International Journal of Advanced Computer Research (ISSN (print):2249-7277 ISSN (online):2277-7970)
Volume-3 Number-2 Issue-10 June-2013

A Study of Video Object Tracking based on Automatic Background Segmentation and updating using Different Technique

Pushpender Prasad Chaturvedi[1], Amit Singh Rajput[2], Aabha Jain[3]

Abstract

Video object tracking play an important role in security surveillance in current scenario. The explosion of successful digital device, the ease of use of high quality and economical video cameras, and the increasing need for computerized video analysis has generated a great deal of interest in video tracking methods. There are three techniques for video analysis: exposure of interesting moving target, tracking of such target from frame to frame, and analysis of target tracks to identify their activities. The successful video object tracking system faced a problem of false detection of moving video object. The false video object detection arises due to drastic change of background of moving video. For the maintenance of background updating various authors proposed a method for automatic background updating. In this paper we study of different video object tracking method using background updating factor.

Keywords

Video Tracking, Noise Filter, Segmentation.

I. Introduction

Background segmentation plays an important role in video object tracking. The automatic changing of background creates a difficulty for capturing and tracking of object, due to automatic background loss of frame and generates error. The generation of error creates a false detection of object tracking. Video tracking is one of the most important applications in computer vision, and has been widely applied to traffic surveillance system, suspicious person monitoring system etc. In practical application, since the camera moves and rotates, it needs to track objects in a dynamical background.

Pushpender Prasad Chaturvedi, Department of Electronics &Communication, RITS, Bhopal, India.

Amit Singh Rajput, Department of Electronics & Communication, RITS, Bhopal, India.

Aabha Jain, Department of Electronics & Communication, RITS, Bhopal, India.

How to select the initial target objects automatically and establish objects' motion model, and how to update object and background models at each frame are the key in real-time visual tracking with an active camera. Recent years have seen significant progress in background segmentation using portion and clustering technique for video object tracking [1,2].
Background segmentation is to exploit features in a low-dimensional space for object detection. However, the computational complexity is likely to increase significantly as a result of low dimensionality of features. Since object tracking can be posed as a binary classification problem with the goal to separate the target object from the background, a discriminative object representation scheme greatly facilitates this task. Therefore, feature selection is of crucial importance for generating an effective low-dimensional discriminative subspace. The creation of noise due to frame loss of video and free environments of camera distance of object. Noise induced a problem during updating a background process of video tracking. Video segmentation are crucial factor in video tracking, the part of segmentation generates temporal video segmentation [3]. The majority of algorithms process uncompressed video. Usually, a similarity measure between successive images is defined. When two images are sufficiently dissimilar, there may be a cut. Gradual transitions are found by using cumulative difference measures and more sophisticated threshold schemes. Based on the metrics used to detect the difference between successive frames, the algorithms can be divided broadly into three categories: pixel, block-based and histogram comparisons. Pair-wise pixel comparison evaluates the differences in intensity or colour values of corresponding pixels in two successive frames. A step further towards reducing sensitivity to camera and object movements can be done by comparing the histograms of successive images. The idea behind histogram-based approaches is that two frames with unchanging background and unchanging (although moving) objects will have little difference in their histograms. In addition, histograms are invariant to image rotation and change slowly under the variations of viewing angle and scale[5,6]. The segmentation process attempt form two or more process such as edge detection, feature extraction etc. The rest of

paper is organized as follows. In Section II discuss related work of video tracking. The Section III video background process. The section IV video noise filterand video segmentation followed by a conclusion in Section V.

II. Related Work

In this section we discuss some related work with video object tracking method based on filter and segmentation method. The process of video segmentation and updating process proposed by different approach.Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn and Andrew Zisserman entitled "The PASCALVisual Object Classes (VOC) Challenge"Authors has evaluated methods for both detection and classification which analyses whether the methods are statistically different, what they are learning from the images, and what the methods find easy or confuse[1]. The objectives of the VOC challenge are first to provide challenging images and high quality annotation, together with a standard evaluation methodology and second to measure the state of the art each year (the competition component). The goal of the VOC challenge is to investigate the performance of recognition methods on a wide spectrum of natural images.Karim Ali, David Hasler and Francois Fleuret entitled "FlowBoost – Appearance Learning from Sparsely Annotated Video" propose a learning method to learn a complex appearance model from a sparsely labeled training video with temporal consistency [2]. Authors use repeatedly a Boosting procedure to improve appearance-based model. This method is demonstrated to reduce the labeling requirement by one to two orders of magnitude. A temporal consistency has been used to label the unlabeled data which iteratively improves the appearance based classifier. A novel approach that propagates a sparse labeling of a training video to every frame in a manner consistent with the known physical constraints on target motions is proposed by the authors. ZdenekKalal, ZdenekKalal and KrystianMikolajczyk entitled "P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints" a novel paradigm for training a binary classifier from labeled and unlabeled examples that we call P-N learning is proposed [3]. The learning process is guided by positive (P) and negative (N) constraints which restrict the labeling of the unlabeled set. P-N learning evaluates the classifier on the unlabeled data, identifies examples that have been classified in contradiction with structural constraints and augments the training set with the corrected samples in an iterative process. P-N learning is applied to the problem of on-line learning of object detector during tracking. The formalized P-N learning theory enables to guide the design of structural constraints that satisfy the requirements on the learning stability. Joao Carreira, Fuxin Li and CristianSminchisescu entitled "Object Recognition by Sequential Figure-Ground Ranking" a problem for segmentation and recognition in different categories of objects in images is proposed [4]. An approach to visual object-class segmentation and recognition based on a pipeline that com-bines multiple figure-ground hypotheses with large object spatial support generated by bottom-up computational processes that do not exploit knowledge of specific categories and sequential categorization based on continuous estimates of the spatial overlap between the image segment hypotheses and each putative class., is presented. The recognition techniques that estimate the spatial layout of objects are classified as bottom-up or data-driven and top-down or model-based. Bottom-up recognition techniques use no prior shape knowledge to obtain the object regions. They often either categorizes among a set of predefined region hypotheses or directly classify pixels. Another set of bottom-up approaches decides the object category directly at the level of image pixels or superpixels based on features extracted over a supporting neighborhood. Top-down methods produce object segmentations that are often qualitative and can miss image detail. ZdenekKalal, KrystianMikolajczyk and Jiri Matas entitled "Tracking-Learning-Detection" problem of tracking of an unknown object in a video stream, where the object changes appearance frequently moves in and out of the camera view is proposed [5]. A novel tracking framework (TLD) that explicitly decomposes the long-term tracking task into tracking, learning and detection is proposed by the authors. The tracker follows the object from frame to frame. The detector localizes all appearances that have been observed so far and corrects the tracker if necessary. The learning estimates detector's errors and updates it to avoid these errors in the future. When a video is streamed and is processed at frame-rate process runs indefinitely long then it is referred as long term tracking. To enable the long-term tracking, there are a number of problems which need to be addressed. The key problem is the detection of the object when it reappears in the camera's field of view. The long-term tracking can be approached either from tracking or from detection perspectives. Tracking algorithms estimate the object motion. Trackers require only initialization, are fast and produce smooth trajectories. An object detector can be trained from a single example and an unlabeled video stream using the following strategy: (i) evaluate the detector, (ii) estimate its errors by a pair of experts, and (iii) update the classifier.SamueleSalti, Andrea Cavallaro and Luigi Di Stefano entitled "Adaptive Appearance Modeling for Video Tracking: Survey and Evaluation" a unified conceptual framework for appearance model adaptation that enables a principled comparison of different approaches is proposed [6]. A key component for achieving long–term tracking is the tracker's capability of updating its internal representation of targets (the appearance model) to changing conditions. In this paper an extensive experimental comparison of trackers that perform appearance model adaptation has been conducted. Long–term tracking in real-world conditions is made difficult by several factors, including illumination and pose changes, occlusions, deformable targets, distracters and clutter. Video trackers rely on an internal representation of target appearance, the appearance model, which is compared to measurements extracted from incoming frames at candidate target positions to estimate the most likely target location. To create the appearance model and the measurements, trackers project image regions at candidate target positions onto lower dimensionality feature spaces that highlight relevant information for the tracking task. A unified conceptual framework that identifies the common building blocks of trackers that perform adaptive appearance modeling is proposed. The frame-work is general and can accommodate in its model also video trackers designed to work with a fixed appearance model.Qing Wang, Feng Chen, WenliXu and Ming-Hsuan Yang entitled "Object Tracking via Partial Least Squares Analysis" a problem of object tracking as a binary classification problem [7]. Object tracking is an important problem in image analysis with numerous applications. It is concerned with low-level visual processing and high-level image analysis, and is widely used in image understanding, human-computer interaction, surveillance, and robotics, to name a few. In this problem the correlation of object appearance and class labels from foreground and background is modeled by partial least squares (PLS) analysis, for generating a low-dimensional discriminative feature subspace. To solve this problem an algorithm is proposed which exploits both the ground truth appearance information of the target labeled in the first frame and the image observations obtained online, thereby alleviating the tracking drift problem caused by model update. Authors proposed a tracking algorithm in which an object is represented by multiple appearance models learned online using partial least squares analysis. The proposed algorithm utilizes an adaptive discriminative representation to account for the nonlinear appearance change of an object over time. To reduce tracking drift, a two-stage particle filtering method is presented which makes use of both the static appearance information obtained at the outset and image observations acquired online. Yi We I and Zhao Long entitled "Robust objects tracking algorithm based on adaptive background updating" an algorithm called Continuously Adaptive Mean Shift (CAMSHIFT) is proposed to solve the false tracking problem [8]. It uses the disparity of global and local motion to detect the motion area. Then, it segments each object by an improved K-Mean clustering algorithm. Finally, it tracks the object by the improved adaptive background updating CAMSHIFT algorithm continuously in real time. Visual tracking is one of the most important fields in computer vision, and has been widely applied to traffic surveillance system, suspicious person monitoring system, etc. In practical application, since the camera moves and rotates, it needs to track objects in a dynamical background. How to select the initial target objects automatically and establish objects' motion model, and how to update object and background models at each frame are the key in real-time visual tracking with an active camera. The advantage of this algorithm is simplicity, but it is only applicable to the situation of camera fixed and cannot obtain complete target object in complex environment. Continuously Adaptive Mean Shift algorithm (CAMSHIFT) is a popular algorithm for visual tracking, providing speed and robustness with minimal training and computational cost. It bases on Mean Shift, transforms the visual tracking problem to the cost function's extreme value problem and can fit the real-time requirement. While it performs well with a fixed camera and static background scene, it can fail rapidly when the camera movessince it relies on static models of both background and the tracked object.

III. Video Background Process

The background of video plays an important role in video segmentation and object tracking. The automatic background updating of video increase the efficiency of video object tracking and reduces the frame loss of video. Various researchers proposed a background updating algorithm for video tracking some are performs better performance for video tracking. A general background subtraction algorithm applies a Kalman filter (or α-blending) to the pixel intensities to find the background.

$$B_{t+1} = B_t + (\alpha_1(1 - M_t) + \alpha_2 M_t)D_t, \quad (1)$$

Frame and B_t, and M_t is a binary moving object hypothesis mask. Where B_t, represents the background model at time t, D_t is the difference between the present such an approach works well when foreground objects appear infrequently, but when the background is occluded by an object for a significant time, the algorithm begins to fail. Another problem is that Mt is usually generated from D_t by thresholding and applying morphological operators. Such self-feedback can make the filtering unstable. For ex-ample, a single detection failure or a sudden illumination change can result in a permanent failure (or a ghost) which may even grow in size until it covers up the entire image. Sudden illumination changes commonly occur in many field video images because most video cameras have an auto-iris feature. Various augmentations have been applied to the back-ground subtraction, for example, to use temporal median instead of the α-blending [4]. More recently, Batista et al. introduced various augmentations including the use of multi-layer background models and dynamic thresholding [2]. Such augmentations significantly improve the robust-ness, but the problem of self-feedback is still there. Therefore, we incorporate an external cue (corner features) to generate more robust M_t. In addition, we also made the following modifications to Equation 1:

•the temporal median approach is combined with the α-blending;

•an illumination correction procedure is added to deal with sudden/temporary illumination changes; and

We use an update equation

$$B_{t+1} = \begin{cases} I_c(B_t) M_t = 1 \\ I_c((1-\alpha)B_t + \alpha N_t \quad M_t = 0, \end{cases} \quad (2)$$

Where $I_c()$ is an illumination-correction function and N_t is the temporal median of the recent, say 15, frames. Note that our background update rate is about 2 frames per second and the 15 frames spans about 7 to 8 seconds.

The illumination-correction is applied to each of the RGB value since the auto-iris can also change the color distribution (hue):

$$I_c(R, G, B) = (k_R R, k_G G, k_b B), \quad (3)$$

Where,

k_R, k_G&k_B are determined by voting on R_C/R, G_C/G, and B_C/B over all the pixels in the images. (R_C, G_C, B_C) are the pixel values of the current frame.

For M_t, we start with the standard procedure which is to 1) threshold the difference, 2) apply morphological operators (or threshold after over-smoothing), and 3) perform connected component analysis to fill holes, remove small regions, and find object blobs. After the object blobs are found, we apply an additional validation step to remove the ghosts. We assume that within all the non-ghost foreground region there exists at least one valid corner, i.e., a corner feature which is not found from the background image. For more details on the valid corner, The illumination challenge caused by an auto-iris camera. The two white vehicles in the bottom changes the entire scene darker and it causes significant false alarms. However, the error is minimized by applying the illumination correction. Here also discuss another background updating model. Each pixel in the scene is modeled by a mixture of k Gaussian distributions. The probability that a certain pixel has a value of X_N at time N can be written as

$$p(X_N) = \sum_{j=1}^{k} w_j \eta(X_N; \theta_j) \ldots\ldots(4)$$

Where w_k is the weight parameter of the k^{th} Gaussian component. $\eta(X;\theta_k)$ is the Normal distribution of k^{th} component represented by

$$\eta(X;\theta_k) = \eta(X;\theta_k,\textstyle\sum_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\sum_k|^{\frac{1}{2}}} e^{\frac{1}{2}(X-\mu_k)^T \sum_k^{-1}(X-\mu_k)} \ldots\ldots(5)$$

Where μ_k is the mean and $\sum_k = \sigma_k^2 I$ is the covariance of the k^{th} component [8].

The K distributions are ordered based on the fitness value ${}^{w_k}/_{\sigma_k}$ and the first B distributions are used as a model of the background of the scene where B is estimated as

$$B = arg_b min\left(\textstyle\sum_{j=1}^{b} w_j > T\right) \ldots\ldots(6)$$

The threshold T is the minimum fraction of the background model. In other words, it is the minimum prior probability that the background is in the scene. Background subtraction is performed by marking a foreground pixel any pixel that is more than 2.5 standard deviations away from any of the B distributions. The first Gaussian component that matches the test value will be updated by the following update equations,

$$\hat{w}_k^{N+1} = (1-\alpha)\hat{w}_k^N + \alpha\hat{w}\text{ÊÞφṔ} \ldots\ldots(7)$$

Where ω_k is the k^{th} Gaussian component, $1/\alpha$ defines the time constant which determines change. If none of the K distributions match that pixel value, the least probable component is replaced by a distribution with the current value as its mean, an initially high variance, and a low weight parameter. According to their papers [1, 2, 3], only two parameters, α &T,

needed to be set for the system. The details of its robustness were explained in their papers [1, 2, 3]; however, with a simple discussion, we can see its incapability. Firstly, if the first value of a given pixel is a foreground object, there is only one Gaussian where its weight equals unity. With only one-color subsequent background values, it will take $log_{(1-\alpha)}(T)$ frames until the genuine background can be considered as a background and $log_{(1-\alpha)}(0.5)$ frames until it will be the dominant background component. For example, if we assume that at least 60% of the time the background is present and α is 0.002 (500 recent frames), it would take 255 frames and 346 frames for the component to be included as part of the background and the dominant background component, respectively. The situation can be worse in busy environments where a clean ckground is rare. This paper presents a solution to the problem in the next section. Secondly, ρ is too small due to the likelihood factor. This leads to too slow adaptations in the means and the covariance matrices, therefore the tracker can fail within a few seconds after initialization. One solution to this is to simply cut out the likelihood term from ρ.

IV. Video Filter and Segmentation

Filter is important tools in video tracking for estimation of frame loss and reduction of AWGN noise. Various filters are used for video processing such as Gaussian filter, kamala filter and partial least square filter. The working mode of filter in video is spatial and temporal. In this section we discuss we discuss some filter used in video tracking. Instead of extracting complex features from a connected component, the raw shape of a connected component itself is an important distinguishable feature for classifying structured video and random or irregular components[4]. Together with the shape of connected component, the surrounding area of a connected component can also play an important role for video and background classification, similarly because of the structured video and non-structured non-video surrounding areas. Neighborhood surrounding areas for video and non-video regions. We refer connected component with its neighborhood surrounding as convideo. Based on the above mentioned hypothesis, our feature vector of connected component is composed of shape and video information[8]. Detail description of the feature vector is presented below. In order to improve the segmentation results, a nearest neighbor analysis by using class probabilities is performed for rending the class label of each connected component. For this purpose, a region of 70 _ 70 (empiricallychosen) is selected from document image by keeping targeted connected component at center. The probabilities of connected points within the selected regions are already computed during classification [14].

V. Conclusion and Future Work

In this paper we study of video object tracking based on background segmentation process. Background segmentation play important role in object tracking system. The correct background updating function improved the performance of video tracking algorithm. Video background segmentation implies by different approach such as background subtraction, particle least technique and many more. Such method creates a difference between actual video and background of motion video. In the process of study we also found that noise filter process for video object tracking system, now a day various filter are used such as particle filter, Gaussian filter and kamala filter. The filtration mechanism of video improves the performance of video tracking. In future we modified the partial least square filter for processing of video object tracking.

References

[1] G. D. Hager and P. N. Belhumeur, “Efficient region tracking with parametric models of geometry and illumination,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 10, pp. 1025–1039, Oct. 1998.
[2] J. Ho, K.-C. Lee, M.-H.Yang, and D. Kriegman, “Visual tracking using learned linear subspaces,” in Proc. IEEE Comput. Vis. Pattern Recognit. Conf., vol. 1. Jul. 2004, pp. 782–789.
[3] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, “Robust online appearance models for visual tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 10, pp. 1296–1311, Oct. 2003.
[4] Z. Kalal, J. Matas, and K. Mikolajczyk, “P-N learning: Bootstrapping binary classifiers by structural constraints,” in Proc. IEEE Comput. Vis. Pattern Recognit. Conf., San Francisco, CA, Jun. 2010, pp. 49–56.
[5] J. Kwon and K. Lee, “Visual tracking decomposition,” in Proc. IEEE Comput. Vis. Pattern Recognit. Conf., San Francisco, CA, Jun. 2010, pp. 1269–1276.
[6] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo, “Robust visual tracking based on incremental tensor subspace learning,” in Proc. IEEE Comput. Vis. Pattern Recognit. Conf., Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.
[7] R.-S. Lin, D. A. Ross, J. Lim, and M.-H. Yang, “Adaptive discriminative generative model and

its applications," in Proc. Adv. Neural Inf. Process. Syst. Conf., 2004, pp. 801–808.

[8] L. Matthews, T. Ishikawa, and S. Baker, "The template update problem," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 6, pp. 810–815, Jun. 2004.

[9] X. Mei and H. Ling, "Robust visual tracking using 1 minimization," in Proc. IEEE Int. Comput. Vis. Conf., Sep. 2009, pp. 1436–1443.

[10] R. Rosipal and N. Kramer, "Overview and recent advances in partial least squares," in Latent Structures Feature Selection. New York: Springer-Verlag, 2006, pp. 34–51.

[11] R. Rosipal and L. Trejo, "Kernel partial least squares regression in reproducing kernel Hilbert space," J. Mach. Learn. Res., vol. 2, pp. 97–123, Mar. 2002.

[12] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," Int. J. Comput. Vis., vol. 77, nos. 1–3, pp. 125–141, 2008.

[13] H. Saigo, N. Kramer, and K. Tsuda, "Partial least squares regression for graph mining," in Proc. ACM SIGKDD Int. Knowl. Discovery Data Mining Conf., 2008, pp. 578–586.

[14] W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis, "Human detection using partial least squares analysis," in Proc. IEEE Int. Comput. Vis. Conf., Oct. 2009, pp. 24–31.

[15] O. Williams, A. Blake, and R. Cipolla, "Sparse bayesian learning for efficient visual tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1292–1304, Aug. 2005.

[16] H. Wold, "Partial least squares," in Encyclopedia of Statistical Science, vol. 6, S. Kotz and N. L. Johnson, Eds. New York: Wiley, 1985, pp. 581–591.