

Exploring the techniques for object detection, classification, and tracking in video surveillance for crowd analysis

Aparna Gullapelly

Department of CSE, Koneru Lakshmaiah Education Foundation,
Deemed to be University, Hyderabad, Telangana 5000765, INDIA
aparnag932012@gmail.com

Dr. Barnali Gupta Banik

Department of CSE, Koneru Lakshmaiah Education Foundation,
Deemed to be University, Hyderabad, Telangana 5000765, INDIA
barnaliguftabanik@klh.edu.in

Abstract - Video surveillance (observation) will be going through a massive variation in the most recent decade, and loads of research is experiencing around there. As Tracking of objects will be an essential portion of such frameworks, it gets essential to survey all state of art techniques & methods that exist to classification, detection of an object, & tracking for crowd analysis. This manuscript comprises of all models in the classified method. After object identification, classification is done to track that later. The Tracking of objects will be the technique to find the target in the following video frames. There are additionally a few difficulties in video observation that make it difficult like the object tracking on a few parameters such as problems of environment, illumination variation, pose variation, motion blur, matching of coordinates, & objects tracking, etc. Great characterization is deliberated for survey & complete review of all probable methods, so every perspective is subsequently enlightened in this article.

Keywords: Object Detection, Classification, Tracking, Crowd Analysis

1. Introduction

These days security necessities are expanding step by step, and human-based observation frameworks are insufficient to manage it. In this way, intelligent video observation frameworks become essential to have high security. The video observation frameworks that incorporate the detection of object & Tracking to execute security dimensions. Observation is examination and checking of activities & performance of doubtful objects for public security. The automatic video observations like in abnormal performance detection of objects & make economic advantage by saving currency on “human-based security frameworks.” As deliberated in numerous difficulties are associated with building up a video observation framework like illumination, issues of environment, motion blur, tracking a variety of an object, & occlusion, etc. & to execute tracking with these difficulties in real-time make tracking tedious.

In general, the video observation framework includes two significant structure blocks, like motion estimation & detection. The detection of the object will be the initial step that is directly affected by background data. Since there will be relevant immaterial and repetitive data in video across time & space, video information should be compressed at initial in video observation applications. Pressure might be accomplished by limiting spatial and transient redundancies existing in the video. In previous days, video information is compacted by lessening frame size in video quality.

The 2D symmetrical changes & motion compensation strategies are associated with late video coding principles to eliminate temporal & spatial redundancies. In the suggested strategy, 2D discrete cosine change is utilized for video compression due to its high energy compaction. The detection & estimation of motion is two main building parts of the video observation framework.

In the detection of motion, the moving object will be recognized by removing the variations in object limits, though, in the estimation of motion, the motion vectors have registered to evaluate places of moving articles.

The “optimal motion vector” is investigated by discovering the co-relocation of directions in reference outline for a block in the current frame. The optimal stream vector will be determined by utilizing the Horn-Schunck method for moving object recognition. Meanwhile, it assumes perfection in-stream over the entire picture outline; it will be more sensitive to noise.

The RLOF (Robust local optical flow feature tracking technique) has superb “long term feature tracking execution, yet its computational intricacy is much as contrasted with KLT. The foundation deduction is the strategy for removing the closer view object for motion investigation in video observation. Non-stationary foundations and illumination variations are bottleneck issues in the background deduction strategy.

Practically speaking, the worldwide imperatives of optical stream-based methods have disregarded those consequences in tracking error. In the majority of the foundation deduction strategies, trackers of objects have affected by foundation data, which prompts false detection. Moreover, an efficient classifier will be needed to segregate targets in cluttered conditions. To deal with these constraints, a new methodology is introduced in this manuscript that viably recognizes the objective in complex conditions without foundation impacts. The main commitments of this manuscript might be summarized as follows.

- (1) In compression of video, input video outlines have compressed by “2D discrete cosine change” with satisfactory blocking artifacts to decrease storage necessities.
- (2) In object discovery, the coordinating component focuses have determined by estimating the relationship coefficients among target template & compressed video edge.
- (3) Then “posterior probabilities” are figured, and the highest probability densities have evaluated by computing peak connection coefficients over the whole picture frame. These profoundly coordinating element points have limited dependent on the Bayesian standard.
- (4) Lastly, the coordinating element points have localized in progressive video outlines by inserting the most extreme probability densities over input outlines.

This paper organized the techniques for crowd analysis, as section 2 describes the object detection methods; section 3 discusses the object classification methods, and object tracking methods are presented in section 4. Finally concluded in section.

2. OBJECT DETECTION METHODS

The detection of the object will be categorized as - Appearance-based & Motion-based object detection. The appearance-based techniques incorporate elements such as size, color, edge, & any other static elements. The motion-based techniques incorporate the acceleration, trajectory, velocity & direction of the object.

2.1 Motion-based object detection

Primarily, “motion-based techniques” have categorized as- “Pooling, Frame Differencing, and Optical Flow technique.”

2.1.1. Pooling

Region of interest pooling (also known as **RoI** pooling) is an operation widely used in object detection tasks using convolutional neural networks. For example, to detect multiple cars and pedestrians in a single image. Its purpose is to perform max pooling on inputs of non-uniform sizes to obtain fixed-size feature maps (e.g., 7×7).

There have few broadly utilized pooling technologies like- “Average pooling, Maximum pooling (also known as.” Max pooling) [14]

2.1.2FD (Frame Differencing)

The “Frame Differencing (FD)” will also be a picture subtraction technique like pooling. However, in this kind of tracking like variance among previous & current frames will be taken, represented in (Eq. 3 and Eq4).

$$\text{Eq (3) Foreground Pixel: } |F_C - F_P| \geq \text{Threshold}$$

$$\text{Eq (4) Background Pixel: } |F_C - F_P| < \text{Threshold}$$

Whereas, F_P be previous frame & F_C be the current frame.

For good accuracy, the threshold must be active also adaptive to the background method. Grouping of FD & pooling might be used to identify the abnormal event

2.1.3 Optical Flow technique

The “optical flow technique” represents the movement pattern in the picture of 2 consecutive frames because of the removal of an item or envisioning gadget and shown in the form of the relocation vector field. These algorithms generally utilize frequency data, correlation, & gradient. As it deliberates total pixels, therefore, the optical flow will not be a better decision for real-time applications without any change.

Many techniques utilized optical stream and attempted to improve it depend on suspicion that “grey values of pixel” does not modify whereas having “pixel displacement” from 1 to another point in sequential edges are shown below.

$$\text{Eq.(5) } I(x + a, y + b, t+1) = I(x, y, t)$$

Whereas I be image in 2 dimensions, a, b are “displacement in coordinates,” and x and y at time t .

The “convolution of optical flow with gradient data outcomes” in development tracking

2.2 Appearance-based object detection

Principally, the “appearance-based techniques” might be categorized as Feature Extraction & Template Matching.

2.2.1 Template matching

It is an easy detection process & able to identify the static article in a group of frames. It takes picture parts recognized as a template that will be to look or identify. The single or group of templates might be taken after the next phase will be matching that incomplete template picture. After setting the threshold, corresponding standards are made. Whenever coordinated then, an imperative item is found. However, it develops delayed in the instance of a significant ROI. To make it secure as a framework is to be prepared with a large dataset of a specific item.

2.2.2 Feature Extraction

The “feature extraction (FE)” will be fundamentally the technique to take trademark data from picture & then show it in low dimensionality space. There are a few feature extraction strategies like deformable templates, Gabor highlights, Haar-like features, “unitary image transforms, Gabor features, speeded up Robust Features (SURF),” contour profiles, “SIFT (Scale Invariant Feature Transform) features,” Gradient features, Zernike moments, etc. The “local binary pattern (LBP) features” have utilized in “local feature extraction “dependent on the surface. The work used LBP for detection of face.

3. Object Classification Methods

The classification of the object will be the next phase after the detection of the object. An object is distinguished from other regarding 4 features - color, shape and size, motion & texture.

3.1 Colour-based classification

As it does not base on the scale of pictures, size, & orientation, it offers many accurate outcomes than other techniques, however, with increment in “computational time.” It provides a wide range of object classifications. The color remains persistent—that why it will be simple to detect objects with support of color. Therefore, “color-based histogram “might be used in “real-time applications like forest delineation,” whereas texture data will also be used.

3.2 Size and shape-based classification

This model takes low computational accuracy & time. The data concerning the shape will be gathered from the box around point, ROI, & blob. Therefore, it finds a particular blob, box in picture frames. It will not be appropriate for dynamic conditions & to identify internal movements.

3.3 Motion-based classification

This classification classifies objects on its velocity, acceleration, and few times variations in position. The computational time & accuracy will be enough; however, it is not appropriate for static objects.

3.4 Texture-based classification

The object might also be categorized by its texture. The texture knows the features of the surface. There will be different texture features based on that object are categorized like smoothness, roughness, coarseness, directionality, brightness change, contrast, etc. This kind of classification also needs many computations to provide accurate outcomes.

4. Object Tracking Methods

The Tracking of objects will be characterized as the way and direction of the object in picture plane through discovering & locating the place of the object in consequent frames. It provides area whereas the object lies in each frame. Then, the association will be attained among the placements of objects in different frames. The object will be tracked in 2 ways: back & forward tracking.

In “forward tracking,” the position of the object will be assessed from before picture frames investigation after their division. It discovers the area of the object in every frame by any recognition technique. If recognition is in the form of a point, the translation will be attained as a trajectory. In “backtracking,” data about the placement of object & related areas attained in the past frame will be upgraded in each frame. After dividing the present frame into background, foreground, communication is built up with prior edges, utilized for handheld items.

The Tracking will be commonly ordered into three sections: Kernel Tracking, Point Tracking, and “Silhouette-based object tracking.”

4.1 Point Tracking (PT)

The PT is utilized, when the object will be provided as a point or group of points after the detection of an object, then to find those points in consequent frames. It will be hard to utilize it if there should arise an occurrence of an impediment because once the item is blocked, then it might result in “false detection.” The PT will be categorized into two kinds: Probabilistic or statistical & Deterministic techniques.

4.1.1 Deterministic Method-

This technique deliberates a few factors to found relation among the present & past frames. These factors [26] are -

- Maximum velocity
- Smooth motion
- Common motion
- Proximity
- Proximal uniformity

a) Rigidity

These factors have not constrained to only the deterministic model. However, they might also be utilized with probabilistic methods. Then forming the factor communication between two frames it might also be protracted to several frames. Several factors might be taken in various frames to form a connection so that Tracking provides better outcomes.

4.1.2 Probabilistic Method-

These techniques take suspicions into deliberation through the Tracking. The calculation of state space will be done to analyse subsequent state based on the acceleration of object, position, & velocity. This technique might be categorized further based on the object that will be to be tracked –“Single and Multiple Object Tracking.”

a) Single Object Tracking

This kind of Tracking might be done with any of these techniques-

i) Kalman Filter

The “Kalman Filter (KF)” will be utilized to calculate unknown variables in the framework with more accuracy by examining the sequence of dimensions with noise in framework factors over time. Extended KF & Unscented KF have utilized in “nonlinear cases.” However, assessment is affected when there will be “high non-linearity.” KF might also track various objects by considering them as distinct. However, it will be a challenge to found better communication between different points.

ii) Particle Filter

As KF is only restricted to Gaussian variables & assess them very poorly. This difficulty is overwhelmed by “particle filtering.” The “particle filtering” as utilized in [29], might assess internal states from incomplete observations. Del Moral [33] created the name particle filtering in 1996. In this approach, without the requirement of expectations, it produces the samples from any kind of “probability density function (PDF),” named as “Importance sampling.” Then, these groups of examples with distinct values have signified as particles. Every particle comprises one group of values for every state variable [34]. Dark particle becomes extra weight [35]. However, weight collapse will be severe because of weight dissimilarity. It might be resolved over resampling.

There are numerous resampling methods accessible like Residual, Multinomial, Systematic, and Stratified. This technique contains three steps [36]-

b) Multiple Object Tracking The “multiple objects tracking” utilizing the statistical method needs numerous communications of diverse objects & state approximation.

Traditional models are dependent on the “Global Nearest Neighbor (GNN) approach” [37, 38] that discovers better distribution from earlier tracks. It will be best, whereas observations have generated by the single target. However, there are probabilities of false recognition when various objects have near to each other.

There are two models for manifold object tracking-

i) Joint Probability Data Association Filter

Better communication of points among numerous objects might be recognized when all are isolated; however, it will be tough to describe the communication among objects that are close to the neighborhood. The “Joint Probability Data Association Filter (JPDAF)” has an issue of combination. At close points, the probabilities of improper communication have more & even a particular inappropriate communication might lead to a framework failure. Therefore, proper data relationships among state & object assessment must be appropriately projected [39]. JPDAF works on a system of KF & based on probability assessment. As the JPDA technique enables upgrading of the track by the weighted sum of total explanations, indefinite situations of relationship are evaded.

ii) Multiple Hypothesis Tracking

As observation frameworks need “multiple objects are tracking,” to found manifold communications among objects and their state to resolve information association difficult. For this persistence, “Multiple Hypothesis Tracking (MHT)” works as an alternative, utilized in [39, 40].

In the “MHT tedious data association, “choices have postponed until much data is not received. MHT is categorized into 2 methods: Track oriented MHT (TOMHT) & Hypothesis oriented MHT [42].

5. Conclusion

This manuscript gives a summary of whole methods on classification, detection of an object, and tracking in video observation. This manuscript mainly concentrates on the working standards of the whole of these technologies. As the detection of object & Tracking handles with many difficulties. Each technique has its benefit; however, one specific model will be incapable of handling all difficulties alone.

Therefore, it is essential to improve a technique to track an object that can track objects effectively & track object precisely & proficiently in regards to computation. As video observation will be on-demand, this manuscript is supported to know all fundamentals to plan a tracking method that adapts to different situations for crowd analysis.

References

- [1] G. L. Foresti, L. Marcenaro and C. S. Regazzoni, "Automatic detection and indexing of video-event shots for surveillance applications," in *IEEE Transactions on Multimedia*, vol. 4, no. 4, pp. 459-471, Dec2002
- [2] S. Abdurrahman, "Smart video-based surveillance: Opportunities and challenges from image processing perspectives," 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, 2016, pp.10-10.
- [3] S. Ezatzadeh and M. R. Keyvanpour, "Fall detection for elderly in assisted environments: Video surveillance systems and challenges," 2017 9th International Conference on Information and Knowledge Technology (IKT), Tehran, 2017, pp.93-98.
- [4] S. Little, K. Clawson, A. Mereu and A. Rodriguez, "Identifying and addressing challenges for search and analysis of disparate surveillance video archives," 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013), London, 2013, pp. 1-6.
- [5] K. Kalirajan, M. Sudha, V. Rajeshkumar, and S. S. Jamaesha, "Adaptive visual tracking system using artificial intelligence," in *Proceedings of the IEEE/OSA/IAPR International Conference on Informatics, Electronics and Vision (ICIEV'12)*, pp.954–957, IEEE, May2012.
- [6] E. Sankaralingam, V. Thangaraj, S. Vijayamani, and N. Palaniswamy, "Video compression using multiwavelet and multistage vector quantization," *International Arab Journal of Information Technology*, vol.6, no.4, pp.385–393, 2009.
- [7] P. Wang and H. Qiao, "Online appearance model learning and generation for adaptive visual tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.21, no.2, pp.156–169, 2011.
- [8] B. Qi, M. Ghazal, and A. Amer, "Robust global motion estimation oriented to video object segmentation," *IEEE Transactions on Image Processing*, vol.17, no.6, pp.958–967, 2008.
- [9] O. Brouard, F. Delannay, V. Ricordel, and D. Barba, "Spatiotemporal segmentation and region tracking of high definition video sequences based on a markov random field model," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'08)*, pp.1552–1555, October2008.
- [10] D. Jansari and S. Parmar, "Novel object detection method based on optical flow," in *Proceedings of the 3rd International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'13)*, Kuala Lumpur, Malaysia, 2013.
- [11] T. Senst, V. Eiselein, and T. Sikora, "Robust local optical flow for feature tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.9, pp.1377–1387, 2012.
- [12] M. Piccardi, "Background subtraction techniques: a review," 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), 2004, pp. 3099-3104 vol.4
- [13] Aseema Mohanty and Sanjivani Shantaiya. Article: A Survey on Moving Object Detection using Background Subtraction Methods in Video. *IJCA Proceedings on National Conference on Knowledge, Innovation in Technology and Engineering (NCKITE 2015)* NCKITE 2015(2):5-10, July 2015
- [14] P. Hafner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [15] J. Kang, D. V. Anderson and M. H. Hayes, "Face recognition for vehicle personalization with near infrared frame differencing," in *IEEE Transactions on Consumer Electronics*, vol. 62, no. 3, pp. 316-324, August 2016.
- [16] M. Zhu and H. Wang, "Fast detection of moving object based on improved frame-difference method," 2017 6th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, 2017, pp.299-303.
- [17] N. Srivastav, S. L. Agrwal, S. K. Gupta, S. R. Srivastava, B. Chacko and H. Sharma, "Hybrid object detection using improved three frame differencing and background subtraction," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, 2017, pp.613-617.
- [18] P. K. Bhaskar, S. P. Yong and L. T. Jung, "Enhanced and effective parallel optical flow method for vehicle detection and tracking," 2015 International Symposium on Mathematical Sciences and Computing Research (iSMSC), Ipon, 2015, pp.138-143.
- [19] D. Guo, A. L. van de Ven and X. Zhou, "Red Blood Cell Tracking Using Optical Flow Methods," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 991-998, May2014.
- [20] T. Senst, V. Eiselein and T. Sikora, "Robust Local Optical Flow for Feature Tracking," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 9, pp. 1377-1387, Sept.2012.
- [21] V. John, K. Yoneda, Z. Liu and S. Mita, "Saliency Map Generation by the Convolutional Neural Network for Real-Time Traffic Light Detection Using Template Matching," in *IEEE Transactions on Computational Imaging*, vol. 1, no. 3, pp. 159-173, Sept.2015.
- [22] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, Dec.2006.
- [23] Z. Wang and R. Boesch, "Color- and Texture-Based Image Segmentation for Improved Forest Delineation," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3055-3062, Oct.2007.
- [24] E. Marchiori, M. Marchiori and J. N. Kok, "Forward-tracking: a technique for searching beyond failure," *Proceedings Eighth IEEE International Conference on Tools with Artificial Intelligence*, 1996, pp.324-331.

- [25] K. Chaudhary, Y. Mae, M. Kojima and T. Arai, "Autonomous acquisition of generic handheld objects in unstructured environments via sequential back-tracking for object recognition," 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, 2014, pp.4953-4958.
- [26] AlperYilmaz,OmarJavedandMubarakShah,"ObjectTracking:A Survey", ACM Computing Surveys, December 2006.
- [27] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," in Proceedings of the IEEE, vol. 92, no. 3, pp. 401-422, Mar2004.
- [28] A. Kamann, J. B. Bielmeyer, S. Hasirlioglu, U. T. Schwarz and T. Brandmeier, "Object tracking based on an extended Kalman filter in high dynamic driving situations," 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, 2017, pp.1-6.
- [29] L. ChongYi, L. Cheng, F. LinYu and Y. JingTing, "Target tracking based on extended Kalman particle filter," 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2017, pp.1715-1719.
- [30] L. Marcenaro, M. Ferrari, L. Marchesotti and C. S. Regazzoni, "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching," Proceedings. International Conference on Image Processing, 2002, pp. III-341-III-344vol.3.
- [31] C. M. Bukey, S. V. Kulkarni and R. A. Chavan, "Multi-object tracking using Kalman filter and particle filter," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017, pp. 1688- 1692.
- [32] L. Fu, "Particle Filter Pedestrian Tracking Algorithm Based on Selected Region RGB Histogram," 2018 International Conference on Robots & Intelligent System (ICRIS), Changsha, China, 2018, pp. 287-290.
- [33] Del Moral, P.; Guionnet, A. Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* 9 (1999), no. 2,275—297
- [34] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," in IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, Feb2002.
- [35] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," in IEEE Transactions on Signal Processing, vol. 51, no. 10, pp. 2592-2601, Oct.2003.
- [36] Kaijen Hsiao, Henry de Plinval-Salgues and Jason Miller," Particle Filters and Their Applications", Cognitive Robotics April 11,2005
- [37] A. Sinha, Z. Ding, T. Kirubarajan and M. Farooq, "Track Quality Based Multitarget Tracking Approach for Global Nearest-Neighbor Association," in IEEE Transactions on Aerospace and Electronic Systems, vol. 48, no. 2, pp. 1179-1191, APRIL2012.
- [38] M. Shi, Q. Ling, Z. Yu and J. Zhu, "Association using modified Global Nearest Neighbor in the presence of bias," Proceedings of the 32nd Chinese Control Conference, Xi'an, 2013, pp.4688-4691.
- [39] X. Liu, K. Wang, D. Li, J. Xu and J. Pan, "A two-stage gating algorithm for joint probability data association filter," IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS, Beijing, 2010, pp.381-384.
- [40] K. Ahmadi and E. Salari, "A novel Multiple Hypothesis Testing (MHT) scheme for tracking of dim objects," 2015 IEEEInternationalConference on Electro/Information Technology (EIT), Dekalb, IL, 2015, pp. 365-369.
- [41] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," in IEEE Aerospace and Electronic Systems Magazine, vol. 19, no. 1, pp. 5-18, Jan.2004.
- [42] Y. Kosuge, M. Kojima and S. Tsujimichi, "Multiple manoeuvre model track-oriented MHT (multiple hypothesis tracking)," SICE Annual, 1999. 38th Annual Conference Proceedings of the, Morioka, 1999, pp.1129-1134.