Object Detection and Tracking from Video Sequence using MATLAB

S. Saranya
Research Scholar
Department of Computer Science
Muthayammal College of Arts & Science, Namakkal, Tamilnadu.

A. Anusha Priya
Associate Professor
Department of Computer Science
Muthayammal College of Arts & Science, Namakkal, Tamilnadu.

Abstract

This paper aims to detecting and tracking objects in a sequence of color images taken from a video camera. Motion detection and tracking algorithm is presented for monitoring the pedestrians in an outdoor scene from a fixed camera. A mixture of Gaussians is used to model each pixel of the background image and thus adaptive to the dynamic scene. Color chromaticity is used as the image representation, which results in the illumination-invariant change detection in a day lit environment. To correctly interpret those objects that are occluded, merged, split or exit from the scene, a scene model is created and the motion of each object is predicted. A Bayesian network is constructed to reason about the uncertainty in the tracking. The results for detecting and tracking the moving objects in the PETS sequences are demonstrated.
Keywords: Matlab, Object detection, Network, Robust

I. Introduction

A. Previous Work

Tracking non-rigid objects in real world scenes contains several difficulties for computer vision. Problems include static and dynamic occlusion, variation of lighting condition, failure of foreground detection, etc. Therefore, many successful tracking systems work only in constrained environments, in which the targets are sparsely distributed and the background is less dynamic [12]. Frame differencing is a technique widely used for the change detection in dynamic images. It compares each incoming frame with a background image and classifies those pixels of significant variation into foreground. The background can be modeled with a single adaptive Gaussian [12] and learnt during an initialization period when the scene is empty. This method is efficient only in less dynamic scenes but has difficulties with vacillating backgrounds (e.g. swaying trees), background elements moving, and illumination changes.

A more robust method is to model the background by a mixture of adaptive Gaussians [11]. However, it may fail in tracking a background pixel under fast illumination changes, e.g. flood of sunlight, shadows or artificial lights switching on/off. This causes spurious “foregrounds” and can lose targets in such cases. The reason is that most the existing applications use intensity-based image representations, e.g. (R, G, B) or I, which are the result of interaction between illumination from light sources and reflectance of object surfaces. To be able to identify and track the same object surface (e.g. a background pixel) under varying illumination, it is desirable to separate the variation of the illumination from that of the surface reflection.

In the existing algorithms that track occluded objects, Intille et al [5] compared the properties of detected foreground regions with those of the previously tracked objects. An object is identified as being occluded if two objects are found to match the same foreground region. Haritaoglu et al [3] identified the dynamic occlusion when the predicted bounding boxes of two objects overlap the same foreground regions. However, these are rule-based methods that are brittle and globally lack of consistency [10]. In addition, these algorithms did not consider static occlusions, e.g. buildings, trees, or road sign. Therefore, objects may suddenly disappear at some sites in a scene and many new objects appear at some other sites. Therefore, the trajectory of an object tends to be short and segmental.

B. Our Approach

In this paper a mixture of Gaussians is used to model each pixel of the background image and thus adaptive to the dynamic scene. The combination of color chromaticity- and intensity-based image representations results in the illumination-invariant change detection in a daylit environment. To correctly interpret those objects that are occluded, merged, split or exit from the scene, a scene model is created and the motion of each object is predicted. A Bayesian network is constructed to reason about the uncertainty and noise in the observation.

II. OBJECT AND SCENE MODELING

A. Blob Models

Each foreground blob detected at the current frame is ideally associated with an object or a group of interacting objects. It is characterized by Positions: its bounding box and coordinate of its centroid. Colors: its color template pyramid. Sizes: the number of foreground pixels. Status: allocated to an object (ALLOCATED) or not (UNALLOCATED). Which represent colors along the red to green, and blue to yellow axes [1]. The 3-D color information has been mapped to 2-D, and for each blob a 2-D m□m color histogram is generated as a template (m=16). This 2-D template needs less storage than a 3-D one, and so is faster to handle. The color template for a blob needs to be compared with that for each object that has been tracked in the previous frames. Intuitively some tracked objects will have radically different colors from others, and an efficient search is desirable. This is achieved by having a pyramid of color templates at four resolutions, and by trying to match initially at the coarsest resolution, and only going to higher resolution for objects considered similar [1].

B. Object models

Each tracked object is recorded in an object database and ideally corresponds with a blob in the new frame.
An object is described by:

- The characteristics of its matched blob
- Positions: its bounding box and coordinate of its
- centroid.
- Colors: its color template pyramid.
- Sizes: the number of foreground pixels.
- The tracking information
- Status: NEW (objects entering the scene), TERMINATED (objects leaving the scene), UPDATED (an object optimally matched to a blob), MERGED (colliding objects), SPLIT (objects colliding and then separate), CCLUDED (objects hidden by static occlusions), MISSING (cannot be interpreted).
- Dynamic model: direction, velocity, acceleration.
- History: last position, frame of being first seen, frame of being last seen.
- Prediction: predicted position (estimated by a first-order motion model), predicted status, predicted bounding box.
- Interacting object: index of another object that interacts or is assumed to interact the underlying object.

C. Scene Models

Because the camera is fixed, a scene model can be constructed for a specific camera position. Whilst this is currently done manually, we are investigating automatic methods for learning the scene model [6]. This helps reasoning about the termination and occlusion of objects by scene elements. Three types of static occlusions in a scene are identified (Fig. 1): Border occlusions (BO) due to the limits of the camera field-of-view (FOV).Long-term occlusions (LO). These are locations where objects may leave the scene earlier than expected, corresponding to the termination of a record in the object database. These occlusions often ave one side touching the border of the image and make objects leave the scene from the other side, at a distance away from the border of the image, e.g. buildings and vegetation. The long-term occlusion may also exist in the middle of an image, e.g. at the doors of a building. Without some prior knowledge of these longterm occlusions, an object disappearing at a LO may later be mismatched with other objects that are present nearby.

Short-term occlusions (SO). These are the locations where an object may be temporarily occluded by a static occlusion, e.g. a tree or a road sign. Prior knowledge of these occlusions helps avoid missing existing objects and creating "new" objects. All the occlusions are stored in an occlusion database.

Each occlusion is characterized by its:□Type (BO, LO or SO).Bounding box, representing its location and dimension. The overlap of these static occlusions with the predicted bounding box of an object can be used to predict object termination and occlusion. Currently a rectangular bounding box is being used for each static occlusion to LO & SO.

A more accurate representation of these occlusions, e.g. using polygons, is Straightforward but not so important here, because these occlusion bounding boxes are used only for the prediction of some events (termination and occlusion). The determination of such events also depends on the result of tracking (e.g. an object fails to find a corresponding blob), because objects may pass in front of an occlusion.

D. Model-based prediction

After the status of an object is determined at each frame, the object is subject to a process of status prediction which is based on a first-order motion model and scene model.

An object is labelled as PREDICTIVE TERMINATED, if its predicted bounding box overlaps a long-term occlusion (LO) or the outer of the border occlusion (BO). An object is labelled as PREDICTIVE OCCLUDED, if its predicted bounding box overlaps a short-term occlusion (SO). An object is labelled as PREDICTIVE MERGED, if its predicted bounding box overlaps that for another object. Then the index of the second object is recorded as the "Interacting Object" of the first object. An object is labeled as PREDICTIVE SPLIT, if its status is MERGED and its predicted bounding box does not overlap that of its "Interacting Object".

III. OBJECT TRACKING

A staged and ordered matching process has been used to make a correspondence between the blobs detected at the current frame and the objects tracked at the previous frame.

A. Blob to object matching

The visible blobs are first compared with the tracked objects. To determine blob-to-object correspondence, a match score for every blob and object combination is computed as the weighted sum of several distance measures, as in [5]. Each distance measure reflects the difference of some characteristic between the blob and object. It is limited by an allowable tolerance for possible matching and normalized by the tolerance value. Because the objects are assumed to have no drastic change in some selected characteristics between two consecutive frames, the distance measure is low for a possible match. The characteristics that have been considered include:

1) Predicted position. It is given the greatest weight. For an object identified as NEW at the previous frame, this characteristic is replaced by the object position.
2) Color. The distance measure between the color template, T, of a blob and that, T, of an object is calculated.

B. Unmatched Objects

After sorting out the UPDATED objects, there remain some objects that do not have a correspondence with any detected blob. This may arise from the objects leaving the scene, the occlusion of objects by scene elements, the merging of multiple objects, or the failure of foreground detection. The ambiguity here can be partly relieved by using domain knowledge. For example, if it is known that an unmatched object was very close to a long-term occlusion in the last frame, it is quite possible that this object left the scene in the current frame. However, there exist uncertainties in such domain knowledge:

– Not all of the objects close to a long-term occlusion will leave the scene (they may walk in front of it).
– An object may merge with another one near the border of a long-term occlusion.
– The foreground detection may fail (i.e. the corresponding blobs are missing) at any position in a scene.

Given the uncertain and incomplete information, the object tracking can be inferred through a process of deduction. A Bayesian network [8] is a framework for representing and using domain knowledge to perform probabilistic inference. It is a directed acyclic graph in which nodes represent random variables and arcs represent causal connections among the variables. Associated with each node is a probability table that provides conditional probabilities of the node's possible states given each possible state of its parents. In the case that a node has no parents, conditional probabilities degenerate to priors. When values are observed for a subset of the nodes, posterior probability distributions can be computed for any of the remaining nodes. Bayesian networks have been used in object tracking and behavior identification.

C. Unmatched blobs

After checking the object database, the detected blobs that have not been interpreted are most likely split or new objects. Another Bayesian network has been used to infer the posterior probabilities of the query nodes, "new at t" and "split at t", given the observed values of the evidence nodes. In order to have efficient computation, the "distance to BO or LO" is approximated with a set of discrete values: touching, close and far. It is noted that most of the split objects was previously merged unless the objects entered the scene in a group. This is reflected in the high conditional probability of "split at t" given "merged at t-1". For a split object, it is most likely labeled as POTENTIALLY SPLIT in the blob-to-object matching stage and tends to be significant smaller than the merging group. Once the status of all the objects is determined; the records in the object database need to be updated. The record of an UPDATED, SPLIT or NEW object is replaced by the characteristics of the corresponding blob. For a MERGED or OCCLUDED object, its position is updated according to its visible history and the first order motion model; its color and size remain unchanged. The record of a MISSING object is kept unchanged until this object is re-tracked or automatically terminated after "missing" for three consecutive frames.

IV. RESULTS

To assess the significance of the detection and tracking algorithm, we have applied it to the PETS2001 sequences which include significant lighting variation, occlusion and scene activity. The sequences were spatially sub-sampled to half-PAL (384*288 pixels) and mporal sub-sampling has been investigated in our experiments. The results presented below use 2.5 fps for foreground detection and 5 fps for object tracking. These rates provided a reasonable trade-off between computational efficiency and robust detection and tracking. The corresponding result sequences ("31_1.avi" and "31_2.avi" using intensity-based model, and "31_3.avi" and "31_4.avi" using the combination of color- and intensity-based models) start at frame 1500 and end at frame 3000. The foreground pixels in the color-based results are those that go beyond [-3.5 , +3.5] of the most probable Gaussians. The foreground pixels in the intensity-based results arise from a global threshold on the difference between the observation and the mean of the most probable Gaussian. The threshold level is selected as 10% of the maximum intensity so as to produce “blobs” of similar sizes to those in the corresponding color-based results. In order to rule out isolated “foreground” pixels and fill gaps and holes in “foreground” regions, a 13 closing (dilation-erosion) opera merged at t-1 split at t new at t distance to BO or LO unmatched blob potentially split predictive split blob size decreases.

A. *Object Detecting*

The tracking results using Dataset 2 (camera 1, testing) at 5 fps. The corresponding result sequence ("31_5.avi") starts at frame 1 and ends at frame 701. The results of the first five frames are noisy and not included, because the Gaussian mixture model needs to learn the initial parameters for each distribution. Fig. 6 shows the manually selected occlusions in the scene, in which No. 0, 2 and 3 are short-term occlusions and No. 1 is a long-term occlusion. The building in the right is a potential long-term occlusion but not used here. When an object (No. 0,white bounding box) is passing by a short-term occlusion (No. 0, in pink bounding box). At frame 351 (Fig.7(a)), the predicted bounding box (invisible here) of object 0 overlaps occlusion 0, the predicted status is set to PREDICTIVE OCCLUDED. At frame 386 (Fig. 7(b)) when no corresponding blob is found, object 0 is determined as OCCLUDED by the Bayesian network. Its position is updated according to the first-order motion model based on its visible history. Therefore, the object bounding box (grey) is not observed but estimated. When one object is merged with another. At frame 416 when the predicted bounding boxes (invisible) of objects 0 and 2 overlap, the predicted status is set to PREDICTIVE MERGED for both the objects. At frame 426 (Fig. 8(b)) when only one blob is detected, object 0 is matched to that blob because its properties are less influenced by the merging. Object 0 is determined as a normal UPDATED object. On the other hand, object 2 is determined as MERGED by the Bayesian network and its interacting object is set to No. 0. Its position is updated.

Predicted according to the first-order motion model based on the visible history (note the grey bounding boxes in Figs. 8(b)(c)). At frame 441 object 2 are matched to a newly detected blob and thus retraced as an UPDATED object. Table 2 shows the tracking errors, when the objects are interacting with each other or the scene elements, for the first four testing sequences (Datasets 1-2, cameras 1-2). The result indicates that the current algorithm is proper to track two interacting objects or object groups. Its performance degrades when multiple objects are clustered in a local region, such as that in Dataset 3. Events Merged Split Occluded.

V. CONCLUSIONS

The combination of the colour- and intensity-based Gaussian mixture models can better adapt to fast illumination changes when detecting foregrounds. The scene model and motion prediction provide relatively reliable evidence in inferring and tracking objects through Bayesian networks, especially when ambiguity in observation arises. Future work includes considering multiple objects that interact within a group, using dynamic Bayesian networks and even continuous variables to infer object status, using multi-view co-operation to interpret the incomplete observation from each single view.

REFERENCES

[1] S. A. Brock-Gunn and T. Ellis, “Using colour templates for target identification and tracking”, Proc. BMVC’92, 207- 216, 1992.
[2] H. Buxton and S. Gong, “Visual surveillance in a dynamic and uncertain world”, Artificial Intelligence, 78:431-459, 1995.
[3] M.Balaji, E.Aarthi, K.Kalpana, B.Nivetha, D.Suganya “Adaptable and Reliable Industrial Security System using PIC Controller” 2017/5, Journal International Journal for Innovative Research in Science & Technology, Volume 3, Issue 12, Page 56-60.
[4] Haritaoglu, D. Harwood and L. S. Davis, “W4: real-time surveillance of people and their activities”, IEEE Trans. On PAMI, 22(8): 809-830, 2000.
[5] T. Huang, D. Koller, J. Malik, et al., "Automatic symbolic traffic scene analysis using belief networks", Proc. AAAI'94, 966-972, 1994.
[6] A.S.Syed Fiaz, R.Pushpapriya, S.Kirubashini & M.Sathya “Generation and allocation of subscriber numbers for telecommunication“,March2013, InternationalJournal of Computer Science Engineering and Information Technology Research, Vol No: 3; Issue No: 1, pp. 257-266.
[7] M. Usha, J. Akilandeswari and A.S.Syed Fiaz “An efficient QoS framework for Cloud Brokerage Services”, Dec. 2012, International Symposium on Cloud and Service Computing, pp: 76-79, 17-18, IEEE Xplore.
[8] A.S.Syed Navaz, A.S.Syed Fiaz, C.Prabhadevi, V.Sangeetha & S.Gopalakrishnan “Human Resource Management System” Jan – Feb 2013, International Organization of Scientific Research Journal of Computer Engineering, Vol 8, Issue 4, pp. 62-71.
[9] A.S.Syed Navaz, S.Gopalakrishnan & R.Meena “Anomaly Detections in Internet Using Empirical Measures” February 2013, International Journal of Innovative Technology and Exploring Engineering, Vol 2 – Issue 3. pp. 58-61.
[10] A.S.Syed Navaz, H.Iyyappa Narayanan & R.Vinoth.” Security Protocol Review Method Analyzer (SPRMAN)”, August – 2013, International Journal of Advanced Studies in Computers, Science and Engineering, Vol No – 2, Issue No – 4, pp. 53-58.
[11] S. S. Intille, J. W. Davis and A. F. Bobick, “Real-time closed-world tracking”, Proc. CVPR’97, 1997.
[12] Makris and T. Ellis, “Finding paths in video sequences”, Proc. BMVC’2001, pp. 263-272, 2001.
[13] K. Murphy, “Bayes net toolbox for Matlab”, http://www. cs.berkeley.edu/~murphyk/Bayes/bnt.html.
[14] J. Pearl, Probabilistic Reasoning in Intelligent Systems:Network of Plausible Inference, Morgan Kaufmann, SanMateo, CA, 1988.
[15] P. Remagnino, T. Tan and K. Baker, "Agent orientated annotation in model based visual surveillance", Proc. ICCV'98, 857-862, 1998.
[16] Sherrah and S. Gong, "Tracking discontinuous motion using Bayesian inference", Proc. ECCV'2000, 150-166, 2000.
[17] G.Naveena, A.Anusha Priya, A Certain Investigation on Cluster Based Medium Access Control and QoS Aware Routing Protocol for Heterogeneous Networks, 2016, International Journal for Scientific Research & Development, Volume 4, Issue 7, Pages 1117-1122.
[18] A.S.Syed Navaz & G.M. Kadhar Nawaz, “Ultra-Wideband on High Speed Wireless Personal Area Networks” August – 2014, International Journal of Science and Research, Vol No – 3, Issue No – 8, pp.1952-1955.
[19] A.S.Syed Navaz & Dr.G.M. Kadhar Nawaz & A.S.Syed Fiaz “Slot Assignment Using FSA and DSA Algorithm in Wireless Sensor Network” October – 2014, Australian Journal of Basic and Applied Sciences, Vol No –8, Issue No –16, pp.11-17.
[20] A.S.Syed Navaz, S.Ravishankar & M.Sudarson, “Tracking Objects Using Multiple Acoustic Video Systems DISCOVERY The International Daily journal, September 2015, Volume – 37, Issue - 168, pp - 59-64.

[21] A.S.Syed Navaz & A.S.Syed Fiaz, “Load Balancing in P2P Networks using Random Walk Algorithm” March – 2015, International Journal of Science and Research, Vol No – 4, Issue No – 3, pp.2062-2066.
[22] A.S.Syed Navaz, J.Antony Daniel Rex, S.Jensy Mary. “Cluster Based Secure Data Transmission in WSN” July – 2015, International Journal of Scientific & Engineering Research, Vol No - 6, Issue No - 7, pp. 1776 – 1781.
[23] A.S.Syed Navaz, P.Jayalakshmi, N.Asha. “Optimization of Real-Time Video Over 3G Wireless Networks” September – 2015, International Journal of Applied Engineering Research, Vol No - 10, Issue No - 18, pp. 39724 – 39730.
[24] A.S.Syed Fiaz, N.Asha, D.Sumathi & A.S.Syed Navaz “Data Visualization: Enhancing Big Data More Adaptable and Valuable” February – 2016, International Journal of Applied Engineering Research, Vol No - 11, Issue No - 4, pp.–2801-2804.
[25] A.S.Syed Navaz & Dr.G.M. Kadhar Nawaz “Flow Based Layer Selection Algorithm for Data Collection in Tree Structure Wireless Sensor Networks” March – 2016, International Journal of Applied Engineering Research, Vol No - 11, Issue No - 5, pp.–3359-3363.
[26] A.S.Syed Navaz & Dr.G.M. Kadhar Nawaz “Layer Orient Time Domain Density Estimation Technique Based Channel Assignment in Tree Structure Wireless Sensor Networks for Fast Data Collection” June - 2016, International Journal of Engineering and Technology, Vol No - 8, Issue No - 3, pp.–1506-1512.
[27] M.Ravi & A.S.Syed Navaz "Rough Set Based Grid Computing Service in Wireless Network" November - 2016, International Research Journal of Engineering and Technology, Vol No - 3, Issue No - 11, pp.1122– 1126.
[28] A.S.Syed Navaz, N.Asha & D.Sumathi “Energy Efficient Consumption for Quality Based Sleep Scheduling in Wireless Sensor Networks” March - 2017, ARPN Journal of Engineering and Applied Sciences, Vol No - 12, Issue No - 5, pp.–1494-1498.
[29] A.Anusha Priya G.Vijayalakshmi, Perceiving Kernel-Level Rootkits Using Data Structure Invariants, 2015/7, International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 7, Pages 6719-6724.
[30] N.Thavamani, A.Anusha Priya, A QOD-Slanting Scattered Steering Procedure for Fusion Wireless Set of Connections, 2015/7, International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 7, Pages 6752-6757.
[31] M.Ravi A.M.Nirmala, P.Subramaniam, A.Anusha Priya, Enriched Performance on Wireless Sensor Network using Fuzzy based Clustering Technique, 2013, International Journal of Advanced Studies in Computer Science and Engineering, Volume 2, Issue 3, Pages 11-17.
[32] A Anusha Priya, A Mohanapriya, An Effective Scrutiny of Static and Dynamic Load Balancing In Cloud, August 2016, International Journal of Emerging Technology in Computer Science & Electronics, Volume 23 Issue 4, Pages – 27 -30.
[33] P.Vijayakumar, A.Anusha Priya, Stabilize the Movement of Nodes on anycast Routing with jamming responsive in mobile ad-hoc networks, September 2015, international Journal of Engineering Sciences & Research Technology, Volume 4 Issue 9,Pages – 207 -214.