, which helps to balance weight in the classification of features. A similarity value is used to classify object with foreground. The confidence value between candidate and template is computed. The confidence value is used with similarity function in section 5.

4.1.2 Sparse representation with LBP

In this section, a sparse representation based generative model is proposed. The histogram of sparse coefficient is formulated for tracking problem. It helps to deal with location information of patches and it treats the occlusion problem. The LBP information is used to represent the local information of the object. The normalized patches $M$ are obtained by using sliding windows on normalized images. Each image patch is resized to a vector and the sparse coefficient of LBP vector patch is computed from every vector.

Similarly,

$$
\min_{\beta_i} \left\| x' - K \beta \right\|^2 + \lambda \left\| \beta \right\|^2
$$

(4)

where $K$ is the histogram of the LBP candidate. $K$ is formed by using k-means algorithm and it consists of most representative patterns of the target object in the first frame. As shown in fig. 3, the sparse coefficient vector $\beta_i$ of each patch is concatenated to form histogram by using following operation.

$$
\beta = [\beta_1, \beta_2, ..., \beta_i]
$$

(5)

where $\beta \in R^{(JxN)x1}$ is the histogram of a candidate.
The similarity function is used to compute the similarity between LBP candidate and the template LBP candidate,

\[ L_c = \sum_{j=1}^{N} \min(\psi_j^c, \varphi_j^c) \quad (6) \]

where \( \psi_j^c \) and \( \varphi_j \) are the weighted histogram of the c-th LBP candidate and LBP template. The weighted histogram \( \psi_j^c \) is generated by using element wise multiplication of each patch with sparse coefficient of candidate \([10]\).

The similarity function gives the similarity value between candidate and template. The feature-based patches achieve better accuracy instead of using gray-scale patches. The features are more robust to illumination and in occlusion condition. The occlusion handling step is defined in the section 4.1.4. In this process, the patches are used to match the candidate with templates. The sliding window is used to generate the patches of the candidate. The patch wise matching allows us to deal with occlusion problem.

4.1.3 Occlusion handling scheme

In order to handle the occlusion problem, the weighted histogram is modified \([10]\). The patch with large reconstruction shows the failed matching state between template and candidate. The threshold value is a set between reconstruction errors of patch which determine the occlusion. The weighted histogram contains indicator occlusion. If the reconstruction error is greater than the threshold, then the weighted histogram is set to be 0 and it indicates the occlusion.

5. JOINT FEATURE MODEL

In this paper, we propose a joint feature model for object tracking. In our tracking algorithm, the final confidence value based on the HOG feature similarity values and the similarity values of the LBP contribute together. The maximum likelihood function is constructed using candidate and template similarity and it achieves robust performance.

\[ F_c = H_c L_c \quad (7) \]

where \( L_c \) and \( H_c \) give the higher confidence value of given features. The multiplication formula of joint feature model is more effective in illumination condition. The object classification as foreground from background is done by using the confidence value of the histogram of the gradient. It helps to locate the object in consecutive frames. In this work, the confidence value \( H_c \) works as a weight for the similarity function. The local similarity function helps to decide the target object. The joint feature sparse representation finally describes the target object. The maximum from likelihood is defined as final tracking candidate.

6. UPDATE METHOD

In visual object tracking, object appearance changes gradually during the tracking process, the update of template is more important and necessary. In the HOG based on features, we keep templates remaining in the entire video sequences. It aims to classify object from the background. In LBP feature-based sparse representation, in every 2 frames the histogram is updated to capture new appearance of the target object \([10]\).

7. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed tracker, we compare our results with other state of the art methods as shown in fig. 6.

7.1 Quantitative Comparison

In our experiment, we evaluate IVT \([4]\), L1 tracker \([11]\), Frag tracker \([13]\), MIL tracker \([23]\), VTD \([25]\) and PN learning algorithm \([24]\) algorithms using the center location error and overlap rate. The Table 1 and Table 2 and Fig. 6 show the centre location error and overlap rate of the evaluated algorithms.

7.2 Qualitative Comparison

Our proposed method shows better performance with the help of features. As shown in fig. 4, in appearance model \([10]\), \([11]\), \([28]\), the object passes through the occlusion with changed texture; the proposed tracker tracks the object. In condition of texture and illumination change, the LBP and histogram of features are more robust in object tracking. As shown fig. 4, the object passes through texture change and in occlusion condition, the appearance models are failed and it loses the target object because of the less information about object.

As shown in fig. 5, the video sequence (a) girl consists of partial occlusion and (e) face occlusion which is appeared in (a) frame no 428 and (e) frame no268. The IVT \([4]\), L1 tracker \([11]\), Frag tracker \([13]\), MIL tracker \([23]\), VTD \([25]\) and PN learning algorithm \([24]\) are not able to keep the tracking rectangle over target object. In our tracking scheme,
the tracking result will not shift the tracking rectangle to the background due to elimination of negative templates. The (b) frame 397 has rotation state, expect IVT and our proposed method other method has less performance and they have failed to adopt the rotation state. The affine transformation with better LBP match adopts the rotation state. The foreground object is correctly matched with positive templates as result; the tracking window has stability over object. As shown in Fig. 5, the (b) car sequence has illumination state. The IVT and our proposed method perform well. The complex background as shown in fig. 5 (c) stone is still challenging because there are many similar object with target object. The IVT and proposed method performed well due to proper matching features.

The proposed feature-based tracking method still tracks the object, even if the texture changes condition. The grayscale information is not enough to track the object in the certain condition such as effect of internal and external factors. Our tracker uses target object templates and their features and it will only classify the object from the background. The negative template update strategy causes tracking failure because it considers positive templates as a negative template. In our experiment, the negative templates are neglected. The proposed algorithm is implemented in MATLAB. In video image sequences, the location of the object is manually labelled in the first frame.

In feature sparse representation, each image is normalized to 32x32 in our experiments. We use 600 sample images in the particle filter frame work. The input of our tracker is updated after frames work. The regularization constant $\lambda$ is set to 0.4 in all experiments. We used seven challenging image sequences. In our experiment, we evaluate tracker against six state of the art algorithms including IVT [4], L1 tracker [11], Frag tracker [13], MIL tracker [23], VTD [25] and PN learning algorithm [24].

Table 1. Average centre error pixel [26]. The best and second best results are shown in green and red fonts.

|       | IVT [4] | L1 [11] | PN [24] | VTD [25] | MIL [23] | FRAG [13] | Ours |
|-------|---------|---------|---------|---------|---------|-----------|------|
| Car11 | 2.10    | 33.25   | 25.11   | 27.05   | 43.46   | 63.92     | 2.85 |
| Occlusion2 | 10.25 | 11.11   | 18.58   | 10.40   | 14.06   | 15.49     | 4.46 |
| Caviar1 | 45.20  | 118.99  | 5.59    | 3.90    | 48.49   | 5.70      | 2.17 |
| Stone  | 2.25    | 19.222  | 7.98    | 31.30   | 32.35   | 65.89     | 5.43 |
| Girl   | 45.24   | 119.93  | 5.59    | 3.90    | 48.49   | 5.69      | 2.17 |

Table 2. Overlap Rate of Tracking Methods [26]. The best and second best results are shown in green and red fonts.

|       | IVT [4] | L1 [11] | PN [24] | VTD [25] | MIL [23] | FRAG [13] | Ours |
|-------|---------|---------|---------|---------|---------|-----------|------|
| Car11 | 0.80    | 0.43    | 0.37    | 0.43    | 0.17    | 0.08      | 0.70 |
| Occlusion2 | 0.58 | 0.67    | 0.49    | 0.59    | 0.61    | 0.60      | 0.83 |
| Caviar1 | 0.27   | 0.27    | 0.70    | 0.53    | 0.25    | 0.68      | 0.82 |
| Stone  | 0.65    | 0.29    | 0.41    | 0.41    | 0.32    | 0.15      | 0.49 |
| Girl   | 0.42    | 0.32    | 0.57    | 0.51    | 0.51    | 0.68      | 0.60 |
Fig. 5. Quantitative evaluation in terms of center location error (in pixel). The proposed method is compared with six state-of-the-art methods on 5 challenging image frames.

(d) Tracking results on sequence stone with complex background

(e) Tracking results on sequence faceocc with severe occlusions

8. CONCLUSION

In this work, a feature-based sparse representation method for video object tracking is proposed. The feature-based tracking method is applied with sparse representation to address the problem of appearance change of the object. The histogram of the gradient and LBP features templates are used in the particle filter framework. Later, the feature sparsity is attained by solving $l_1$ minimization problem. The target object is represented by selecting minimum reconstruction error. The proposed method exploits the strength both HOG method and LBP features. The joint feature sparse representation contributes together to enhance the performance of the algorithm. Experiment on challenging videos demonstrates that our tracking method performs favourably against state of the art methods.

ACKNOWLEDGEMENT

This research was supported partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by MEST (2015-024474) and by the technological innovation R&D Program of SMBA (S2173771).

REFERENCES

[1] D. Comaniciu, V. R. Member, and P. Meer, “Kernel-based object tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 5, May. 2003, pp. 564-575.
[2] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, “Color-based probabilistic tracking,” in Proc. Eur. Conf. Computer Vision, 2002, pp. 661-675.
[3] S. Avidan, “Ensemble tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, Feb. 2007, pp. 261-271.
[4] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, “Incremental learning for robust visual tracking,” Int. J. Computer Vision, vol. 77, nos. 1-3, 2008, pp. 125-141.
[5] H. Grabner and H. Bischof, “On-line boosting and vision,” in Proc. IEEE Conf. Comput. Vision Pattern Recognt., Jun. 2006, pp. 260-267.
[6] S. Wang, H. Lu, F. Yang, and M.-H. Yang, “Superpixel tracking”, in Proc. IEEE Int. Conf. Computer Vision, Nov. 2011, pp. 1323-1330.
[7] A. Adam, E. Rivlin, and I. Shimshoni, “Robust fragments-based tracking using the integral histogram,” in Proc. IEEE Conf. Computer Vision Pattern Recognition, Jun. 2006, pp. 798-805.
[8] S. Avidan, “Support vector tracking,” IEEE Trans. Pattern Anal. Machine Intelligent, vol. 26, no. 8, Aug. 2004, pp. 1064-1072.
[9] Q. Yu, T. B. Dinh, and G. G. Medioni, “Online tracking and reacquisition using co-trained generative and discriminative trackers,” in Proc. Eur. Conf. Computer Vision, 2008, pp. 678-679.
[10] W. Zhong, H. Lu, and M.-H. Yang, “Robust object tracking via sparsity-based collaborative model,” in Proc. IEEE Conf. Comput. Vision Pattern Recogn., Jun. 2012, pp. 1838-1845.
[11] X. Mei and H. Ling, “Robust visual tracking using ℓ1 minimization,” in ICCV, 2009, pp. 1436-1443.
[12] B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, and C. Kulikowski, “Robust and fast collaborative tracking with two stage sparse optimization,” in ECCV, 2010, pp. 624-637.
[13] A. Adam, E. Rivlin, and I. Shimshoni. “Robust fragments-based tracking using the integral histogram,” In CVPR, 2006, pp. 798-805.
[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. “Robust face recognition via sparse representation,” 2009, pp. 210-227.
[15] J. Yang, J. Wright, T. S. Huang, and Y. Ma. “Image super-resolution via sparse representation,” TIP, 2010, pp. 2861-2873.
[16] B. Liu, J. Huang, L. Yang, and C. Kulikowski. “Robust tracking using local sparse appearance model and k-selection,” In CVPR, 2011, pp. 1313-1320.
[17] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo, “Robust Visual Tracking Based on Incremental Tensor Subspace Learning,” in Proc. ICCV, 2007, pp. 1-8.
[18] B. Han, W. Robert, D. Wu, and J. Li. “Robust Feature based Object Tracking,” SPIE DSS 2007, Orlando, FL, USA, Apr. 9-13, 2007.
[19] P. Tissainayagama and D. Suterb, “Object tracking in image sequences using point features,” Pattern Recognition, vol. 38, issue 1, Jan. 2005, pp. 105-113
[20] Zuren Feng Shuai Wang Qin Nie, “A Multiple Features Image Tracking Algorithm,” Fifth International Symposium on Computational Intelligence and Design, Oct. 28-29, 2012, pp. 77-80.
[21] Huiyu Zhoua, Yuan Yuanb, and Chunmei Shi, “Object tracking using SIFT features and mean shift,” Computer Vision and Image Understanding, vol. 113, issue 3, Mar. 2009, pp. 345-352.
[22] S. Fazli, H. M. Pour, and H. Bouzari, “Particle Filter Based Object Tracking with Sift and Color Feature,” Second International Conference on Machine Vision, Dec. 28-30 2009, pp. 89-93.
[23] B. Babenko, M.-H. Yang, and S. Belongie. “Visual tracking with on-line multiple instance learning,” In CVPR, 2009, pp. 983-990.

[24] Z. Kalal, J. Matas, and K. Mikolajczyk. “P-N learning: Bootstrapping binary classifiers by structural constraints,” In CVPR, 2010, pp. 49-56.
[25] J. Kwon and K. M. Lee. “Visual tracking decomposition,” In CVPR, 2010, pp. 1269-1276.
[26] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. “The PASCAL Visual Object Classes Challenge,” Results, 2010.
[27] G. R. Bradski, “Computer Vision Face Tracking for Use in a Perceptual User Interface,” In: Intel Technology Journal, 1998, pp. 13-27.
[28] Dong Wang. “Online Object Tracking With Sparse Prototypes,” IEEE Transaction on Image Processing, vol. 22, no. 1, Jan. 2013, pp. 314-325.

Abhijeet Boragule
He received the Bachelor in Computer Science from Shivaji University, Kolhapur City, India in 2012. He is currently a MS student in Electronics and Computer Engineering department of Chonnam National University, Korea. His interesting research fields are multimedia and pattern recognition.

JungYeon Yeo
She received a B.S. degree in Mathematics from Chonnam National University, Korea in 2014. She is currently an M.S. student in the Electronics and Computer Science department of Chonnam National University, Korea. Her research interests include multimedia image processing, and pattern recognition.

GueeSang Lee
He received a B.S. degree in Electrical Engineering and an M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received a Ph.D. degree in Computer Science from Pennsylvania State University in 1991. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video technology.