



UWS Academic Portal

Dense compounded features and data association based multiple people tracking

Al-Faris, Mahmoud ; Chiverton, John; Ndzi, David; Alhabo, Mohanad

Published in:

2021 International Conference on Communication & Information Technology (ICICT)

DOI:

[10.1109/ICICT52195.2021.9568438](https://doi.org/10.1109/ICICT52195.2021.9568438)

Published: 26/10/2021

Document Version

Peer reviewed version

[Link to publication on the UWS Academic Portal](#)

Citation for published version (APA):

Al-Faris, M., Chiverton, J., Ndzi, D., & Alhabo, M. (2021). Dense compounded features and data association based multiple people tracking. In *2021 International Conference on Communication & Information Technology (ICICT)* (pp. 25-30). IEEE. <https://doi.org/10.1109/ICICT52195.2021.9568438>

General rights

Copyright and moral rights for the publications made accessible in the UWS Academic Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact pure@uws.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Al-Faris, M., Chiverton, J., Ndzi, D., & Alhabo, M. (2021). Dense compounded features and data association based multiple people tracking. In *2021 International Conference on Communication & Information Technology (ICICT)* IEEE. <https://doi.org/10.1109/ICICT52195.2021.9568438>

“© © 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Dense Compounded Features and Data Association Based Multiple People Tracking

1st Mahmoud Al-Faris

mahmoud.al-faris1@myport.ac.uk

2nd John Chiverton

john.chiverton@port.ac.uk

3rd David Ndzi

david.ndzi@uws.ac.uk

4th Mohanad Alhabo

mohanad.alhabo@gmail.com

^{1,4} Informatics & Telecommunications Public Company, Ministry of Communications, Iraq

² School of Energy and Electronic Engineering, University of Portsmouth, UK

³ School of Computing, Engineering and Physical Sciences, University of the West of Scotland, UK

Abstract—People tracking is a wide area of research in computer vision and machine learning. The challenging problem of multiple objects tracking (MOT) is further complicated by factors such as occlusion, varying number of targets, illumination variations and objects' appearances which may be similar.

In this paper, a significant approach is proposed for MOT with a single static camera based on dense valuable global and local features. This includes cascaded HOG-III features, texture, and motion information of objects in order to build a robust tracking system. To speed up the system, a simple data association method is employed using Hungarian algorithm to associate candidate response to the target objects. Genetic algorithm is also used to provide a heuristic data association based multiple objects tracking. A comparison between two association algorithms is made based on the tracking results.

TUD-crossing and TUD-campus datasets are used for validation purposes. A system's performance can be evaluated using a wide metrics of MOT performance indicators (MOTA, MOTP). The results reach (82.89 and 81.96) in terms of TUD-crossing dataset and (79.30 and 73.06) in terms of TUD-campus dataset respectively. The experiments show that the proposed method can be suitably employed during scale changes or in the presence of a cluttered background environment, in addition, our method achieves competitive results in comparison with state of the art approaches.

Keywords—Tracking, Association, Features, MOT

I. INTRODUCTION

Computer vision is a widely researched area that can be utilised in different fields such as detection and analysis [1], action recognition [2] and tracking. Tracking an object can be used in a wide range of practical applications like surveillance, sport analysis, medical diagnosis, traffic analysis, pattern recognition, human computer interaction etc [3]. Tracking objects in a video depends on several factors, such as the amount of prior knowledge about the object, and the number and the type of parameters being tracked (e.g., location, scale and contour around the object) [4].

A typical tracking system consists of three key components: an appearance model, which can evaluate the likelihood that the object is at some particular location; a motion model, which relates the locations of objects over time; and a search strategy for finding the most likely location in the current image [4], [5].

Multiple Object Tracking (MOT), or Multiple Target Tracking (MTT), is an extension of the well-studied single object tracking [6], [7]. MOT plays an important role in computer vision. The task of MOT, given an input video [8], is largely partitioned into: locating multiple objects; maintaining their identities; and yielding their individual trajectories.

The various applications above have sparked enormous interest in this topic. However, compared with Single Object Tracking (SOT) which primarily focuses on designing

sophisticated appearance models or motion models to deal with challenging factors such as scale changes, out-of-plane rotation and illumination variations, multiple object tracking additionally requires maintaining the identities among multiple objects. Besides the common challenges in both SOT and MOT, the further key issues making MOT challenging can include: frequent occlusions, initialization and termination of tracks, small size of objects, similar appearance among objects, and interaction between multiple objects [6], [7], [8].

This paper proposes a multiple-object tracking system to track people in an indoor environment based on valuable dense features. A combination of different features is introduced to handle the issues mentioned above and to provide shape and motion information of a person that's being tracked. Two methods of association over time are compared, namely the Hungarian algorithm and a genetic algorithm.

The rest of paper is structured as follows: first, we give an overview and research potential problems of MOT systems. Second, The proposed method is explained in detail. Last, we describe the datasets, experiments and results in addition to a comparison with a number of state-of-the-art approaches.

II. RELATED WORK

In general, MOT system consists of two processing levels including observation and tracking. The first, describes the tracked object with unique features which can be used to recognise, identify, detect and track objects. The second, links the extracted observation to the trajectories. Accordingly, MOT methods could be classified into Bayesian and data association methods. Bayesian methods estimate the trajectories with Bayesian theory using a prediction process. Whereas, association methods associate the observations to the trajectories using an affinity function between them [6].

An observation based model, in turn, can be classified into appearance, motion exclusion and an integrated model which combines both of them based on the used features [6], [8].

The main categories of MOT approaches are direct tracking and tracking-by-detection approaches [9]. The first deals with finding objects in the first frame and extracting effective features, thereafter, looking for extracted features in the successive frames using tracking techniques. The second, objects are being detected in each frame and then tracked across frames using association with features of the detected objects [3]. The latter category has the ability to deal with occlusion issues although it requires more computational processes due to its need for detection in each single frame [10].

MOT system can be optimised either by improving detection, association, or feature extraction processes to alleviate frequent challenges.

Many studies have been carried on multiple object tracking. A Kalman filter has been used by [11], [12], [13] to introduce MOT system. The studies were based on RANSAC,

background and shape, and mixture of Gaussian matching algorithms respectively. Another study was proposed by [14] using confidence-based relative motion network. The method was utilised tracklet confidence and relative motion network data associations for tracking.

Other researchers focused on data association based multiple object tracking model to improve association errors in future steps. Joint probabilistic data association method was improved by [15] to track various numbers of objects instead of a fixed number of targets. Good results were provided but with an incorrect decision in a cluttered environment. A multiple objects tracking for pedestrians surveillance was introduced by [16] used energy minimisation association algorithm. Moreover, a multiple hypothesis tracking (MHT) algorithm was used by [17] to preserve multiple hypothesis associations of previous observations with a target. The method was verified via simulation for moving object tracking. MHT algorithm was also used by [18] with dynamic track management to provide multiple object tracking system and applied to vehicle tracking.

The aforementioned approaches could improve MOT performance, at the same time, more computation complexity can be added into the system due to the use of a graph structure and an object tracker. Therefore, valuable dense features of appearance and motion information are used in this work with a simple data association algorithm to create less complex MOT system.

Some researchers have followed the same trend such as [19], where three independent features including colour histogram, spatial and motion information were used to model each target for MOT. The same was introduced by [10], where Kanade-Lucas-Tomasi (KLT) tracker was used in conjunction with speeded-up robust features (SURF) and gradient weighted optical flow (GWOFF) features to implement the MOT system. MOT system based on multiple visual cues was introduced by [20]. This method presented locality sensitive histograms using a sparse representation and color information. A tracking-by-detection MOT was presented by [21]. This was based on the continuous confidence of detectors and online trained classifier.

In this paper, inspiration has been taken from these approaches to build an MOT model based on a pool of robust features including HOG-III [22], optical flow [23] and texture information [24] with a simple data association algorithm for identifying each target.

III. METHODOLOGY

Our proposed MOT system is shown in Fig.1. Individual tracks are started from manually provided initial detection responses which could come from motion detection or similar processes, see e.g. [25]. An affinity function is then calculated for each frame based on the similarity between the candidate and target. Different features are used to describe appearance and motion information. Thereafter, the assignment stage is achieved using either Hungarian or genetic algorithm.

A. HOG-III

HOG-III was used by [26] for human detection. In this paper, HOG-III features are used in an efficient way by discarding the duplicated information and keep the useful information to be fed into MOT system. HOG-III consists of a combination of complementary features these are HOG features, colour features and bar-shape features combined with a cell based histogram stand on ignoring frequent information.

1) HOG

Histogram of oriented gradients was introduced by [27] and used for people detection. It is widely used in computer vision for different applications such as human action recognition [28]. In this paper, HOG features are computed and utilised in MOT system and can be explained by the following steps:

- Gradient computation: in this stage, interesting saliency can be detected by filtering to compute the horizontal and vertical gradients. The kernel is used in the vertical and horizontal directions. Later, the magnitude G and orientation θ of the gradient could be calculated using the vertical and horizontal gradients (I_x, I_y) .
- Orientation binning: the histogram of individual cells are computed using \mathcal{Q} orientation bins on the interval of $[0^\circ, 180^\circ]$. The corresponding bin of each pixel is found based on the orientation and related magnitude of the pixel.
- Histogram blocks: cells are grouped into larger blocks to achieve illumination and in-variance representation. There are two kinds of block geometry including rectangular and circular HOG. In this paper, the rectangular geometry is used.

2) Histogram of colour (HOC)

To benefit from the colour information in the tracking system, we convert each salience area in each frame of each RGB image to Hue-Saturation-Intensity space. This provides an opportunity to separate the mixture of RGB information and to avoid intensity information which is used already with the HOG process. Therefore, Hue and Saturation information is maintained in HSI space, to be processed based on histograms of colour.

3) Histogram of bar-shape (HOB)

Bar-shape information is related to the second order gradients as mentioned in [22], [29]. The second-order gradients can be computed as follows:

$$r^* = \max_{\theta} \frac{\partial^2 I}{\partial^2 u}, \quad \theta^* = \arg \max_{\theta} \frac{\partial^2 I}{\partial^2 u} \quad (1)$$

where I is the intensity value of the input image, and $u = (\cos \theta, \sin \theta)$ is the unit direction vector. They can be solved by equating the derivatives to zero to obtain:

$$\theta^* = \frac{1}{2} \tan^{-1} \left(\frac{2I_{xy}}{I_{xx} - I_{yy}} \right) \quad (2)$$

$$r^* = I_{xx} \cos^2 \theta^* + 2I_{xy} \cos \theta^* \sin \theta^* + I_{yy} \sin^2 \theta^*. \quad (3)$$

where I_{xx}, I_{xy}, I_{yy} are the second-order derivatives of I with respect to the corresponding orientations. After we get the second-order gradient r_{xy}^*, θ_{xy}^* at each pixel (x, y) , by following the same steps of HOG computation but with second order gradients. Histograms of bar-shape featured result in the representation of the distribution of bar-shapes in each frame [22].

4) Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix (GLCM) is a statistical approach used to identify texture features of pixels. This method identifies certain texture properties in an image in terms of spatial distribution of gray levels.

GLCM matrix could be generated after calculation of how often a pixel with the intensity value (a) occurs in a specific

spatial related to a pixel with the value (\mathbf{b}). Different variables affect the distribution in the matrix such as distance, angular relations and directions (horizontal, vertical, diagonal, anti-diagonal) between the pixels. Many kinds of texture features were proposed and extracted from gray level co-occurrence matrices such as in [30].

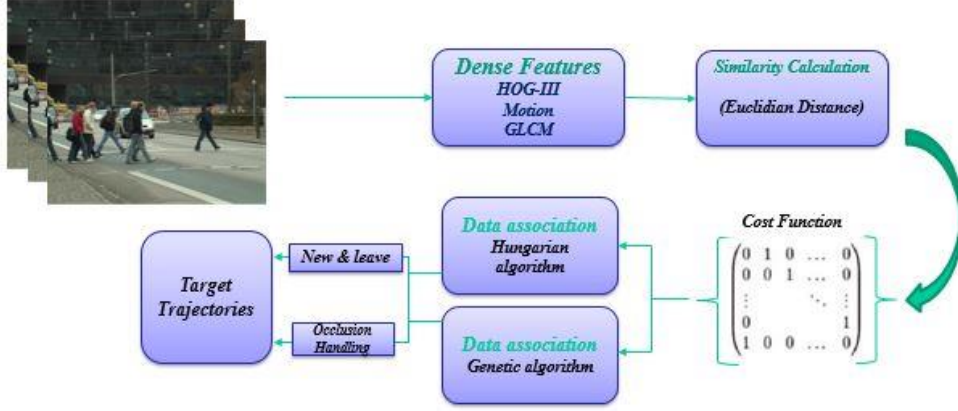


Figure 1: Framework design of the proposed method

To calculate a GLCM, suppose an image has $N \times M$ dimensions. The horizontal space is $M_x = 1, 2, 3 \dots N$ and the vertical space is $M_y = 1, 2, 3 \dots M$, in addition, suppose that i and j are different pixels values is graded through Z number of levels and occur in the relation given by offset, Thereafter, with the distance $\Delta x, \Delta y$, GLCM can be calculated such that:

$$P_{\Delta x, \Delta y}(i, j) = \sum_{x=1}^N \sum_{y=1}^M \begin{cases} 1 & \text{if } I(x, y) = i \\ & \text{and } I(x + \Delta x, y + \Delta y) = j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

GLCM can also be calculated based on distance and direction instead of $\Delta x, \Delta y$.

$$P(i, j | \Delta x, \Delta y) = \frac{p(i, j | \Delta x, \Delta y)}{\sum_i \sum_j p(i, j | \Delta x, \Delta y)} \quad (5)$$

Eventually, contrast features of texture information can be extracted from GLCM matrix using:

$$Contrast = \sum_{i,j=0}^{Z-1} P_{ij}(i - j)^2. \quad (6)$$

where P_{ij} is a GLCM element.

5) Optical flow

In this work, Horn-Schunck [23] method is used to estimate optical flow information for each target. This algorithm minimises distortion in flow and assuming smoothness in the flow over the whole image and formulates the flow as an energy function over D domain (image). Horn-Shuck computes estimated velocity field that minimises the following equation:

$$E = \int_D [(I_x u + I_y v + I_t)^2 + \alpha^2 \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right]] dx dy, \quad (7)$$

where D is a spatial domain indicates to the image, I_x, I_y and I_t are the derivatives of image intensity values in terms of (x, y) and *time* dimensions respectively, u and v is the horizontal and vertical optical flow vectors, α is a regularisation constant

which is proportional to the flow smoothness, $\frac{\partial u + \partial v}{\partial x}$ and $\frac{\partial u + \partial v}{\partial y}$ are spatial derivatives of velocity component. This method minimises the above equation to attain the velocity field, $[u \ v]$, for each pixel in the image from the Gauss Seidel equations that solve the appropriate Euler-Lagrange equations:

$$u_{x,y}^{k+1} = \bar{u}_{x,y}^k - \frac{I_x [I_x \bar{u}_{x,y}^k + I_y \bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}, \quad (8)$$

$$v_{x,y}^{k+1} = \bar{v}_{x,y}^k - \frac{I_y [I_x \bar{u}_{x,y}^k + I_y \bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}. \quad (9)$$

where k is the iteration number, $[u_{xy} \ v_{xy}]$ is the estimated velocity of the pixel at (x, y), v_y^0 and v_x^0 denote to initial velocity which is zero; $\bar{v}_{x,y}^k$ and $\bar{u}_{x,y}^k$ are neighborhood averages of $v_{x,y}$ and $u_{x,y}$.

Although, this method is more sensitive to noise, however, it has the advantage of a high density of flow vectors [23].

B. Data association

In this system, one class corresponds to the candidate objects. The other corresponds to the target objects. Euclidean distance is used to return a matrix containing distances between each pair of candidate and target observations (minimum distance means much similar), in which, refers to similarity likelihood measurements that represent the entries of assignment algorithm [31]; used here as:

$$D(i, j) = \sqrt{(i_x - j_x)^2 + (i_y - j_y)^2}. \quad (10)$$

where (i_x, i_y) and (j_x, j_y) are the features vector parameters of a target and a candidate respectively.

Then, the association algorithm does the best assignment corresponds to matching measurements [32].

This work goes further to present a comparison between two data association algorithms (Hungarian & genetic algorithms).

1) Hungarian Algorithm

To assign candidates to targets in the process of multi-objects tracking, variant of the Hungarian assignment algorithm is used [33]. This method determines which tracks that were missing and which detection should begin a new track. This method returns the indices of assigned tracks [34].

Hungarian algorithm can solve assignment problems in polynomial time $O(n^3)$. The first step for this algorithm is a square cost matrix C which is based on non-negative matrix C_0 as:

$$C_0 = (d_{ij}), \quad (11)$$

where d_{ij} is the matching costs of candidate and target observations.

In the case that C_0 is not square, the matrix C needs to be set up to the appropriate rows or columns to zero.

$$C = \begin{cases} C_0 & n_e = n_m \\ \begin{bmatrix} C_0 & 0_{n_e \times (n_e - n_m)} \\ 0_{(n_m - n_e) \times n_m} & C_0 \end{bmatrix} & n_e > n_m \\ \begin{bmatrix} C_0 & 0_{n_e \times (n_e - n_m)} \\ 0_{(n_m - n_e) \times n_m} & C_0 \end{bmatrix} & n_e < n_m, \end{cases} \quad (12)$$

The output of Hungarian algorithm is the optimum assignment matrix X as following expression [35]:

$$X = \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij}, \quad (13)$$

$$X_{ij} = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if } e_i \text{ is assigned to } m_j. \end{cases} \quad (14)$$

where e_i and m_j are the target and measurement objects respectively.

2) Genetic algorithm

Genetic algorithm (GA) is proposed in the data association stage as an alternative assignment algorithm in order to assign candidates to targets. This can provide a comparison between both association algorithms at the end. Genetic algorithm is usually mimics biological evolution to solve optimisation problems. It repeatedly uses a population of individuals and selects individuals from a current population to be used as parents producing the next generation's children. over different consecutive generations, the population will ultimately be an optimal solution which in our case the optimal assignment.

The basic steps of genetic algorithm can be simplified as: Fitness function based optimisation, A population of chromosomes, Specify chromosomes that going to reproduce, Crossover process and Mutation of new generation chromosomes [36].

Parents are chosen in this algorithm based on the fitness value of each member of the current population, however, other individuals that have low fitness value are directed to the next population. For example, if f is a positive fitness function, then the probability that the chromosome C_1 is chosen to be selected can be:

$$P(C_1) = \left| \frac{f(C_1)}{\sum_{i=1}^N f(C_i)} \right|. \quad (15)$$

Children are produced from parents either by a single parent mutation or a pair of parents crossover, to replace the current population and start the next generation [37].

C. Occlusion issues

MOT system suffers from different issues that need to be handled such as objects occlusion, miss detection and false detection. In our case, if the target is assigned to a candidate, this means there is no occlusion. Whereas a potential issue that can happen and needs to be handled is when a target is not assigned; this indicates either an occluded target or a missing detection.

In order to handle this issue, the target's state will be considered a 'passive' if the target is not assigned. The target's state will change back to an 'active' if the target gets assigned during an occlusion time threshold (20 frames).

Furthermore, the target is considered as entering or leaving based on its position of the locations of which are selected manually. If the target is not assigned at the current time and was located in the leaving area at previous time, then, the target status will be 'leave'. Whereas if the detection response is not being assigned and located in the entering area, then, the person considered as new and a new target should be added to the set of tracks.

D. Dataset

In this study, TUD-crossing and TUD-campus datasets [38] are used for validation purposes. These datasets provide 50% annotations for all occluded pedestrians and have been used before to evaluate detection by tracking and tracking by detection systems.

These datasets consist of walking pedestrian in an outdoor environment with a poor light condition. TUD-Crossing consists of 201 frames sequence with 640X480 resolution and showing a road crossing from a side view, in which, targets are frequently occluded by each other in addition to other objects. Detection responses are already provided with the video sequence. TUD - Campus is a short sequence dataset with side-view pedestrians consisting of 71 frames with 640x480 resolution and the targets are frequently occluded by each other.

System performance can be evaluated using widely metrics of clear MOT [39]. These metrics include accuracy score (MOTA) which combines false positive, missed targets and identity switched errors; and precision score (MOTP) which is the average distance between ground truth and predicted targets; in addition to false positive (FP) and false negative (FN).

These metrics could be calculated such as:

$$MOTP = \frac{\sum_i \sum_t d_t^i}{\sum_t c_t} \quad (16)$$

which is the error in estimated position for the whole frames, averaged by the total number of matches.

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \quad (17)$$

where m_t , fp_t , mme_t are the number of misses, false positives and mismatches, respectively, for time t . It's obvious that MOTA is derived from three error ratios which are:

$$\bar{m} = \frac{\sum_t m_t}{\sum_t g_t} \quad (18)$$

the ratio of misses in the sequence, computed over the total number of objects present in all frames,

$$\bar{fp} = \frac{\sum_t fp_t}{\sum_t g_t} \quad (19)$$

the ratio of false positive, and

$$m\bar{m}e = \frac{\sum_t mme_t}{\sum_t g_t}. \quad (20)$$

the ratio of mismatches.

IV. RESULTS

Two different association algorithms are used in the experiments in order to achieve further engagement in the

tracking system. GA and Hungarian algorithms are used to show their effects on the tracking system in terms of the accuracy and computation complexity.

Table I includes a comparison between the two algorithms. It's obvious that genetic algorithm outperforms Hungarian algorithm regarding tracking accuracy (MOTA and MOTP) as it achieves 82.89% and 81.96% respectively in terms of TUD-Crossing dataset. Furthermore, it achieves 79.3% and 73.06% in terms of TUD-Campus dataset.

Table I: Comparison of Hungarian and Genetic algorithms in terms of TUD-crossing and TUD-campus datasets

Dataset	Algorithm	MOTA	MOTP	FN	FP
TUD-Crossing	GA	82.89	81.96	2.60	1.40
	Hungarian	80.10	72.60	1.30	1.60
TUD-Campus	GA	79.30	73.06	1.40	1.70
	Hungarian	74.80	73.00	1.86	1.29

However, Genetic algorithm took longer computation time compared to Hungarian algorithm due to multiple iterations that required to generate a population of points.

The experiments reveal that the proposed approach is more robust against the variability of objects' style, appearance and scales. It achieved competitive results in comparison with state-of-the-art methods as included in Table II.

Table II: Our approach compared to state of the art approaches

Dataset	Method	MOTA	MOTP	FN	FP
TUD-Crossing	[19]	78.30	66.00	1.4	8.38
	[20]	72.00	76.00	26	1
	[21]	84.30	71.00	14.1	1.4
	Proposed	82.89	81.96	2.6	1.4
TUD-Campus	[19]	78.18	69.00	0	13
	[20]	72.00	74.00	25	2
	[21]	73.30	67.00	26.4	0.1
	Proposed	79.30	73.06	1.4	1.7

In terms of TUD-Crossing, our method outperforms the approaches in [19] and [20] regarding MOTA and MOTP. Although the MOTA scores are a little lower than in [21] due to a big similarity in objects appearance but still our approach provides a robust result in tracking with a flexible system.

For TUD-Campus, our method outperforms [19], [20] and [21] regarding MOTA and MOTP.

Moreover, our method assures that a robust tracking system could be introduced by using a simple data association with a robust pool of combined features. The results show that our method performs well in the presence of scale and background changes. In addition to the effective engagement with objects occlusion that's found in many scenes of datasets.

V. CONCLUSION

A tracking system is presented in this work for multiple people tracking. This system is based on two main steps those are observation and association levels.

A pool of dense valuable features is used in the observation level include HOG-III, texture, and optical flow to provide suitable appearance and motion information of the target. Hungarian and Genetic algorithms are used in the association level one each time. A comparison between the two algorithms is introduced in terms of tracking accuracy and computational time. This method handled different issues including occlusion, varying objects number and objects' similarity.

The proposed method is compared to several state-of-the-art approaches. The results demonstrated the efficiency of our method and its competitive achievement on all tested videos.

VI. REFERENCES

- [1] M. Al-Faris, J. Chiverton, D. Ndzi and I. Ahmed, "A review on computer vision-based methods for human action recognition," *Journal of Imaging*, vol. 46, p. 6(6), 2020.
- [2] M. Al-Faris, J. Chiverton, Y. Yang and D. Ndzi, "Multi-view region-adaptive multi-temporal DMM and RGB action recognition," *Pattern Analysis and Applications*, vol. 23, pp. 1587--1602, 2020.
- [3] A. Sharma, "Robust tracking for multiple objects in a video," in *Computing for Sustainable Global Development (INDIACom)*, IEEE, 2016, pp. 2437--2442.
- [4] R. C. Philip, S. Ram, X. Gao and J. J. Rodriguez, "A comparison of tracking algorithm performance for objects in wide area imagery," in *Image Analysis and Interpretation (SSIAI)*, Southwest Symposium, 2014.
- [5] B. Sahbani and W. Adiprawita, "Kalman filter and Iterative-Hungarian Algorithm implementation for low complexity point tracking as part of fast multiple object tracking system," in *System Engineering and Technology (ICSET)*, IEEE, 2016.
- [6] L. Fan, Z. Wang, B. Cail, C. Tao, Z. Zhang, Y. Wang, S. Li, F. Huang, S. Fu and F. and Zhang, "A survey on multiple object tracking algorithm," in *Information and Automation (ICIA)*, IEEE International Conference, 2016.
- [7] A. Yilmaz, O. Javed and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol. 38, p. 13, 2006.
- [8] W. Luo, J. Xing, X. Zhang, X. Zhao and T.-K. and Kim, "Multiple object tracking: A literature review," *arXiv preprint arXiv:1409.7618*, 2014.
- [9] S. Yi, H. Wang, Z. He, Y. Li and W. Chen, "Multiple Object Tracking Based on Robust Network Flow Model," in *Computational Intelligence and Security (CIS)*, 12th International Conference, 2016.
- [10] V. Buddubariki, S. G. Tulluri and S. and Mukherjee, "Multiple object tracking by improved KLT tracker over SURF features," in *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Fifth National Conference, 2015.
- [11] E. Apostolidis, V. Mezaris and I. Kompatsiaris, "Fast object re-detection and localization in video for spatio-temporal fragment creation," in *Multimedia and Expo Workshops (ICMEW)*, IEEE International Conference, 2013.
- [12] L. Marcenaro, M. Ferrari, L. Marchesotti and C. S. Regazzoni, "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching," in *Image Processing Proceedings, IEEE International Conference*, 2002.

- [13] D. Kangin, D. Kolev and G. Markarian, "Multiple video object tracking using variational inference," in *Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2015.
- [14] S.-H. Park, K. Lee and K.-J. Yoon, "Robust online multiple object tracking based on the confidence-based relative motion network and correlation filter," in *Image Processing (ICIP), IEEE International Conference*, 2016.
- [15] M. Schutz and K. Dietmayer, "A flexible environment perception framework for advanced driver assistance systems," in *Advanced Microsystems for Automotive Applications*, Springer, 2013, pp. 21--29.
- [16] L. Wei and L. Xingwei, "Multiple object tracking based on energy minimization," in *Computational and Business Intelligence (ISCBI), 4th International Symposium*, 2016.
- [17] R. Ding and W.-H. Chen, "Moving object tracking in support of unmanned vehicle operation," in *Automation and Computing (ICAC), 19th International Conference*, 2013.
- [18] A. H. A. Rahman, H. Zamzuri, S. A. Mazlan, M. A. A. Rahman and M. A. Zakaria, "Tracking uncertain moving objects using dynamic track management in Multiple Hypothesis Tracking," in *Connected Vehicles and Expo (ICCVE), International Conference*, 2014.
- [19] D. Riahi and G.-A. Bilodeau, "Multiple object tracking based on sparse generative appearance modeling," in *Image Processing (ICIP), IEEE International Conference*, 2015.
- [20] D. Riahi and G.-A. Bilodeau, "Multiple feature fusion in the dempster-shafer framework for multi-object tracking," in *Computer and Robot Vision (CRV), IEEE Canadian Conference*, 2014.
- [21] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. and Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, pp. 1820--1833, 2011.
- [22] Y. Jiang and J. and Ma, "Combination features and models for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [23] B. K. Horn and B. G. and Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, pp. 185--203, 1981.
- [24] N. Bagri and P. K. Johari, "A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval," *International Journal of Advanced Science and Technology*, vol. 80, pp. 41--52, 2015.
- [25] J. Chiverton, X. Xie and M. Mirmehdi., "Automatic Bootstrapping and Tracking of Object Contours," *IEEE Transactions on Image Processing*, vol. 21, pp. 1231--1245, 2012.
- [26] Y. Jiang and J. and Ma, "Combination features and models for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society Conference*, 2005.
- [28] M. Al-Faris, J. Chiverton, L. Yang and D. Ndzi, "Appearance and motion information based human activity recognition," in *IET 3rd International Conference on Intelligent Signal Processing (ISP)*, London, 2017.
- [29] N. Dalal, Finding people in images and videos, Institut National Polytechnique de Grenoble-INPG, 2006.
- [30] R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. 3, pp. 610--621, 1973.
- [31] M. M. Deza and E. and Deza, *Encyclopedia of Distances*, Springer, 2009.
- [32] M. Rowan and F. D. Maire, "An efficient multiple object vision tracking system using bipartite graph matching," in *FIRA Robot World Congress*, 2004.
- [33] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the society for industrial and applied mathematics*, vol. 5, pp. 32--38, 1957.
- [34] S. P. Nandashri D, "An Efficient Tracking of Multi Object Visual Motion using Hungarian Method," *International Journal of Engineering Research & Technology*, vol. 4, 2015.
- [35] F. Lutteke, X. Zhang and J. Franke, "Implementation of the hungarian method for object tracking on a camera monitored transportation system," in *Robotics; Proceedings of ROBOTIK ; 7th German Conference*, 2012.
- [36] J. Carr, "An introduction to genetic algorithms," *Senior Project*, pp. 1--40, 2014.
- [37] D. Hermawanto, "Genetic algorithm for solving simple mathematical equality problem," in *Proc. Int. conf. on Computational and Mathematics*, 2013.
- [38] M. Andriluka, S. Roth and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Computer Vision and Pattern Recognition, CVPR 2008. IEEE Conference*, 2008.
- [39] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1--10, 2008.