A Survey on Correlation Filters for Object Detection

Disha Davis[1] Nicy Johnson[2]
[1]Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]IES College of Engineering, Thrissur, India

Abstract— Object detection is a sequence of segments comprises the objects of interest, which are taken as preprocessing and expansively applied in various vision tasks like face detection, face recognition, video object co-segmentation. Out of these, Correlation filters are the special classifiers which are designed for accurate object detection. Normal correlation filters require scalar valued image features and only perform on objects with consistent appearance but newer filters work with feature spaces that establish some invariance to small deformations. However, most of existing approaches only utilizes the proposals to compute the location; the aim is to propose a fastest algorithm for making the object detection as an efficient task. This paper accentuates correlation techniques and feature extraction methods used by researchers for object detection.

Key words: Object Detection, Correlation Filters, Feature Extraction, Classifiers

I. Introduction

Today, computer vision has become an integrated part of people's lives. It has, and continues to influence many aspects of daily life and has allowed better working in robotics, real time videos and can also be used for image enhancement, object detection etc. Image processing is one of the rapidly growing technologies and is one of the important parts of computer vision. Of these, image enhancement or the object detection method has become the recently important task due to its many possible applications, such as object recognition, face recognition and video co-segmentation. However, regardless of the ongoing growth in the field of object detection, it is still a challenging job as the process of detecting object is affected not only by internal factors like accuracy, efficiency, but also by external factors, such as noises, illuminations and occlusions. Several ways has been introduced to detect the object that is associated with the image or the video. One of the important methods is by using correlation. So, a survey was done among different proposals and this paper consists survey among different methods for object detection.

II. Motivation

Object detection is a prominent task related to image processing that deals with identifying occurrences of objects in a particular classes such as humans, buildings or cars in digital images and videos. At present, it has become a relevant task and researchers are finding the faster methods for accurately detecting the object. Every object class has its own special features that help in classifying the class. Several methods had been introduced in order for detection of an object. But the procedures and the algorithms used in each method are different. The main concern is that, how to efficiently and accurately detects the object without causing any delays. In views of that, handling the noises, illuminations, background clutters and occlusions becomes more important. So that the filters must be designed in order to increase the accuracy, efficiency, speed and performance thereby reducing the human time. There are many popular real world applications that are directly or indirectly related to object detection. With the availability of large amounts of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy. Object detection through automated systems is useful when we need to label an object that gets resided in an image or a video more efficiently.

III. Literature Survey

Correlation filters are considered as the special classifiers which are used for shift invariant object recognition and also robust for patters distortion. The idea behind the correlation filter is first proposed by Hester and Casasent namely multivariant technique for multiclass pattern recognition [1]. Correlation with an example template works well if the target does not change notably from image to image and video to video. Unfortunately, the appearance of the target does change across images in most of the domains, due to variations among target instances and changes in imaging condition. It highly focuses on formulating the theory and later on more practical applications. It is a technique for multiclass optical pattern recognition of different standpoint views of an object is described. Each multiclass depiction of an object is recounted as an orthonormal basis function expansion, and a single averaged matched spatial filter is then obtained from a weighted linear combination of these functions. The technique is manifested for a terminal missile guidance application using IR tank imagery. Though it is a simple method and later works had been proposed in order to obtain high efficiency.

D. S. Bolme et al. [2] proposed a newer type of correlation filter called average of synthetic exact filters that differs from the traditional filter is that it will specify the entire correlation output on each training image by tuning filters for particular tasks. This is in marked contrast to prior methods which only specify a single output value per training image. It has two important properties. First, an entire correlation response surface is specified for each training instance during filter construction. Second, the resulting filters are simply averaged. The resulting filters are less susceptible to over-fitting the training data than other methods, and can therefore be trained over larger and more inclusive training sets. As a result, they outperform previous methods. One consequence of completely specifying the correlation output is a perfect balance between constraints and degrees of freedom for each training image, and therefore a complete "exact filter" is determined for every training image. ASEF is shown to out sail all these methods, locating the eye within the radius of the iris approximately 98.5%. The work had been extended to locate more points on faces, such as the nose, eyebrows, and mouth with excellent results.

To work with more variety problems including face detection, face verification, automatic target recognition, and medical image registration, Vitomir Struc et al. [3] proposed principal directions of synthetic exact filters. It exhibits some desirable properties such as relatively short training times, computational simplicity, high localization rates and real time capabilities. Instead of only relying on the average of a set of synthetic exact filters, as it is the case with the ASEF filters, eigenvectors of the correlation matrix of the SEFs as correlation templates (or filters) are employed hence, the name PSEFs. After determining the location of the facial regions in all images, 640 images are selected from the LFW database and manually label the locations of the left and right eye. Through these transformations, the initial set of 640 images are augmented to a set of 25600 images (of size 128×128 pixels) and employed for training of the ASEF and PSEF filters. For testing purposes, the same random transforms are applied to 3815 images from the FERET database. Here, only 12 modifications of each facial regions are produced, which results in 45780 facial images being available for the assessment.

Hamed Kiani Galoogahi et al. [4] proposed multi channel correlation filters that can be efficiently posed as a correlation/convolution between a multi-channel image and a multi-channel detector/filter which results in a single channel response map indicating where the pattern (e.g. object) has occurred. To demonstrate the effectiveness of this strategy, it is evaluated across a number of visual detection/localization tasks where it exhibit superior performance to current state of the art correlation filters, and superior computational and memory efficiencies compared to state of the art spatial detectors. This approach is applied across a myriad of detection and localization tasks including: eye localization, car detection and pedestrian detection and demonstrated superior performance to current state of the art single-channel correlation filters, and superior computational and memory efficiency in comparison to spatial detectors (e.g. linear SVM) with comparable detection performance.

Samya Bagchi et al. [5] proposed object recognition algorithm using maximum margin correlation filter & SVM. It considers the problem of detecting objects in two dimensional images and proposed a new technique which uses support vector machine (SVM) along with maximum margin correlation filter. This algorithm detects objects well and is robust with respect to scale changes. Introduction to support vector machine (SVM) helps maximum margin correlation filter to compacts with non-linearly separable data. It also detects same object, if it is found several times at various scales, thus it helps avoiding detection of same object and finally selects the best version of it. Application of SVM on top of this filter also helps us to detect and select the best peak corresponding to the object. It also helps in detecting object by choosing the best among several scaled versions. This algorithm can be useful for a good number of applications like automatic target recognition, robot vision etc. which requires some more attention.

Ryan Tokola et al. [6] proposed ensembles of correlation filters for object detection which jointly learning an ensemble of correlation filters that collectively capture as much variation in object appearance as possible. Traditional correlation filters for object detection are well organized and provide good localization, but needs scalar valued image features and only perform well on objects with consistent appearance. Some filters work with feature spaces that introduce invariance to few deformations, but more difficult detection problems require more than one filter. During training the filters adapt to the needs of the training data with no restrictions on size or scope. Then the performance that overreaches the state of the art in several challenging experiments is demonstrated. It does not need to determine a priori the size of the filters, the function of the filters, or the origin of the filters. These filters are flexible and powerful enough to be used as the base of many different object detection models. These filters significantly outperform the baseline methods. Training consists of effortlessly parallelizable least square optimizations.

Hamed Kiani Galoogahi et al. [7] proposed correlation filters with limited boundaries. It demonstrate that only 1/D proportion of shifted examples are unaffected by boundary effects which has a dramatic effect on detection/tracking performance. a novel approach is proposed to correlation filter estimation that it takes advantage of inherent computational redundancies in the frequency domain, dramatically reduces boundary effects, and it is able to implicitly exploit all possible patches densely extracted from training examples during learning process. Imposing object tracking and detection results are shown in terms of both accuracy and computational efficiency. Moreover, the proposed approach implicitly learns correlation filters over an embedded dense sampling strategy which is inherited from the shift circular property of the convolution operation in the Frequency domain. This allows one to learn an effective detector/filter by exploiting a huge set of negative examples with very logical memory cost which was shown to be independent of the number of training images and sampled patches. This approach demonstrated superior empirical results for both object detection and real-time tracking compared to current state of the arts.

Correlation filter tracking via bootstrap learning proposed by

Kunqi Gu et al. [8] derives a multi-feature and multi-kernel correlation filter based tracker which fully takes advantage of the invariance-discriminative power spectrums of various features and kernels to further improve the performance. A novel bootstrap learning method is utilized to obtain a strong classifier by fusing these weak kernel correlation filters. Moreover, a new target scale estimation strategy is incorporated into our framework. The efficient and effective scale estimation method is based on target dictionary representation. This method aims at finding an optimal combination of hyper plane regarding specific features that can best classify two classes from a large pool of classifiers and image attributes. Besides, the proposed scale estimation method is based on incremental learning are also effective and efficient. Quantitative and qualitative comparisons with seven state-of-the-art trackers on ten videos have demonstrated the effectiveness and robustness of the proposed tracker.

Eric J. He et al. [9] proposed masked correlation filters for partially occluded face recognition. A challenge is to perform accurate face recognition when there exist partial occlusions of the face such as scarves or sunglasses. Correlation Filters are an occlusion-tolerant object recognition method and it is potentially suited to deal with partial occlusions. A new class is introduced of correlation filters called masked correlation filters, which are designed specifically to handle partial occlusions in face images. The benefits of using masked correlation filters are illustrated using well-known face image data sets. While its design needs prior knowledge of occluded regions, it may be possible to determine occlusions using a soft-biometric detector. It could also be useful in other pattern recognition tasks, such as automatic target recognition.

Yang Liu et al. [10] proposed latent bi-constraint SVM for video-based object recognition which addresses the task of recognizing objects from video input. It is applied to recognize office objects and museum sculptures, and demonstrate its benefits over image-based, set based, and other video-based object recognition. Its technical novelty is threefold. The first constraint expands the training video, associates the object label to all subsequences of each training video. This enforces all subsequences of training video to be correctly classified, enabling the recognition of an object from various viewpoints. The second constraint requires the monotonicity of the score function with respect to the inclusion relationship between subsequences of a video. This is to ensure the consistency of the recognition decisions. The third is the incorporation of the latent variable allows the monotonicity requirement to be satisfied, discarding bad views of an object due to such factors as occlusion and motion blur. Given the training samples, at the core of correlation filtering is to find the optimal filters, which requires the unknown variable distribution estimation.

IV. Conclusion

In this paper latest works in the field of object detection were discussed. Many researchers had contributed and are still working in this field. Though there are a number of problems in existing systems that need to be addressed such as occlusions, noises and background clutters that are associated with the object. However some of the problems like redundant images on various scales, detecting different objects on same image have already been solved but still have room for improvement. It is concluded, correlation filters are introduced to overcome all the existing disadvantages by relating the pixels like difference between foreground and background image, illumination variations etc., thereby it can accurately detect the object by working in a linear fashion. It will have consistent advantages over the state-of-the-arts when applying this technique to several computer vision applications including eye detection, car detection and object tracking. In all-purpose, different object detection approaches and algorithms can be used in order to get effective results in different real life scenarios.

Acknowledgment

I would like to convey my heartfelt gratitude towards my guide Mrs. Nicy Johnson for her constant guidance, encouraging help and inspiring words. I am thankful to the department of computer engineering for their support.

References

[1] C. F. Hester and D. Casasent “Multivariant technique for multiclass pattern recognition”, Applied Optics, vol. 19, no. 11, pp. 1758–1761, 1980.
[2] D. S. Bolme, B. A. Draper and J. R. Beveridge, “Average of synthetic exact filters”, IEEE Conference on Computer Vision Pattern Recognition (CVPR), pp. 2105–2112, 2009.
[3] Vitomir Struc, Jerneja Zganec Gros and Nikola Pavesi “Principal directions of synthetic exact filters for robust real-time eye localization”, Springer, pp.180-192, 2011.
[4] Hamed Kiani Galoogahi, Terence Sim and Simon Lucey, “Multi-channel correlation filters” IEEE International Conference on Computer Vision, pp. 3072–3079, 2013.
[5] Samya Bagchi and Poonacha P. G, “An object recognition algorithm using maximum margin”, IEEE Twentieth National Conference on Communications (NCC), pp.1 – 6, 2014.
[6] Ryan Tokola and David Bolme " Ensembles of correlation filters for object detection”, IEEE Winter Conference on Applications of Computer Vision, pp. 1028-1034, 2015.
[7] Hamed Kiani Galoogahi, Terence Sim and Simon Lucey, “Correlation filters with limited boundaries”, IEEE Conference on Computer Vision Pattern Recognition, pp. 4630–4638, 2015.
[8] Kunqi Gu, Tao Zhou, Fanghui Liu, Jie Yang1 and Yu Qiao, “Correlation filter tracking via bootstrap learning”, IEEE International Conference on Image Processing (ICIP), pp. 459-463, 2016.
[9] Eric J. He, Joseph A. Fernandez, B. V. K. Vijaya Kumar and Mohamed Alkanhal, “Masked correlation filters for partially occluded face recognition”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1293-1297, 2016.
[10] Yang Liu, Minh Hoai, Mang Shao and Tae-Kyun Kim, “Latent bi-constraint SVM for video-based object recognition”, IEEE Transactions on Circuits and Systems for Video Technology, pp. 3044 – 3052, 2017.

