Hybrid approach for Semantic object Detection in Video

R.Renugadevi
PG scholar
Computer Science and Engineering
Sri Vidya College of Engineering and Technology,
Virudhunagar, Tamilnadu

Dr. P. Murugeswari
Professor
Computer Science and Engineering
Sri Vidya College of Engineering and Technology,
Virudhunagar, Tamilnadu

Abstract

Object detection is the process of finding moving objects such as human beings, animals and vehicles in videos. Object detection algorithms typically use extracted features and learning algorithms to recognize instances of an objects. It is commonly used in applications such as security, automated vehicle parking systems, surveillance, and image retrieval . Object detection has applications in many areas of computer vision, image retrieval and video surveillance. The input video is converted into frames. The frames were preprocessed to remove the noises from the frames. The objects are detected from the frames by initially identifying the object regions by applying Background subtraction. The thresholding operation and morphological operations were applied to extract the object alone. The obtained object pixels were then grouped and the identified positions were then represented in different color inorder to obtain them exactly.

Keywords: Surveillance, Background subtraction, Thresholding operation and Morphological operation.

I. INTRODUCTION

Video tracking is the process of locating a moving object over time using a web camera. A variety of uses, some of which are: security and surveillance, video communication human-computer interaction, compression, traffic control, augmented reality, video editing and medical imaging [16]. Video tracking can be time consuming process due to the amount of data in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking. To perform video tracking an algorithm analyzes sequential video frames and outputs the movement of targets between the frames. A variety of algorithms, each having strengths and weaknesses. There are two main components of a visual tracking system: localization and target representation, as well as filtering and data association. Target representation and localization is mostly a bottom-up process. These methods give a variety of tools for identifying the moving object. Locating and tracking the target object successfully is dependent on the algorithm. For example, using blob tracking is useful for identifying human movement because a person's profile changes dynamically. Tracking of video objects is one of the major issues in video surveillance system. To analyze the behavior of specific person, the target's position needs to be correctly located in consecutive frames[3]. The challenges in tracking include illumination change, Occlusion, scale change, or object perspective etc. Kernel-based video object tracking has recently widely investigated for better more robust tracking performance. kernel-based tracking is introduced to minimize the difference between the reference color distribution and the candidate region color distribution in the current frame[2]. The main contributions of this paper are: (i) Computationally efficient use of the projected gradient optimization to help multiple kernels find the best match of the tracked target under predefined constraints. (ii) Since not all of the kernels are reliable owing to occlusion, we need to assign the appropriate weights to them. We combine the velocity consistent and similarity into the weights computation for different kernels. (iii) Effective use of the gradient of the density estimator with respect to the bandwidth parameter to update the scale change of the object [2]. The input is the video frame, and the output is the tracking result for each object in the frame. First, the system does the background subtraction and the foreground objects are extracted [3].

II. EXISTING APPROACHES

LBP features combined with SVM were proposed to identify the objects in the video. Hierarchical SVM and star cascade SVM were proposed which contains some modifications compared to the traditional LBP measures. Boosting classifiers that classifies the object and the background information's from the image. The original LBP operator labels the pixels of an image with decimal numbers, called LBPs or LBP codes that encode the local structure around each pixel. Each pixel is compared with its eight neighbors in a 3 × 3 neighborhood by subtracting the center pixel value; the resulting strictly negative values are encoded with 0, and the others with 1. For each pixel, a binary number is obtained by concatenating all these binary values in a clockwise direction, which starts from the one of its top-left neighbor. The corresponding decimal value of the generated binary number is then used for labeling the given pixel. The derived binary numbers are referred to be the LBPs or LBP codes. One limitation of the basic LBP operator is that its small 3 × 3 neighborhood cannot capture dominant features with large-scale structure. The texture at different scales, the operator was later generalized to use neighborhoods of different sizes. A local neighborhood is defined as a set of sampling points evenly spaced on a circle, which is centered at the pixels to be labeled, and the sampling points do not fall within the pixels are interpolated using bilinear interpolation, thus allowing for any number of sampling points and any radius in the neighborhood. A weak learner is defined to be a classifier which is only slightly correlated with the true classification. In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classifications. While boosting is not algorithmically constrained, most boosting algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. They are typically weighted in some way that is usually related to the weak learners' accuracy. After a weak learner is added, the data is reweighted: examples that are misclassified gain weight and examples that are classified correctly lose weight. A technique for speeding up processing of boosted classifiers, early termination refers to only testing each potential object with as many layers of the final classifier necessary to meet some confidence threshold, speeding up computation for cases where the class of the object can easily be determined. One such scheme is the object detection framework introduced by Viola and Jones: in an application with significantly more negative samples than positive, a cascade of separate Adaboost classifiers is trained, and the output of each stage biased such that some acceptably small fraction of positive samples is mislabeled as negative, and all samples are marked as negative after each stage are discarded [5]. If 50% of negative samples are filtered out by each stage, only a small number of objects would pass through the entire classifiers, reducing computation effort. A formula provided for choosing optimal thresholds at each stage to achieve some desired false positive and false negative rate. In the field of statistics, where AdaBoost is more commonly applied to problems of moderate dimensionality, early stopping is used as a strategy to reduce over fitting. A validation set of samples is separated from the training set, performance of the classified on the samples used for training is compared to performance on the validation samples, and training is finished if performance on the validation sample is seen to decrease even as performance on the training set continues to improve.

III. Architecture

The input video is converted into frames. The frames were preprocessed to remove the noises from the frames. The objects are detected from the frames by initially identifying the object regions by applying Background subtraction. The thresholding operation and morphological operations were applied to extract the object alone. The measurements of the regions that were identified as objects were calculated and based on the calculated values the objects in the video is tracked by plotting a rectangle around the object. The obtained object pixels were then grouped and the identified positions were then represented in different color inorder to obtain them exactly. Finally the performance of the process is measured by measuring the error rate and accuracy [13]. The measure performance of the process shows that the proposed method is capable for the identification of the object positions in a accurate manner compared to the other existing methods used for the identification of the object positions in the video which is due to the intensity based feature values extracted from the image that denotes the object related information in a better manner.

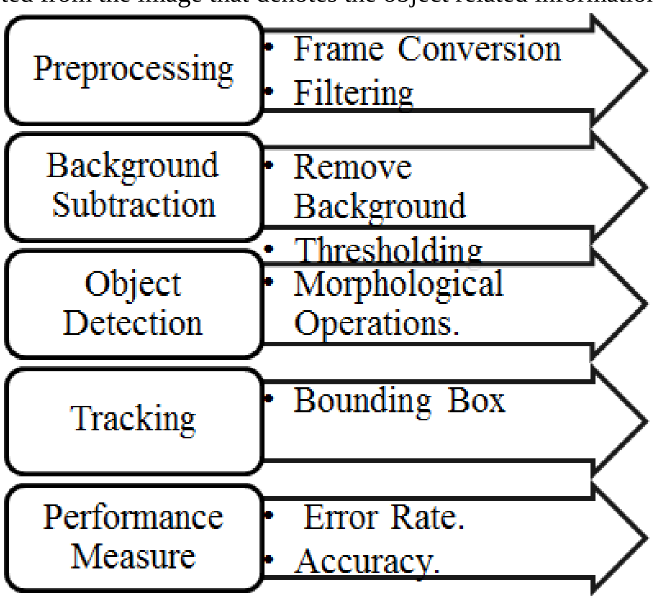

Fig. 1: Architecture for object detection

IV. Object detection

A. Key frame extraction

Key frame extraction is starting and ending points of smooth transition. The frames their position time to measured in frames. The key frame extraction method based on the similarity between contiguous frames. The similarity calculated is using the same method of assignment 2.the reference and each video frame being checked are first divided into 7*7 blocks. Then the mean and standard deviation of each block for each color of RGB is calculated. Index is the square root of the sum of the square of the difference between the mean and SD of each block. A similarity which is bigger than a preset threshold indicates a key frame is found. The threshold is set by looking the similarity value between each pair of contiguous frame. It is assumed that the first frame in the video is also the first key frame.

B. Robust principal component analysis

Principal component analysis is to separate moving object from the background. But without mechanism of robust analysis. The moving object can absorbed in the background model. The background sequence is then modeled by a low rank subspace that can gradually change over time. While the moving foreground objects constitute the correlated sparse outliers.

$$M= L0 + S0 \qquad (1)$$

L0 - Low rank
S0 – Sparse

C. Viola Jones object detection

Viola Jones can be trained to detect a variety of object classes. It was motivated primarily by the problem of face detection. This algorithm is implemented in OPENCV as CVHaar Detect objects. Viola Jones to detect human face's, noses, eyes, mouth and upper body. Use the train Case Cade object Detector function to train a custom classifier to use with the system objects [14]. It's used method for real time object detection. Its training is slow but detection is very fast. Each image contains 10-50 thousands locs/scales. Faces are rare 0-50 per image -1000 times as many non faces or faces.

D. Support vector machine

Machine learning support vector machine are supervised learning models with associated learning algorithms that analysis images and recognize patterns, used for classification and regression analysis. Set of training examples each marked belonging to one of the two categories : an SVM training algorithm builds a model that assigns new examples into one category or the other making in a non probabilistic binary linear classifier [17]. SVM constructs a hyper plane or set of hyper plane s in a high or infinite dimensional space. It can be used for classification, regression or other tasks. A good separation is achieved by the hyper plane that has largest distance to the nearest training data points of an class [16].

E. AdaBoost

Adaptive Boosting is a machine learning meta-algorithm. It can be used in conjunction with many other types of learning algorithms to improve their performance. The output of the other learning algorithms is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. In some problems, however, it can be less susceptible to the over fitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing the final model can be proven to converge to a strong learner. While every learning algorithm will tend to suit some problem types better than others, and will typically have many different parameters and configurations to be adjusted before achieving optimal performance on a dataset, AdaBoost is often referred to as the best out-of-the-box classifier.

F. Morphological operation

Morphological is a theory and technique for the analysis and processing of geometrical structures, based on set theory, lattice theory, topology, and random functions. MM is most commonly applied to digital images, but it can be employed as well on graphs, surface meshes, solids, and many other spatial structures. Topological and geometrical continuous-space concepts such as size, shape, convexity, connectivity, and geodesic distance, were introduced by MM on both continuous and discrete spaces. MM is also the foundation of morphological image processing, which consists of a set of operators that transform images according to the above characterizations. The basic morphological operators re erosion, dilation, opening and closing. MM was originally developed for binary images, and was later extended to grayscale functions and images. The subsequent generalization to complete lattices is widely accepted today as MM's theoretical foundation

G. Thresholding operation

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. Color images can also be threshold. One approach is to designate a separate threshold for each of the RGB components of the image and then combine them with an AND operation. This reflects the way the camera works and how the data is stored in the computer, but it does not correspond to the way that people recognize color. Therefore, the HSL and HSV color models are more often used. It is also possible to use the CMYK color model.

V. CONCLUSION

A effective tracking algorithm that identifies the objects in the video based on background subtraction process. Uses thresholding to separate object and non object regions. The proposed method handles the problem of abrupt motion and scalability effectively compared to the other methods used for the tracking of objects

REFERENCES

[1] Xiaowei Zhou, Student Member, IEEE, Can Yang, and Weichuan Yu," Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation,"
[2] Chun-Te chu, Jenq- Neng Hwang, Hung-Ipai, Kung-Ming LAN, "Robust video object tracking based on multiple kernals with projected Gradients,"
[3] Chun-Te Chu, Jenq-Neng Hwang Shen-Zheng Wang, Yi-Yuan Chen Human, "Tracking by Adaptive Kalman Filtering and Multiple Kernels Tracking with Projected Gradients,"
[4] Jiangxiong Fanga, Jie Yanga, Huaxiang Liub,"Efficient and robust fragments-based multiple kernels tracking,"
[5] Wojek and B. Schiele, "A performance evaluation of single and multi feature people detection,"
[6] Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multi scale, deformable part model,"
[7] Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection,"
[8] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling,"
[9] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features,"
[10] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west,"
[11] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection,"
[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features,"
[13] L. D. Bourdev and J. Brandt, "Robust object detection via soft cascade,"
[14] C. Zhang and P. Viola, "Multiple-instance pruning for learning efficient cascade detectors,"
[15] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second,"
[16] B. Heisele, T. Serre, S. Prentice, and T. Poggio, "Hierarchical classification and feature reduction for fast face detection with support vector machines,"
[17] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake, "Efficient face detection by a cascaded support vector machine expansion,"