DISCRIMINATIVE APPEARANCE MODEL FOR ROBUST ONLINE MULTIPLE TARGET TRACKING

MODELO DE APARIENCIA DISCRIMINATORIO PARA UN SÓLIDO SEGUIMIENTO EN LÍNEA DE MÚLTIPLES OBJETIVOS

Altaf Osman Mulani
SKN Sinhgad College of Engineering, Pandharpur, India
draomulani.vlsi@gmail.com

Rajesh Maharudra Patil
SKN Sinhgad College of Engineering, Pandharpur, India
maharudrapatil.1@gmail.com

Kazi Kutubuddin Sayyad Liyakat
Brahmdevdada Mane Institute of Technology, Solapur, India
drkkazi@gmail.com

ABSTRACT

Multiple target tracking algorithm faces challenges of occlusion, halt, merge and split of the moving objects. The change in appearance of the moving targets complicates the tracker. Hence the discriminative appearance model is needed for the robust multiple target tracking. This paper incorporates tracking-by-detection approach along with Kalman filter based motion model. The appearance of the target proposed in this paper is modeled based on object's texture features. Phase congruency derived by gray level co-occurrence matrix (GLCM) constitutes the appearance model of the moving object. Thus the proposed tracker is invariant to image illumination and contrast variation. Confidence based data association helps for track management in this paper. The proposed tracker is evaluated on the standard benchmark datasets namely CAVIAR, PETS2009 and ETH. The experimental results of the proposed tracker demonstrate zero error in identity matching when tested on ETH dataset.

Key Words: Multiple target tracking, Kalman filter, phase congruency, appearance model, data association

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

RESUMEN

El algoritmo de seguimiento de múltiples objetivos enfrenta desafíos de oclusión, detención, fusión y división de los objetos en movimiento. El cambio de apariencia de los objetivos en movimiento complica el rastreador. Por lo tanto, el modelo de apariencia discriminativa es necesario para el seguimiento robusto de objetivos múltiples. Este documento incorpora el enfoque de seguimiento por detección junto con el modelo de movimiento basado en el filtro de Kalman. La apariencia del objetivo propuesto en este documento se modela en función de las características de textura del objeto. La congruencia de fase derivada de la matriz de coocurrencia de nivel de gris (GLCM) constituye el modelo de apariencia del objeto en movimiento. Por lo tanto, el rastreador propuesto es invariable a la iluminación de la imagen y la variación del contraste. La asociación de datos basada en la confianza ayuda a la gestión de seguimiento en este documento. El rastreador propuesto se evalúa en los conjuntos de datos de referencia estándar, a saber, CAVIAR, PETS2009 y ETH. Los resultados experimentales del rastreador propuesto demuestran cero errores en la coincidencia de identidad cuando se prueban en el conjunto de datos ETH.

Palabras clave: seguimiento de objetivos múltiples, filtro de Kalman, congruencia de fase, modelo de apariencia, asociación de datos

INTRODUCTION

Multiple target tracking (MTT) has been proved to be the greatest challenge in modern surveillance systems. Belonging to own set of advantages, drawbacks and applicable video sequences, state-of-the-art trackers' , T. Kutschbach, E. Bochinski, V. Eiselein, T. Sikora, (2017), Francois, Thierry Chateau, Datta Ramadasan (2009) and Zhongdao Wang, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang (2019) performance varies.

The performance of the tracker is highly degraded by full or partial occlusion, merge, split, illumination changes to name a few. In multiple target tracking (MTT), targets and the detections need to be matched from frame to frame in spite of changes in illumination, color, shape, size, position and motion path. The objects should be reliably tracked on a continuous basis. The identity (ID) must be preserved even if the objects split from an original tracked object and getting separate identities, leaving their individual history behind [25] or merge with other objects. The need to maintain the identity of each detected object adds more complexity to the tracker. Hence, data association on the basis of feature and motion estimation is usually

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

considered of utmost importance in MTT (Zhongdao Wang, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang (2019).

The tracking algorithm estimates the positions of varying number of targets using a series of noisy measurements. This may result in missed detection, inaccurate positions of target and false alarms. Many tracking approaches proposed appearance models considering color (Tang Sze Ling, Liang Kim Meng, Lim Mei Kuan, Zulaikha Kadim, Ahmed A. Baha'a Al-Deen, 2009) texture (Robert M. Haralick, K. Shanmugam, Its'hak Dinstein (2009), SIFT (Sudipta N. Sinha, Jan-Michael Frahm, Marc Pollefeys, Yakup Genc, 2007) and HOG (Anton Milan, Laura Leal-Taix´e, Konrad Schindler, Ian Reid,2015) features. Multi modal cues like appearance, motion and interaction among objects help to infer their motion for multiple applications such as people tracking (S. McKenna, S. Jabri, Z. Duric, H. Wechsler. 2000 and Min Yang, Yunde Jia, 2015), vehicle tracking (Jose C. Rubio, Joan Serrat, Antonio M. Lopez, Danial Ponsa. 2022), sports analysis (P. Nillius, J. Sullivan, S. Carlsson, 2006) etc. 'Tracking-by-detection' paradigm connected across video frames is mostly appreciated and followed. Over the decades work has been done towards proposing the optimum solution for MTT (Seong - Wook Joo, Rama Chellappa, 2006 M. Hofmann, M. Haag, G. Rigoll, 2013)

The proposed algorithm considers any class and any number of objects for tracking in contrast to previous approaches which limited their algorithms to a specific class of objects. The proposed framework does not need any kind of pre-processing. 'Tracking- by -detection' paradigm underlines significance of the detection output. Background subtraction algorithm solves the purpose of detection in the proposed work. Object model formed with the detection methodology and appearance model is updated frame wise. Texture features in terms of normalized GLCM` considered in our work as they are comparatively robust. Normalized GLCM considers texture features in four orientations. Phase congruency is derived from normalized GLCM in six orientations.

Normalized phase congruency value is used for further data association task. Usually illumination change affects the performance of the tracker. The proposed framework is independent to change in illumination due to the tracker's dependency on phase congruency calculation. Considering linear propagation of tracklets in short time samples, we used a simple Kalman filter. State space analogy allows the Kalman filter to predict and estimate the location of the object being tracked. Unassigned observations are given separate identities with preservation of their object models.

Proposed illumination independent appearance model leads to achieve a correct identification count. Combination of appearance model and motion

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

model confirms the reliability and sustainability of the proposed tracker. An application of Euclidean distance criteria for both appearance model and temporal positions enhances the robustness of the proposed algorithm. Retaining object models allow the proposed framework to consider the re - appearance of tracks after a long time. It helps to achieve minimum identity switching error.

The remainder of the paper is organized as follows: Section 2 reviews related work. Section 3 covers the proposed algorithm. Experimental results are demonstrated in section 4. Paper is concluded in section 5.

THEORETICAL FOUNDATION

Appearance modeling has gained more attention in the multiple object tracking literature H. Possegger, T. Mauthner, P. M. Roth, H. Bischof(2014) and Zeyu Fu, Pengming Feng, Federico Angelini, Jonathon Chambers, Syed Mohsen Naqvi (2018). Min Yang and Yunde Jia (2016) proposed a temporal dynamic appearance model.

They used HMM for appearance modeling. HOG features are used as low level appearance features. Shape, motion and appearance models together used for data association. Tang Sze Ling et al (2009) proposed the clustered color tracker which improved the performance in the re - appearance of observations. The average comparison score of clusters computed both in current as well as the previous frame decides data association. Mohammed and Morris (2014) suggested a color-based technique which combines accruing and normalizing histograms. Their approach proved to be robust against varying illuminations but failed in identifying specific color intensities of extreme camera's entire region of symbols. The minimum distance criterion acts as object classifier. Seong-Ho Lee et al (2018) proposed histogram based MOT tracker for online discriminative appearance learning using partial least square method. They preferred to update the tracks with the lower discrimination ability. Data association on the basis of length and continuity of the track along with an appearance affinity with the detection is preferred in their algorithm. Appearance and illumination changes were very well handled by Zeyu Fu et al (2018) using enhanced sequential Monte Carlo probability hypothesis density (PHD) filter based MTT.

Texture has gained more attention in appearance modeling. G. Jemilda and S. Baulkani (2015) used a Gabor filter to extract textural features followed by application of spatio-temporal optimization which includes multiple kernel learning technique which could track an object. Yi Dai et al (2009) proposed marginal likelihood based feature fusion approach. It includes color and texture features to represent the objet followed by the

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Markov model to determine the motion of the object under the track. Fengwei Yu et al (2016) computed data association in tracking on the basis of the appearance features affinity values. A second order statistical method, GLCM for texture extraction gives promising results in matching algorithms (Jyotismita Chaki, Ranjan Parekh, Samar Bhattacharya . 2015 and Bobo Wang, Hong Bao, Shan Yang, Haitao Lou. 2013)

In complex environments, not only pedestrians' changing postures and scale, occlusion, merge, split but also moving background impose more complexity in pedestrian tracking problem. Francois Bardet et al (2009) used MCMC particle filter to compute likelihoods of the moving objects. Occlusions and deep scale changes were handled by global likelihood observation function. Gaussian Mixture Probability Hypothesis Density (GMPHD) filter and kernalized correlation filters (KCF) proposed by T. Kutschbach et al (2017) for multiple object tracking in video data. This method resulted in better quality of object tracking at the cost of increased computational complexity. Tianzhu Zhang et al (2014) proposed multi-task correlation particle filter (MCPF) tracker which considered an appearance model for tracking with particle filter. CNN features or intensity, HOG, color features is used to build an appearance model of an object under the track. Prediction state of particle filter considered an appearance model of each part of the target under the track.

Alex Bewley et al. [39] proposed simple online real time tracking - SORT which incorporates classical tracking methods, i.e. Kalman filter and Hungarian algorithm along with faster region convolution neural networks (FrRCNN) detection framework. In the first stage, FrRCNN extracts features and regions are proposed for second stage where the classification of objects in the proposed region is carried out. Their algorithm is complemented with an appearance model which considers feature map on dimensionality 128 which improves the identity matching of the tracks. Guanglong Du et al [6] developed an algorithm which estimated position and orientation using interval Kalman filters and improved particle filter respectively.

In this paper, we propose a novel appearance modeling approach which is robust to illumination changes in contrast to the state-of-the-art methods. The objects in each frame are represented in terms of their respective normalized GLCMs in four orientations. Formation of normalized GLCM of each detection allows to reduce noise. Phase congruency feature vector derived from this normalized GLCM is robust to illumination changes. Though nonlinear in nature, object's motion is linear at each frame time sample. This allowed us to use the Kalman filter prediction model as the temporal model.

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

RESULTS OF THE INVESTIGATION

Proposed system

Our proposed method is divided into five parts:
a). Processing of Current Frame
b). Appearance Model
c). Temporal Model
d). Data Association
e). Track Initialization and Prediction

Processing of Current Frame

Processing of the current frame includes foreground detection and object model formation. It is illustrated in Figure 1.

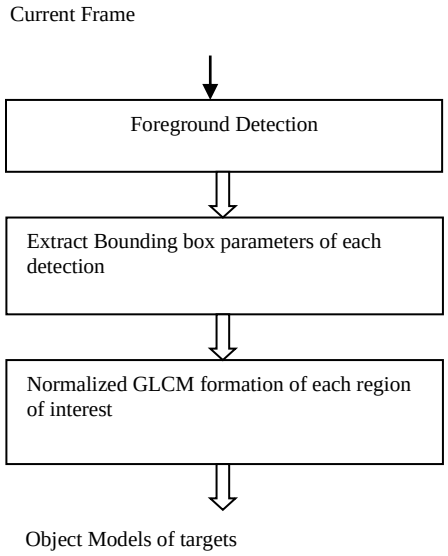

Fig 1 Processing of Current Frame

Foreground Detection

Tracker's performance gets significantly affected by foreground detection quality. Sun, C. Zhao, Z. Yan, P. Liu, T. Duckett, R. Stolkin, (2019) proposed weakly supervised approach which is proved to be highly effective in RGB-D object detection and recognition application with limited human annotations. Background subtraction technique along with morphological

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

operations are incorporated in the proposed algorithm for foreground detection. The subtracted frame is binarized by threshold method. 40 frames are used for the training foreground detection process. Minimum background ratio 0.7 is used for binarization process. Fig 2 shows an original image along with binarized image.

Bounding Box Parameters Extraction

Depending on the threshold of the number of pixels in the region, an object is detected. In our work, we considered 400 pixels as the threshold for the blob area. A bounding box is placed around it. Bounding box parameters are extracted for tracking process. The center of gravity point, i.e. centroid of a bounding box is the point of interest. It is considered for the prediction of the object position in case of non-assignment to the previously detected track.

The bounding box around the blob (as shown in Fig 2) b= (x, y, w, h) represents the spatial extend of an object. (x,y) are the co-ordinates of left most corner. (w, h) are width and height of the bounding box respectively.

(a) **(b)**

Fig 2 Foreground Detection and Bounding Box Formation b = (x,y,w,h) (a) around the target (b) around the blob

Normalized Gray Level Co-occurrence Matrix

The gray level co-occurrence matrix (GLCM) [1] is the spatial gray level dependence matrix. It is based on the statistics of pixel intensity distribution. Thus a co-occurrence matrix is developed with a comparison of neighboring pixels on the basis of intensities (Francois, Thierry Chateau, Datta Ramadasan,2009). The dimension of GLCM is equal to the number of gray levels i.e. pixel brightness values in an image.. GLCM provides texture information. Single pixel statistics do not provide sufficient descriptions of texture for practical applications. The co-occurrence matrix is the second order statistical method which provides numerical data of the texture. It is obtained by considering spatial relationship between pixels. The co-

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

occurrence matrix expresses the relative frequencies of the pixels in an image in Θ direction and distanced from a specific pixel.

GLCM of dimensionality 8x8 each is computed in all four offsets (Θ) i.e. 0^0, 45^0, 90^0 and 135^0 in the proposed algorithm. All four GLCMs are further normalized to get equivalent normalized GLCM of each region of interest. Normalization of GLCM projects texture features onto the unit hyper-sphere in terms of intensity pixel arrangements. (D. Clausi, Y. Zha, 2003; J. Gotlieb, H. Kreyszig, 1990; [7] S. McKenna, S. Jabri, Z. Duric, H. Wechsler, 2000 and J.M.Wu, Y.C. Chen,2013)

3Object Model

We considered extremely challenging situations where almost no priori information regarding the object's shape, color, texture or motion is known. The quiescent objects may appear and disappear anywhere, anytime, whether jointly, adjacently or solitarily.

To track detected objects, the object model of each is built and maintained in the proposed algorithm. Each object is parameterized with spatial layout and its textural appearance. An object in the current frame of the video sequence is extracted as region of interest and represented as detection

$$D = (C_x, C_y, b) \quad (1)$$

where (C_x, C_y) is the centroid of the bounding box and bounding box parameters are represented by b. The proposed object model is

$$O \equiv (D, \mathrm{g}) \quad (2)$$

It consists of D as a bounding box which represents an object in terms of position and spatial location of the detected blob and ‘g’ as its textural appearance in terms of normalized GLCM.

Appearance Model

The selected appearance features of multiple targets under track need to be durable, stable and reliable. The parameters describing image features must be invariant to image illumination and magnification. Image illumination variations affect the performance of gradient based edge detection methods. Approaches using gradient-based feature detection are also sensitive to changes in size and noise. An alternative to gradient-based approaches is the frequency-based function extraction approach.

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Proposed Phase Congruency Based Appearance Model

Phase congruency is a dimensionless entity. Its value is in between 0 and 1, suggesting no importance for high significant values[10]. The proposed phase congruency based appearance model follows following steps.

- The proposed appearance model considers the proposed object model $O \equiv (D, \mathrm{g})$ as the input. Normalized GLCM of the detected blob (g) is considered for further computation of normalized phase congruency.
- Following three issues had been taken care of and resolved in computing normalized phase congruency

1. Filter Form in 2D : We consider an image of the detected object
2. The number of orientations to analyze: In our experimentation, we computed phase congruencies in 6 orientations.
3. The manner in which the findings are combined from each orientation: Normalization of phase congruencies computed in six orientations is computed. The normalization of Fourier transform is scale and rotation invariant. It presents low noise sensitivity. It is independent of brightness or contrast changes in an image and spatial magnification.

- Threshold design for data association
  The normalization of phase congruencies in six orientations significantly describes feature points as an absolute measure. It is used as the threshold for data association. This makes it compatible for proposed computation of nearest neighbor approach.

Temporal Model

Temporal model propagates the target's identity into the next frame. These inter frame displacements of the objects under track are approximated with linear Kalman filter's constant velocity model.

Kalman Filter

Kalman filter for tracking an object in a Cartesian coordinate system, moving with constant velocity is used in the proposed work. We assumed zero initial velocity of the tracked object. A discrete time, the linear State-Space System

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

is implemented with the Kalman filter algorithm (R. E. Kalman, 1960) as follows

$$x(k) = A * x(k-1) + B * u(k-1) + w(k-1) \quad \text{(state equation)}$$
$$z(k) = H * x(k) + v(k) \quad \text{(measurement equation)}$$

where A is a state transition matrix, H is a measurement transformation matrix, B is control input to state transformation.

Kalman filter predicts the track's future location with reduced noise in the detected location and associates multiple detections with these tracks. It Corrects or updates the tracks using the current measurement of the detections.

Kalman Filter Temporal Model

Our work considers following parameters:

$A = [1\ 1\ 0\ 0;\ 0\ 1\ 0\ 0;\ 0\ 0\ 1\ 1;\ 0\ 0\ 0\ 1]$
$H = [1\ 0\ 0\ 0;\ 0\ 0\ 1\ 0]$
$B = [\,]$

Each target stage is modeled as:

$$X = [u, v, \dot{u}, \dot{v}]^T \quad (3)$$

where u and v represent the horizontal and vertical pixel location of the centroid (C_x, C_y) of the target bounding box as per the proposed object model O shown in equation (2).

Kalman filter framework (J.M.Wu, Y.C. Chen, 2013) optimally solves velocity components $\dot{u}$ and $\dot{v}$. In case of no detection and no target association, target's state is simply predicted without correction using Kalman filter's linear velocity model.

Data Association

In the proposed method, simple frame-by-frame data association technique is adopted. At each frame, the existing tracks are updated by detections with our data association technique. Later tracks' data is sequentially grown with associated detections' data. If no association, then track initialization strategy completes the tracking method. Note that we do not have any track termination strategy, as we need every track's information in case of reappearance. Algorithm 1 describes the tracking procedure of the proposed Phase Congruency of normalized GLCM (PCNG) tracker.

Algorithm 1: Tracking Procedure of Proposed PCNG Tracker
Input: Current frame at the (t+1) time step, Detection set $p = Z_{t+1}$, Previous track set $q = X_t$

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Output: New track set X_{t+1}

Steps:

1. Calculate the motion affinity matrix $\Psi_{p,q}$.
2. Use $\Psi_{threshold}$ to decide success of motion affinity.
3. Calculate the appearance affinity matrix $Ð_{p,q}$
4. Use $Ð_{threshold}$ to decide success of appearance affinity.
5. Obtain association-success track subset, association-fail track subset, matched detection subset and unmatched detection subset using steps 2 and 4.
6. Use Kalman filter correction model to update association-success track subset with matched detection set as $(X_t)_{\text{corrected}}$
7. Use Kalman filter to predict association-fail track subset as $(X_t)_{\text{predicted}}$
8. Initialize the unmatched detection set as new track subset $(X_{t+1})_{\text{new}}$
9. Merge the track subsets to generate new track set as

$X_{t+1} = (X_t)_{\text{corrected}} + (X_t)_{\text{predicted}} + (X_{t+1})_{\text{new}}$

Temporal Match

Based on the assumption that tracks have small displacements over consecutive frames in a video sequence, we formed association cost matrix Ψ. Each element in Ψ indicates distance between the n^{th} track and the m^{th} detection. If the distance score Ψ_{nm} is the minimum in the distance set and less than the $\Psi_{threshold}$, then it is further considered for appearance affinity match.

Suppose we have n tracks $X_t = \{X_t^p\}_{p=1}^n$ at time t, and m detections $Z_{t+1} = \{Z_{t+1}^q\}_{q=1}^m$ at time $(t+1)$, where X_t^p is the p^{th} track and Z_{t+1}^q is the q^{th} detection. To form an association between X_t^p and Z_{t+1}^q, an $n \times m$ association cost matrix Ψ is computed.

$$\Psi_{pq} = \rho_M\left(X_t^p, Z_{t+1}^q\right) \quad (4)$$

where $\rho_M\left(X_t^p, Z_{t+1}^q\right)$ is the motion affinity between track X_t^p and detection Z_{t+1}^q based on motion model. The cost ρ_M takes into account the Euclidean distance between the predicted centroid of the track and the centroid of the

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

detection. It also includes the confidence of the prediction, which is maintained by the Kalman filter. Then for data association we computed,

$$\Psi_{min} = \min\ (|\Psi_{pq}|) \quad (5)$$

When $\Psi_{min} \leq \Psi_{threshold}$ (we set $\Psi_{threshold}$ as 20 pixels in our experimentation), the detection $Z_{t+1}^{q}$is checked for appearance match as described further to get associated with track X_t^p. The detection Z_{t+1}^{q} gets associated with the track X_t^p if it satisfies the condition for matching. If the track X_t^p cannot be associated with any of the detection then its position will be predicted for the next frame using prediction model of constant velocity Kalman filter. The position of the track which gets associated with the detection is refined with detection's measurement using correction model of Kalman filter.

Appearance Match

Given an image patch of a detection we computed normalized GLCM (as described in Section 3.1.4) and then phase congruency (as described in Section 3.2.1). The $n \times m$ appearance cost matrix Ð is computed using Euclidean distance between phase congruencies of each track X_t^p and each detection Z_{t+1}^{q}.

$$Ð_{pq} = Eucli\,(\rho_A(X_t^p, Z_{t+1}^{q})) \quad (6)$$

where $\rho_A(X_t^p, Z_{t+1}^{q})$ is the affinity between track X_t^p and detection Z_{t+1}^{q}based on appearance model.

Then the smallest Euclidean distance between the p^{th} track and q^{th} detection in appearance space is:

$$Ð_{min} = \min\ (|Ð_{pq}|) \quad (7)$$

A binary variable is introduced to indicate whether an association is admissible according to the threshold metric $Ð_{threshold}$ decided. We set threshold metric $Ð_{threshold}$= 0.5 in our experimentation.

$$b_{(p,q)} = 1[Ð_{min} \leq Ð_{threshold}] \quad (8)$$

A minimum distance $Ð_{threshold}$ is imposed to accept assignments where the track and detection affinity in appearance model is less than or equal to $Ð_{threshold}$.We found that the $Ð_{threshold}$, the threshold metric of appearance cost implicitly handles the identity switch problem in case of short term partial or full occlusion caused by passing or halted targets.

This dual threshold strategy between consecutive frames helps to generate small but reliable tracklets. Though conservative, this strategy allows only reliable association between two consecutive frames.

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Extension of a track

The binary affinity between track X_t^p and detection Z_{t+1}^q is then calculated based on motion and appearance affinity expressed as:

$$\rho(X_t^p, Z_{t+1}^q) = \max\left(\rho_A(X_t^p, Z_{t+1}^q)\right).\min\left(\rho_M(X_t^p, Z_{t+1}^q)\right) \quad (9)$$

The detection Z_{t+1}^q gets associated with the track X_t^p if and only if $\rho(X_t^p, Z_{t+1}^q) = 1$ is achieved.

Track Initialization and Prediction

Unassociated detections' and unassigned tracks' management is carried out for future frames.

Track Initialization

Suppose there are m' unassociated detections $Z'_{t+1} = \{Z_{t+1}^u\}_{u=1}^{m'}$ at (t+1), where Z_{t+1}^u is u[th] unassociated detection. Such unassociated detections are initialized as new tracks at (t+1) and their object models are preserved for tracking from time (t+2).

Track Prediction

Suppose there are n' unassigned tracks $X'_{t+1} = \{X_{t+1}^u\}_{u=1}^{n'}$ at (t+1), where X_{t+1}^u is u[th] unassigned track. Such unassigned tracks at time (t+1) are predicted for time (t+2) using prediction model of Kalman filter.

Experiments

Qualitative Evaluation

The applicability and effectiveness of the proposed approach in various challenging scenarios are presented here.

Image sequences namely Browse1, Fight_Runaway1, Meet_WalkSplit and Fight_Chase from CAVIAR benchmark dataset are used for testing and demonstration of results. Non salient motion features like occlusion, halt, re-appearance, merge, split and re-identification are handled for performance presentation.

(a)Frame No. 47 | *(b) Frame No. 87* | *(c) Frame No. 340* | *(d) Frame No. 464* | *(e) Frame No. 698* | *(f) Frame No.766*

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Fig 3A: Halt, re-appearance and re-identification conditions tackled in Browse1 data sequence
(a) Tracking two objects (b) Halted object tracked successfully (c) No object detected for tracking (d) Re-appearance and re-identification of an object (e) Lost tracking due to detection flaw (f) Re-identification of an object


(g)Frame No. 361


(h) Frame No. 408


(i) Frame No. 140


(j) Frame No. 215


(k)Frame No. 305

Fig 3B: Merge, Split and Re-identification motion features handled in (i) Fight_Runaway1 data sequence (g) and (h) (ii) Meet_WalkSplit data sequence (i), (j) and (k)


(l) Frame No. 274 Initial identity


(m)Frame No.317 Full occlusion

Fig 3C: Full occlusion handled in Fight Chase data sequence
Fig 3: Non salient motion features handled by the proposed PCNG tracker

Fig 3 represents our PCNG tracker's performance in case of non salient motion features namely halt, merge, split, re-appearance, re-identification and occlusion. The frames shown in Fig 3A are from Browse1 video sequence of CAVIAR dataset. Halt, re-appearance, re-identification cases are successfully handled by our proposed PCNG tracker. Combination of proposed appearance model and motion model offers the correct identity to the objects under track as shown in Fig 3A (a). The halted object (refer figure 3A (b)) can also retain its identity due to the data association technique used in PCNG tracker. The tracker could not associate a particular track in sequential frames due to non-detection or detections not meeting the threshold criteria of motion and appearance model. No track termination policy adopted in the proposed tracker allows the motion and appearance model to match detections' appearances and physical distances with that of the tracks as shown in Fig 3A (d) and (f)

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Proposed normalized phase congruency of the normalized GLCM of the detection's or track's image forms the rigid appearance model. Ψ_{min} value of the motion model avoids the same identification to the two distant but likely looking objects. This helps to minimize identity switching problem in case of merge and split challenge in tracking phenomenon. The results are shown in Fig 3B (g),(h) for Fight_Runaway1 data sequence and (i),(j),(k) for Meet_WalkSplit data sequence of CAVIAR dataset.

Fight_Chase sequence demonstrates occlusion. An individual detected in Fig 3C *(l)* retains its identity even when he fully occludes/overlaps the other as shown in figure 3C *(m).* The identity switch problem during partial or full occlusion in Fight_Chase sequence is handled by $Đ_{threshold}$. It is the threshold metric of appearance cost of the proposed tracker's appearance model. When the target is fully occluded or overlapped by other target, the overlapping target is tracked but not the overlapped one. Hence the proposed tracker generates appearance model only for the overlapping target. This extends the association of the overlapping target.

Any of the previously detected tracks are not deleted in the proposed tracking algorithm. It allows us to track the overlapped/occluded target after occlusion is over. Thus, our tracker successfully handles merge, split, halt and occlusion features of motion.

Quantitative Evaluation

The standard CLEAR MOT (Anton Milan, Laura Leal-Taix´ e, Ian Reid, Stefan Roth, Konrad Schindler, 2016) metrics are used to quantify performance of our proposed PCNG tracker algorithm. The quality of tracking and association is directly depicted by Multi-object tracking accuracy (MOTA) and Number of identity switches (IDS) respectively. The ratio of Correctly matched detections to total detections in ground truth depicts sensitivity of the tracker in terms of Recall. Higher values of MOTA, MOTP, Recall and lower score of IDS highlight better performance of the tracker. A true positive, considers 50% overlap of the tracking bounding box in the result with the corresponding ground truth bounding box.

Comparison with the state-of-the art algorithms for public benchmark databases allows quantitative evaluation of our PCNG algorithm. Public detections are used for evaluation. The tracking results to compare with are extracted from the MOT Challenge website and the public papers. The numbers marked in bold indicate the first place in the ranking of evaluation measures in the result tables. The dash (-) sign in a column of comparison table indicates no value available of that evaluation measure for the algorithm mentioned in respective row.

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Comparison on the basis of ratio of identity switches (IDS) to correctly matched detections

The proposed PCNG tracker is evaluated on ETH-Bahnhof sequence. The proposed tracker is compared with the state-of-the-art trackers for performance evaluation on the basis of the quantitative measure mentioned in Anton Milan, Laura Leal-Taix´ e, Ian Reid, Stefan Roth, Konrad Schindler, (2016) which is the ratio of IDS to correctly matched detections. For the ease of comparison, quantitative results are computed on first 350 frames of the ETH - Bahnhof sequence and tabulated in Table 1.

Algorithm	**ETH-Bahnhof**	**ETH (GT)**
DP [23]	37/1387	25/1648
MCNF[17]	11/1057	5/922
LRMCNF[16]	23/1514	14/1783
CML[13]	5/1728	3/1786
TSML[13]	**1/1790**	**0/1820**
Proposed PCNG	36/2144	**0/2144**

Table 1: Comparison of tracking results on the basis of ratio of identity switches (IDS) to correctly matched detections, on ETH-Bahnhof sequence.

Bing Wang et al (Bing Wang, Kap Luk Chan, 2013) generated initial tracklets based successive shortest path algorithm. Online learned target specific metrics which are adaptive to local segments, refine their initial tracklets into reliable tracklets. The proposed PCNG tracker considers not only the shortest distance in motion model but also the maximum affinity of the detection with the track in appearance model. It results in better response for IDS.

TSML though outperforms our algorithm, it is observed that our algorithm can correctly match much more detections compared to it. Moreover, zero ID switches indicate better performance of our algorithm in long term tracking.

Comparison on the basis of MOT metrics

We also compared our algorithm with state-of-the-art methods on the basis of MOT metrics [12] when evaluated on PETS2009-S2L1 data sequence. Public detections are used for the evaluation and comparison.

Algorithm	**MOTA (%)**	**MOTP (%)**	**Recall (%)**	**Gt**	**IDS**

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Energy minimization[18]	81.4	76.1	-	19	15
DC Tracking[19]	95.9	78.7	-	19	10
KSP[20]	80.3	72.0	-	19	13
MTMM[21]	83.3	71.1	-	19	19
UHMTGDA[25]	97.8	75.3	-	19	8
HJMRMT[2]	98.0	82.8	-	19	10
$(MP)^2$ T[38]	90.7	76.0	-	19	-
DTLE Tracking[27]	90.3	74.3	-	19	22
CEMMT[36]	90.6	80.2	-	19	11
GMCP-tracker[37]	90.3	69.0	-	19	10
OMTD[38]	79.7	56.3	-	19	-
OMAT[39]	92.8	74.3	-	19	8
PMPTCS[40]	76.0	53.8	-	19	-
OGOMT[41]	98.1	80.5	-	19	9
CSL-VOX[42]	89.78	-	98.28	19	6
CSL-DPT[42]	88.13	-	97.64	19	8
CML[13]	92.1	86.4	95.1	19	28
TSML[13]	93.4	86.4	96.0	19	18
TD[13]	93.7	86.3	96.6	19	13
TSML+TD[13]	94.7	86.4	97.2	19	7
TSML + TD+WP[13]	95.3	86.4	97.4	19	4
OURS PCNG	96.10	**96.27**	**99.77**	19	24

Table 2: Comparison table of trackers on PETS2009-S2L1 data sequence

Overall tracking performance of our proposed algorithm on PETS2009-S2L1 video sequence is presented in Table 2. It performs favourably and is reflected through MOTA, MOTP and Recall metrics. Proposed PCNG tracker when compared with listed trackers, achieves the best performance in terms of 96.274% MOTP and 99.77 % Recall. It also achieves competitive performance in terms of MOTA compared to state-of-art methods. It is 96.10% compared to maximum 98%. Use of the same ground truth as defined in [43] helped us in a fair comparison.

CONCLUSIONS

An affinity model for the tracks considering the appearance cues based on similarity measures and motion cues based on coherent dynamic

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

estimation are combined in the proposed method. Tracklet association is performed on the basis of affinity distance matrix. It is found to be consistent over a longer period. Thus, the proposed PCNG algorithm exploits both motion and appearance cues which reduce identity switches with improved accuracy. In this research ETH-Bahnhof dataset is used to show the efficacy of the proposed method in terms of identity switches over five methods listed in the experimentation. The method shows the competitive performance in terms of MOTA, MOTP and Recall when tested on PETS2009-S2L1 dataset.

The proposed method gives 96.27 MOTP and 99.77% recall, which is the highest on PETS2009-S2L1 dataset. The proposed PCNG tracker is capable of handling cases of missed detections and their re-appearance. It is also observed that proposed method is effective in cases of occlusion, merge and split, where appearance or motion cues fail. The proposed tracking algorithm is validated on public benchmarks and performs better than state-of-the-art tracking algorithms.

REFERENCES

Andriyenko, A., & Schindler, K. (2011). 'Multi-target tracking by continuous energy minimization', CVPR.

Andriyenko, A., Schindler, K., & Roth, S. (2012). Discrete-continuous optimization for multi-target tracking. In CVPR, 1926–1933. https://doi.org/10.1109/CVPR.2012.6247893

Butt, A. A., & Collins, R. T. (2013). Multi-target tracking by Lagrangian relaxation to Min-cost network flow. In *CVPR*, 1846–1853. https://doi.org/10.1109/CVPR.2013.241

Clausi, D. A., & Zhao, Y. (2003). Grey level co-occurrence integrated algorithm (glcia): Asuperior computational method to rapidly determine coccurrencepobability texture features. *Computers and Geosciences*, *29*(7), 837–850. https://doi.org/10.1016/S0098-3004(03)00089-X

Deshpande, H. S., Karande, K. J., & Mulani, A. O. (2015, April). Area optimized implementation of AES algorithm on FPGA. In *2015 International Conference on Communications and Signal Processing (ICCSP)* (pp. 0010-0014). IEEE.

Du, Guanglong, Zhang, P., & Liu, X. (2016). Markerless human–manipulator interface using leap motion with interval kalman filter and improved particle filter. *IEEE Transactions on Industrial Informatics*, *12*(2), 694–704. https://doi.org/10.1109/TII.2016.2526674

Godse, A. P., & Mulani, A. O. (2009). *Embedded systems*. Technical Publications.

Gotlieb, C. C., & Kreyszig, H. E. (1990). Texture descriptors based on co-occurrencematrices. *Computer Vision, Graphics, and Image Processing*, *51*(1), 70–86. https://doi.org/10.1016/S0734-189X(05)80063-5

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, *6*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hofmann, M., Wolf, D., & Rigoll, G. Hypergraphs for joint multi-view reconstruction and multi-object tracking, *CVPR* pages 3650–3657. (2013). https://doi.org/10.1109/CVPR.2013.468

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(series 1), 35–45. https://doi.org/10.1115/1.3662552

Kashid, M. M., Karande, K. J., & Mulani, A. O. (2022, November). IoT-Based Environmental Parameter Monitoring Using Machine Learning Approach. In *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1* (pp. 43-51). Singapore: Springer Nature Singapore.

Kovesi, P. (2003). Phase congruency detects corners and edges. In C. Sun, H. Talbot, S. Ourselin & T. Adriaansen (Eds.). *Proceedings of the VIIth Digital Image Computing: Techniques and Applications*, *10–12*.

Kulkarni, P., & Mulani, A. O. (2015). Robust invisible digital image watermarking using discrete wavelet transform. *International Journal of Engineering Research & Technology (IJERT)*, *4*(01), 139-14Liyakat, K. K. S.,

Ling, T. S., Meng, L. K., Kuan, L. M., Kadim, Z., & Baha'a Al-Deen, A. A. (2009). Colour-based object tracking in surveillance application. In *Proceedings of the International Multiconference of Engineers and Computer Scientists*, *I*, IMECS.

Mane, D. P., & Mulani, A. O. (2019). High throughput and area efficient FPGA implementation of AES algorithm. *International Journal of Engineering and Advanced Technology*, *8*(4).

Mandwale, A. J., & Mulani, A. O. (2015, January). Different Approaches For Implementation of Viterbi decoder on reconfigurable platform. In *2015 International Conference on Pervasive Computing (ICPC)* (pp. 1-4). IEEE.

McKenna, S. J., Jabri, S., Duric, Z., Rosenfeld, A., & Wechsler, H. (2000). Tracking groups of people. *Computer Vision and Image Understanding*, *80*(1), 42–56. https://doi.org/10.1006/cviu.2000.0870

Milan, A., Leal-Taixé, L., Schindler, K., & Reid, I. (2015). Joint tracking and segmentation of multiple targets. *Proceedings of the IEEE*.

Milan, A. (2016). Laura leal-Taix´ e, Ian Reid, Stefan Roth, Konrad Schindler. *MOT16: A benchmark for multi-object tracking*, arXiv:1603.00831v2. cs.CV

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023

Mulani, A. O., & Mane, P. B. (2019). High-Speed area-efficient implementation of AES algorithm on reconfigurable platform. *Computer and Network Security*, *119*.

Mulani, A. O., & Mane, P. B. (2014, October). Area optimization of cryptographic algorithm on less dense reconfigurable platform. In *2014 International Conference on Smart Structures and Systems (ICSSS)* (pp. 86-89). IEEE.

Mulani, A. O., & Mane, P. B. (2017). Watermarking and cryptography based image authentication on reconfigurable platform. *Bulletin of Electrical Engineering and Informatics*, *6*(2), 181-187.

Mulani, A. O., Jadhav, M. M., & Seth, M. (2022). Painless Non-invasive blood glucose concentration level estimation using PCA and machine learning. *the CRC Book entitled Artificial Intelligence, Internet of Things (IoT) and Smart Materials for Energy Applications*.

Sinha, S. N., Frahm, J.-M., Pollefeys, M., & Genc, Y. (2011). Feature tracking and matching in video using programmable graphics hardware. *Machine Vision and Applications*. Springer-Verlag, 22(1), 207–217. https://doi.org/10.1007/s00138-007-0105-z

Yang, M., & Jia, Y. (2016). Temporal dynamic appearance modeling for online multi-person tracking. *Computer Vision and Image Understanding*, 153, 16–28. https://doi.org/10.1016/j.cviu.2016.05.003

Warhade, N. S., Pol, R. S., Jadhav, H. M., & Mulani, A. O. (2022). Yarn Quality detection for Textile Industries using Image Processing. *Journal Of Algebraic Statistics*, *13*(3), 3465-3472.

Wang, B., & Luk Chan, K. (2017). Tracklet association by online target-specific metric learning and coherent dynamics estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Volume: 39 , Issue: 3.

Wu, J. M., & Chen, Y. C. (1992). Statistical feature matrix for texture analysis, *Computer Vision, Graphics and Image Processing*. *Graphical Models and Image Processing*, *54*, 407–419.

Zhang, L., Li, Y., & Nevatia, R. (2008). Global data association for multi-object tracking using network flows, IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). https://doi.org/10.1109/CVPR.2008.4587584

Depósito Legal: PPX200002ZU2142 / ISSN:1856-4194. Volumen 22 Edición No .1 – Año 2023