



Implementation of Crf Based Automatic Multiple Person Tracking System With Open Cv on Beagle Board

KEYWORDS

Background Subtraction, Conditional Random Field (CRF), Feature Extraction, Object Tracking, Open CV, Target Tracking.

V.Kamatchi Sundari

Research Scholar, Faculty of Electronics Engineering,
Sathyabama University, Chennai, India

M.Manikandan

Associate Professor, Anna University, Chennai, India

ABSTRACT

Tracking of multiple persons is the process of locating and following the moving people in a test video sequence. The objective of this work is to explore the use of Conditional Random Field (CRF) in the automatic detection and tracking of multiple peoples in the given video. Here the USB Camera is used to take the images (Moving Objects-Human) and transfer the captured pictures/videos from the memory source to computer (PC). Background Subtraction is used as the first module to eliminate background variations and is then followed by CRF algorithm to track and draw bounding box over the moving objects in the consecutive frames. The main purpose of this proposed work is to implement multiple objects tracking algorithm in Beagle Board-XM in which objects are detected and tracked at higher speed when compared to Matlab.

INTRODUCTION

Visual-based object detection and tracking is critical to automatically monitor object activities in video sequences. Tracking is a practice of scanning an image for an object of interest. The exercise of object tracking is pertinent in the tasks like action based recognition, video indexing, human-machine interaction, monitoring of traffic, and vehicle navigation.

Various factors like intricate object shapes, image noises, and complexity in object motion & camera action increase the complexity of tracking system.

Multiple people tracking system by Conditional Random Field (CRF) is very useful particularly in video surveillance contexts, where tracking the position of people over time might benefit tasks such as group and social activities analysis, posture estimation and anomalous detection. Multiple people tracking continue a tough task, mainly in single camera settings, noticeably due to sensor noise, changing backgrounds, heavy crowd, occlusions and similar appearance among persons.

RELATED WORK

Construction of hierarchical feature space is done by selecting the features with higher reliability to represent targets in the test scene (Wang .C & Liu .H, Gao .Y, 2014). The problems of merge, split, fragments and occlusion are discussed in which Histogram intersection testing is used to limit the tracker bounding box expansion (Liu. H & Wang. C, 2013). Probabilistic Relative Distance Comparison (PRDC) model aims to maximize the probability of a pair of true match having a smaller distance than that of a wrong match pair (Zheng.W.S, Gong.S.G & Xiangtan,2011). The problem addressed in this paper is matching people across non-overlapping camera views, is challenging one due to the lack of spatial and temporal constraints and large visual appearance changes. An online detection-based two-stage multi-object tracking method is introduced with a camera (Xing .J, Ai .H & Lao .S, 2009). Particle filter based method effectively deals with interacting targets which are influenced by the proximity and/or behavior of other targets (Khan .Z, Balch .T& Dellaert .F, 2009). The particle fil-

ter take in a Markov random field (MRF) motion prior that helps in maintaining the identity of targets throughout an interaction, considerably reducing tracker failures.

A novel approach for multi-person tracking by detection in a particle filtering framework is discussed (Breitenstein .M.D et al., 2009) with which objects can be tracked in occluded environment.

A detection-based three-level hierarchical association approach is introduced to robustly track multiple objects in crowded environments from a single camera (Huang .C, Wu .B & Nevatia .R,2008). A new adaptive way is proposed to integrate multi cue in tracking multiple human driven by human detections ((Yang .M. et al., 2009). Implementation of several object tracking algorithms is done with different preprocessing methods and their performances are evaluated for different video sequences (Afef .S et al., 2012). Modified Background subtraction technique is proposed to find moving objects in a video sequences (Chinchkhede .D.W & Uke.N.J, 2012). A novel algorithm is developed for real-time detection and tracking of multiple moving objects, which sequentially integrate the entropy difference method with adaptive threshold and the fast level set method (Wanhyun Cho et al., 2011). Ye Lu & Ze-Nian Li (2007) combined the various views of object to form a 3D view of the object for extraction from active video.

Open CV is an Image Processing library created by Intel and maintained by Willow Garage. Open CV, a computer vision library, is extremely popular and has considerable functionality relevant to this work (Gary Bradski & Adrian Kaehler,2008). It also supports Video4Linux, a project to support common video capture devices in Linux. While other computer vision libraries exist, Open CV is the most popular and hence is best supported online.

PROPOSED METHOD

Multi-target tracking is important for many applications such as surveillance and human-computer interaction systems. Its aim is to locate the targets, retrieve their trajectories, and maintain their identities through a video se-

quence; this is a highly challenging problem in crowded environments when the occlusions of targets are frequent. Conditional random fields (CRFs) are a class of statistical modeling method often applied in pattern recognition and machine learning Olivier Barnich & Marc Van Droogenbroeck(2009) which is used in finding the targets in our work and is followed by implementation on the Beagle board installed with a supported (manufacturer supplied or recommended) variant of Ubuntu Linux, and Open CV from source

Figure 1: Overview of main stages

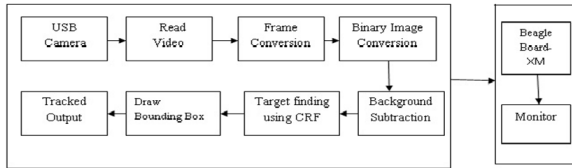


Figure 2: CRF System Architecture

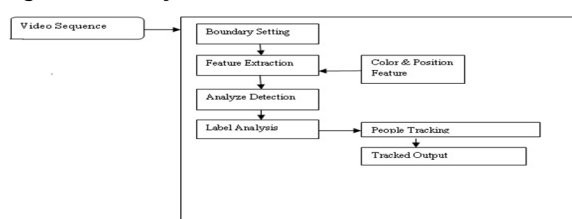


Figure 3: Implementation setup



Figure.1. shows the overview of main stages involved in our proposed system, figure.2. shows the flow for CRF and the setup used for our implementation process is given in figure.3.

Input video is captured by digital camera and then it is converted into frames. Matlab code is used to convert the image pixel values into matrix values and is stored in the work space, which are used as inputs for feature extraction. The images arranged in matrix form are bringing into CRF. Particle filter is used to study the motion of each and every particle within a medium. The coordinates (x, y, z) with reference to time can be defined as a trajectory. The trajectory can be explored to classify the modes of motion.

The background subtraction is a widely used approach for detecting moving objects in video from static camera. The underlying attitude in the approach is that of finding the difference between the current frame and a reference frame, called the background image. The background image should be a depiction of the scene with no moving objects and must be kept regularly updated so as to adjust to the varying lighting conditions and geometry settings (Wren, C et al., 1997, Olivier Barnich & Marc Van Droogenbroeck,2009).

Feature Extraction plays a major role to detect people in sequence of frames. Every object has an explicit feature like color or shape. The segmentation is performed using frame difference; the residual image can be viewed with rectangular bounding box with the dimensions of the object produced from residual image and here features are extracted by the intensity value which is used to describe the color. The intensity values corresponding to pixel val-

ues obtained from the first hit from top side, bottom side, left side and right side are stored. Then a rectangular bounding box is plotted within the limits of the values produced.

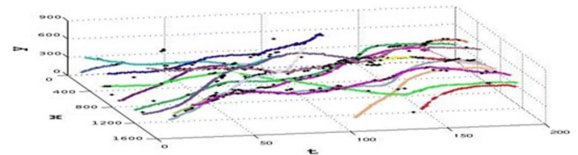


Figure 4: x-y-t Graph of Tracking Results

The above figure.4. shows the tracking results with x-y-t graph, the center of a bounding box and time is represented by (x; y) and t respectively. Every dot in the graph represents a detected bounding box, where different colors represent different tracklets.

CONDITIONAL RANDOM FIELD (CRF)

Conditional random field (CRF) is a statistical model, more often used in pattern recognition and machine learning Olivier Barnich & Marc Van Droogenbroeck(2009). CRF predicts sequences of labels for sequences of input samples.

$$p(L|R) = \frac{1}{Z(R)} \left(\prod_{(i,j)} \prod_{k=1}^{Nf_2} \Phi_k(l_i, l_j, r_i, r_j) \right) \cdot \left(\prod_i \prod_{l=1}^{Nl} \Psi_l(l_i, r_i) \right) \cdot \Omega(L) \quad (1)$$

Human detection step has to be performed on each frame of a test video. R represents the set of detection outputs. Each detection r_i consists of a set of observations, which include an occurrence time t_i (or frame number), as well as some features. In this paper, two important features are used viz., the color descriptor (h_i), and the position of the detection (X_i). X_i is calculated by projecting the bottom center of the detection bounding box into the ground plane. As color descriptor h_i , the multi-resolution color histogram is used in the HSV color space. Multi-resolution color histogram is used to reduce the quantization effects. Moreover, to avoid taking many pixels from the background, the color histograms are computed within an elliptical region enclosed in the detection bounding boxes. Thus, a detection is represented by $r_i = (t_i; X_i; h_i)$. Based on similarity measures, the task of multi-object tracking is done by linking those detections across frames. This task can be formulated as a labeling problem, where labels are assigned for detections according to the identity of the objects. The label field $L = l_i, r_i=1: N_r$ for that purpose, where l_i denotes the label identity for detection r_i . Detections corresponding to the same object should possess the same label, meaning there would be ideally one label per track. Then the label field should be identified which maximizes the posterior probability $p(L|R)$. In a traditional generative model, Maximum A Posteriori (MAP) formulation is used, and can equivalently maximize $p(R|L) p(L)$. Typically, $p(L)$ defines a prior over the label field and $p(R|L) = \prod_i p(r_i|l_i)$ denotes the data likelihood. Note that such an approach is not appropriate for association, in view of the fact that the number of classes are not known in advance and also the term $p(r_i|l_i)$ only entails one detection and not able to be defined in advance. Rather, in this paper, Conditional Random Field (CRF) is adopted.

Unlike the generative approach which represents the joint probability distribution, CRFs present the advantage that no assumptions on the dependencies among the observed variables r_i need to be specified, as the label field conditional distribution is modeled directly. Given this global

model, the main technical points to address are the definition of the functions, the learning of their parameters, and the optimization procedure.

At testing time, the goal is to find the optimal label field by maximizing the objective function given in following equation. It can be shown that it is equivalent to minimizing the energy function

$$U(L) = \sum_{i_k = t_j | \mathcal{L}_{i_k} \neq \mathcal{L}_{t_j}} \beta_{ij}^{potts} \cdot \delta(L_i - L_j) \quad (2)$$

Where $\delta(\cdot)$ denotes the Kronecker function and the potentials between each pair (also called Potts coefficient) are defined by:

$$\beta_{ij}^{potts} = \log \left[\frac{\prod_{k=1}^{N_{f_2}} p(f_k(r_i, r_j) | H_0)}{\prod_{k=1}^{N_{f_2}} p(f_k(r_i, r_j) | H_1)} \right] \quad (3)$$

If β_{ij}^{potts} is negative, then $\delta(|r_i - r_j|) = 1$, meaning that the pair (r_i, r_j) is expected to be correspond to the same object (hypothesis H_1 prevails) to an extent related to the amplitude of β_{ij}^{potts} . On the contrary, when β_{ij}^{potts} is positive, hypothesis H_0 prevails.

TRACKING PROCEDURE

Once the people are located in each individual image, it is necessary to track them across images. This is achieved using a particle tracker developed. For each image the minimization energy obtained above, the centroids of the blobs are located. These points and their frame number are then fed into the CRF tracking algorithm. To assign a track label to a minimization, it essentially finds the closest particle in the previous image and assigns that track to the current case. Constraints are set in the maximum translation a particle can move. Also, the algorithm allows a particle to disappear for several frames and still be tracked. This accounts for quick occlusions when people pass in front of each other. Furthermore if a particle appears for less than two frames, it's not tracked and assumed to be noise. This accounts for quick changes in lighting and shadows. The algorithm works fairly well in putting together tracks for individual particles. It does not get confused when people pass over each other slowly, and in some occasions when people move very quickly such that the displacement is large. However, for the most part, as long as a track is found it is sufficient to make a decision without maintaining identity.

EXPERIMENTAL RESULTS

In this section our algorithm is tested with Matlab and implements the same in Beagle Board-XM with recorded video having frame rate of 21 frames per second. Each frame size is of 340x260 pixels.

The Beagle Board is a single-board computer based development kit with low cost and on low-power from Texas Instruments. This unit includes DM3730 processor, runs on Angstrom (embedded operating system). Its architecture is designed with USB web camera. Open CV package consists of various image processing functions is installed to interface USB web camera. Our code is then implemented on the Beagle board installed with Open CV software.

On each of these boards, we installed a supported (manufacturer supplied or recommended) variant of Ubuntu Linux, and Open CV from source. Where possible, we enabled support for the ARM specific NEON single instruction, multiple data extensions to increase performance.



Figure 5: Input Video Frame



Figure 6: Boundary Setting



Figure 7: Tracking a person

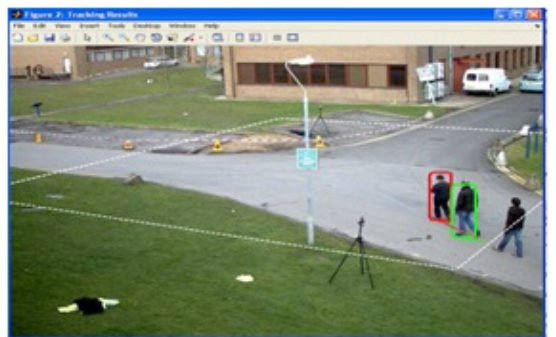


Figure 8: Tracking two persons



Figure 9: Tracking multiple persons

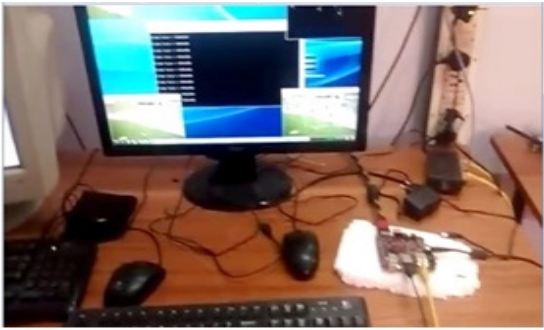


Figure 10: Snapshot of our implementation setup



Figure 11: Snapshot obtained by Open CV Using Conditional Random Field

PERFORMANCE EVALUATION

Open CV has cross platform library function, with the avid features of its used in major activities in Motion analysis. The four advantages of Open CV are speed, resource, portability and cost.

Speed: To execute in Open CV the code is directly provided from machine language code to computer language code. This ensures that the user get more image processing done for computers processing cycles, and not more just mere interpreting. Compared to other platforms with the similar program it can be found that Open CV is faster compared to the rest due to the previous explained advantages. The speed of Open CV is incredibly fast which would result in at least 30 frames per second, resulting in real-time detection.

Resource needed: The source need for an Open CV is very less compared to the other platforms. Only around 70Mb of RAM is required to process any length of program and can support all types of execution.

Portability: Windows, Linux and MacOS all support Open CV. But the main agenda of portability is that any machine OS which supports C can support Open CV.

Cost: Cost is also an important factor and plays an important role. The license to acquire Open CV is free while comparing to other which have to be bought at a high price. This enables all types of user to acquire it and use it in the effective manner.

As a result of above mentioned advantages, the video tracking was found to be effective by the s comparing the performance of Matlab with Open CV. The execution times for the algorithms on with OpenCV and Matlab are tabulated in Table 1. The performance is improved 2.3 to

4.6 times by utilizing the computing power of the onboard Beagle BoardXM core and nearly all the algorithms have met the requirements of real-time.

TABLE – 1
PERFORMANCE COMPARISON

Parameter	Open CV	MATLAB
Average number of frames processed per sec	18	5
Average execution time for a frame	2.77 sec	19.43 sec

BENCHMARKING METHODOLOGY

The Target object a fixed distance of approximately 1.5 meters from the web camera. All tests were performed indoors in a room with a constant lighting level. Timing Data each implementation was benchmarked using calls to C++'s get time of day function. For each implementation we profiled the amount of time taken for various calls within our pose estimation routine over 20 frames.

For each test, we discarded the first 20 frame chunk since this is when the camera automatically adjusts exposure and focus settings. This data was collected for 10 arbitrary 20 frame chunks and averaged to provide. The Logitech web camera we are using captures frames at a 340 x 260 pixel resolution at a maximum of 21 frames per second.

CONCLUSIONS

In this paper, a fast and robust automatic multiple people tracking system has been proposed and implemented in Beagle Board-XM with Open CV. Background subtraction is done with the input video and CRF algorithm is used to find the accurate object in the particular field and thereby objects are tracked successfully. Finally the results obtained in Open CV is compared with MATLAB and hence proved that our system detects very accurately at higher speed.

REFERENCE

- [1] Afef .S, Jammoussi & Ameni .Y.(2012), "Object Tracking System Using Camshift, Meanshift and Kalman Filter." World Academy of Science Engineering and Technology, 6:598-603. [2] Bernardin .K & Stiefelhagen .R.(2008), " Evaluating Multiple Object Tracking Performance: The Clear Mot Metrics." EURASIP Journal of Image Video Processing, 1.1:110-120. [3] Berclaz .J, Fleuret .F, Turetken .E & Fua .P.(2011), "Multiple Object Tracking Using K-Shortest Paths Optimization." IEEE Transaction on Pattern Analysis Machine Intelligence, 33(9): 1806-1819. [4] Breitenstein .M.D, Reichlin .F, Leibe .B, Koller-Meier .E & Gool .L .V. (2009), "Robust Tracking-by-Detection Using a Detector Confidence Particle Filter." IEEE International Conference on Computer Vision ICCV, 1515-1522. [5] Chinchkhede .D.W & Uke.N.J. (2012), "Image Segmentation in Video Sequences Using Modified Background." IEEE International Journal of Computer Science & Information Technology IJCSIT, 4: 93-104. [6] Gary Bradski & Adrian Kaehler.(2008), " Learning OpenCV. Computer Vision with the OpenCV Library." O'Reilly Media [7] Huang .C, Wu .B & Nevatia .R.(2008), " Robust Object Tracking by Hierarchical Association of Detection Responses." International Conference on Computer Vision ECCV, 788-801. [8] Hyung -Bok Kim & Chung.S.(2010), " A Particular Object Tracking in an Environment of Multiple Moving Objects." IEEE International Conference on Control Automation and Systems ICCAS, 92-98. [9] Khan .Z, Balch .T& Dellaert .F.(2009), " MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets." IEEE Transaction on Pattern Analysis Machine Intell.,27(11): 1805-1891. [10] Liu. H & Wang. C.(2013), "Hierarchical Data Association and Depth-invariant Appearance Model for Indoor Multiple Objects Tracking." IEEE International Conference on Image Processing ICIP 2013: 2635-2639. [11] Malhorta .D& Saskatchewan .R. (2005), "A Novel Approach for Finding the Movement of an Object in Video Sequences by an Artificial Neural Network for 2.5D Object Modeling." IEEE Conference on Computer Vision and Pattern Recognition CVPR, 67-73. [12] Namitha .M & Prabhakar.S.(2013), "Certain Approaches of Real Time Object Tracking in Video Sequences on Embedded Linux Platform." International Journal of Innovative Technology and Exploring Engineering IJITEE, 2:102-104. [13] Olivier Barnich & Marc Van Droogenbroeck(2009), "ViBe: A Universal Background Subtraction Algorithm for Video Sequences." ICASSP. [14] C. Sutton, & A. McCallum.(2007), " An Introduction to Conditional Random Fields for Relational Learning ." MIT Press. [15] H. M.Wallach. (2004), "Conditional random fields: An introduction." Technical Report MS-CIS-04-21, University of Pennsylvania. [16] Wang .C & Liu .H, Gao .Y.(2014), " Scene-Adaptive Hierarchical Data Association for Multiple Objects Tracking." IEEE Conference on Computer Vision and Recognition, 21(6): 697-712. [17] Wanhun Cho, Sunworl Kim & Gukdong Ahn.(2011), " Detection and Tracking of Multiple Moving Objects in Video Sequence Using Entropy Mask Method and Fast Level Set Method." IEEE International Conference on Computer Vision ICCV, 1-6. [18] Wren, C., Azarbayejani, A., Darrell, T., Pentland & P. Pfinder (1997), "Real-Time Tracking of the Human Body." IEEE Transactions on Pattern Analysis and Machine Intelligence. [19] Xing .J, Ai .H & Lao .S.(2009), " Multi-Object Tracking Through Occlusions by Local Tracklets Filtering and Global Tracklets Association with Detection Responses." IEEE Conference on Computer Vision and Pattern Recognition CVPR, 1200-1207. [20] Yang .M, Lv .F, Xu .W & Gong .Y.(2009), " Detection Driven Adaptive Multi-Cue Integration for Multiple Human Tracking." IEEE International Conference on Computer Vision ICCV,1554-1561. [21] Ye Lu & Ze-Nian Li (2008), " Automatic object extraction and reconstruction in active video. Pattern Recognition." The Journal of Pattern Recognition Society, 41: 1159-1172. [22] Zheng.W.S, Gong.S.G & Xiangtan (2011), " Person Re-identification by Probabilistic Relative Distance Comparison." IEEE Conference on Computer Vision and Pattern Recognition CVPR, 649-656.