Robust Multi-object Tracking with Semantic Color Correlation

Noor M. Al-Shakarji
University of Missouri Columbia
MO 65211, USA
nmahyd@mail.missouri.edu

Filiz Bunyak
University of Missouri Columbia
MO 65211, USA
punyak@missouri.edu

Guna Seetharaman
U.S Naval Research Laboratory
Washington, DC 20375, USA
Gunasekaran.Seetharaman@rl.af.mil

Kannappan Palaniappan
University of Missouri Columbia
MO 65211, USA
palaniappank@missouri.edu

Abstract

Multi-object tracking is an important computer vision task with wide variety of real-life applications from surveillance and monitoring to biomedical video analysis. Multi-object tracking is a challenging problem due to complications such as partial or full occlusions, factors affecting object appearance, object interaction dynamics, etc. and computational cost. In this paper, we propose a detection-based multi-object tracking system that uses a two-step data association scheme to ensure time efficiency while preserving tracking accuracy; a robust but discriminative object appearance model that compares object color attributes using a novel color correlation cost matrix; and a framework that handles occlusions through prediction. Our experiments on UA-DETRAC multi-object tracking benchmark dataset consisting of challenging real-world traffic videos show promising results against state-of-the-art trackers.

1. Introduction

Visual tracking is the process of locating objects of interest in a video sequence and maintaining their identity over time. Tracking is an important task in computer vision with many real-life applications e.g. surveillance and monitoring, video summarization, robot navigation etc., and more recently autonomous driving. Two main categories of visual tracking are single object tracking and multi-object tracking. This work focuses on multi-object tracking (MOT).

Tracking-by-detection [1, 6, 30] is the most popular category in multi-object tracking approaches. It consists of a detection module that localizes the objects of interest in a frame, and an association module that links these detected objects in time, maintaining their identity and producing object trajectories. The process can be performed online [29, 35, 14] by only using information gathered from past frames, or offline (batch mode) [15, 39, 28] by exploiting information from the whole video including past and future frames. Some applications (i.e. online video surveillance, navigation, autonomous driving etc.) require use of online approaches because of their real-time nature, others can use online or offline approaches.

Performance of the tracking-by-detection methods are greatly influenced by performance of the object detection methods they use. False positives (false detections), false negatives (undetected objects), under-segmentation (merging of neighboring objects), and over-segmentation (object fragmentation) are major sources of tracking errors. Recent advances in object detection [7, 13, 32] are thus important for the success of the tracking task.

Association links the detected objects in time. Tracking-by-detection frameworks formulate tracking as an object-to-track assignment problem. Local data association performs assignment considering information between adjacent frames [5, 3], whereas global data association takes multiple frames into account [38, 36]. Local assignment is more sensitive to detection errors, tends to produce short fragmented trajectories (tracklets), and may cause drift under occlusion [9, 37]. Association is done comparing object descriptors with suitable cost functions (similarity or dissimilarity measures). Features used to describe and compare objects can vary. Selection of discriminative features is very important in order to reduce association ambiguities. Features used in tracking should be able to discriminate different objects while being robust to factors such as illumination, viewpoint, pose etc. changes. Appearance similarity of the detected objects is a challenge for the association process since it may cause matching ambiguities. It is important to integrate additional information such as

978-1-5386-2939-0/17/$31.00 ©2017 IEEE

scene structures [22], target context [23, 27, 4], or target state prediction [16, 25, 17] to resolve association ambiguities. However, these additional processes adversely affect computational cost of tracking.

In this paper, we propose a robust detection-based multi-object tracking system that relies on an efficient but discriminative object description and a two-step combined local and global data association scheme. The main contributions of the paper are local assignment using spatial distance combined with a global assignment process integrating spatial, temporal, and appearance features to ensure time efficiency while preserving tracking accuracy; and an appearance model that combines shape and texture information using HoG descriptor with object color attributes with a novel color correlation cost matrix that ensures reliable color matching under changing illumination conditions. Experimental results show that using the proposed color matching within the tracklet linking step improves tracking performance. The paper is organized as follows. The next section will describe the tracking steps in term of tracking initialization, objects position prediction, local and global data association. Followed by experiment results and conclusion.

2. Multi-Object Tracking

Multi-object tracking (MOT) is the process of locating objects of interest in a video sequence, maintaining their identity over time, and producing set of trajectories corresponding to their motion. Our MOT system is a tracking-by-detection framework consisting of three stages: (1) detection validation and filtering, (2) local data association, and (3) global data association (Figure 1). The system uses the detection masks produced by the CompACT detector framework [7]. For each frame I_t of a given video sequence $V = \{I_1, I_2, ..., I_Q\}$ of length Q, there are set of $N(t)$ detected objects $D_t = \{d_{t,1}, d_{t,2}, ..., d_{t,N(t)}\}$ where $d_{t,i}$ represents object i in frame I_t. Each detected object $d_{t,i}$ is encoded with the vector $(d_{t,i}[x], d_{t,i}[y], d_{t,i}[w], d_{t,i}[h], s_{t,i})$, where the entries represent position, width, height, and detection score respectively. Tracking determines a set of associations $\mathbb{A} = \{\theta_2, \theta_3, ..., \theta_Q\}$, where each θ_t represents an assignment matrix from tracks to detections at frame I_t. *Detection filtering* step is applied to eliminate potential false detections produced by the detector. Objects with low detection scores are removed to reduce false positive score. *Local data association* is an online step that links the detected objects in the current frame to the existing tracks. Hungarian Algorithm [18] is used to resolve associations on a cost matrix relying on spatial distances between tracks and detections. *Global data association* is an offline tracklet to tracklet (rather than object to tracklet) linking step. Tracklets (short tracks) terminated early because of problems such as occlusions or weak detection scores, are linked

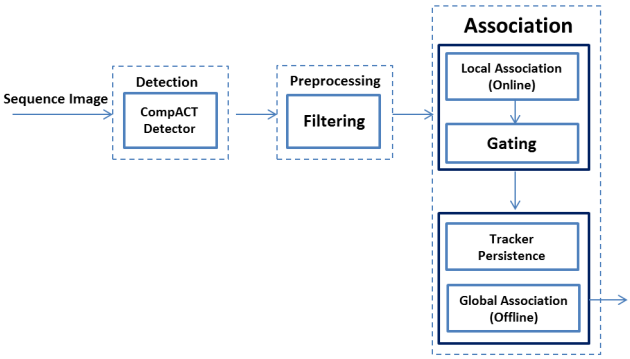

Figure 1. The framework of Multi-Object Tracking.

to tracklets formed after object recovery using object appearance cues and tracklet motion history. Appearance is described and compared using HoG descriptor [8] and CN (color name) histogram along with a novel weighted CN histogram dissimilarity measure. Below sections summarize main steps involved in the proposed tracking system.

2.1. Track Initialization

When a new track is formed, three groups of information corresponding to the new track are initialized: (1) detected object's location, width, height, and appearance; (2) counters for age, visible frames, and invisible frames; and (3) Kalman filter parameters $KF_i^t = \{x_i^t, P_i^t\}$ where x_i^t represents state estimate for the object i at frame t and P_i represents associated covariance matrix.

2.2. Track Position Prediction

Association is done between detected object positions D_t and predicted track positions T_t at frame t. Constant velocity Kalman filter is used to predict track positions. The process involves two steps: (1) *prediction step* that uses the current state estimate $\{x_i^{t-1}, P_i^{t-1}\}$ from $t-1$ to predict the estimate for time t; and (2) *update step* where the current prediction $\{x_i^t, P_i^t\}$ is combined with current observation (detection) information $\{d_i^t, R_i^t\}$ to refine the state estimate for future predictions, where d_i^t is the position and R_i^t is the observation covariance matrix of the detected object i .

2.3. Local Data Association using Spatial Distance

For the tracking-by-detection systems, the biggest challenge is the association of the noisy object detections $D^t = \{d_1, d_2, ..., d_N\}$ in a video frame with the previously tracked objects $T^{t-1} = \{T_1, T_2, ..., T_M\}$. Detected objects are assigned to existing tracks by minimizing a cost matrix using Munkres Hungarian algorithm [18]. Elements of the cost matrix are computed as:

$$C(i,j) = \log \left\| d_i(x,y), T_j(x,y) \right\|_2 \quad (1)$$

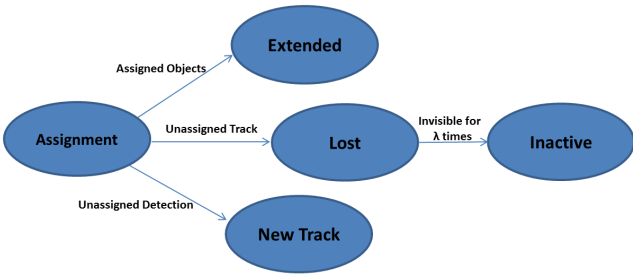

Figure 2. Decision State after assignment.

where $d_i(x, y)$ and $T_j(x, y)$ are the centroids of detected objects and predicted tracks respectively. We use circular gating around predicted track positions to eliminate highly unlikely associations, to reduce computational cost, and to reduce false matches. Minimization on the cost matrix results in the assignment $T \times 2$ matrix $\mathbb{A}_t$ containing the indices of the corresponding detection and track pairs. $\mathbb{A}_t$ determines track states for frame t. The four possible states $\{$*new track, extended track, lost track, inactive track*$\}$ are illustrated in Figure 2. State descriptions and associated parameter updates are summarized in Table 1.

2.4. Global Data Association using Spatial Distance and Appearance Similarity

Various problems during object detection and data association stages may cause tracks to terminate early resulting in short tracklets rather than full tracks. Global data association is used to link these tracklets to generate longer tracks. Global data association is an expensive process because it involves optimization of *all* possible matches not just the ones between consecutive frames. Reducing the number of objects that need to be associated is helpful in reducing this computational cost. In this paper, a refinement process relying on spatial distance, start and end frames, motion direction, and appearance model of the tracklets are used to filter out infeasible tracklet matches before the global assignment step. Given a tracklet K_i and the initial set of tracklets $J = \{K_1, K_2, ..., K_k\}$, the refinement process filters out the tracklets with the following properties from the set J as infeasible forward matches for K_i. Where forward linking refers to cases where tracklets are appended to the end of K_i.

1. All tracklets K_j that are initialized on frame borders or other source positions (tracklets entering the scene).

2. All tracklets K_j that are born before the death of tracklet K_i.

3. All tracklets K_j that start beyond the spatial distance T_s from the last position of tracklet K_i.

4. All tracklets K_j that start beyond the temporal distance T_t from the last frame of tracklet K_i.

5. All tracklets K_j whose appearance distances from tracklet K_i are above T_a. Appearance is described in terms of color using CN (color name) distribution and in terms of shape and texture using HoG descriptor.

Refinement process for backward linking where K_i gets inserted to the end of another tracklet is done in a similar way. Once all potential match sets are refined. Global data association determines the tracklet to tracklet associations by minimizing spatial and appearance distances. Details on our appearance model and appearance comparison scheme are given in Section 2.4.1.

2.4.1 Appearance Model for Tracklet Linking

It is important to have discriminant features to filter out infeasible detection-to-tracklet or tracklet-to-tracklet associations. Tracked objects' appearance models provide powerful information to refine the set of candidate matches. However, appearance is also sensitive to external factors such as illumination changes, shadows, partial occlusions, pose, viewing angles etc. It is important to develop appearance models that can discriminate different tracked objects, while being invariant against factors such as illumination, pose etc. In this work, we propose an appearance model that combines shape and texture information using HoG descriptor with object color attributes using our novel color correlation cost matrix. HoG [8] is a widely used powerful descriptor that describes shape and texture through histogram of gradient orientations in local image regions. HoG is particularly suitable for description of rigid objects such as cars in the UA-DETRAC dataset [33]. We record the HoG descriptor for all new tracks at the time of track initialization. Mean square error (MSE) is used to compute the distance between HoG descriptors.

We incorporate color information through an extension of the CN (Color Names) model proposed in [31]. Linguistic study described in [19] shows that English language has eleven basic color terms: black, blue, brown, gray, green, orange, pink, purple, red, white and yellow. [31] explores this concept to generate CN (Color Names) map. CN model associates RGB color values with linguistic color labels and reduces the 256^3 RGB color space to just 11^1 CN color space. Biggest challenge in color description is sensitivity to illumination. Illumination variations and shadows may alter color (and intensity) of an object making the information not reliable (*e.g.* shadow may change blue to black, or yellow to brown). When performing color appearance comparison, it is important to consider similarity of individual color codes and likelihood of colors switching from one value to another (*e.g.* while blue, yellow, and orange are

Table 1. State description and associated parameter updates.

State	Parameter Updates	Description
New Track	- Track ID - Kalman filter state $KF_i^t = \{x_i^t, P_i^t\}$ - Counters: age, visible frames, invisible frames - Appearance: CN color histogram & HoG descriptor	Detected objects not assigned to any existing tracks start new tracks.
Extended Track	- Kalman filter state $KF_i^t = \{x_i^t, P_i^t\}$ - Counters: age, visible frames	Detected objects successfully matched to existing tracks extend those tracks.
Lost Track	- Counters: age, invisible frames	Tracks not assigned to any detected objects are assigned to lost state.
Inactive Track	Save - Track ID - Full trajectory (previous and last seen positions) - Appearance: CN color histogram & HoG descriptor - Counters: age, born/death frames	After spending λ time steps in lost state, unmatched tracks are terminated.

three distinct color names, distance between blue and yellow is higher than distance between yellow and orange, and transition likelihood from blue to yellow is lower compared to transition probability from yellow to orange). In [20], O'Pele et.al. proposed a color distance weight matrix fusing CIEDE2000 color dissimilarity [26] with CN color pair probabilistic distance matrix. In this work, we have built and used an 11×11 CN-to-CN color correlation weight matrix $\mathcal{W}_{CN}$ to account for similarity between different CN values during color distribution comparison. The elements of the color correlation matrix are computed as follows:

$$\mathcal{W}_{CN}(c_i, c_j) = 1 - 2 \times \underset{k}{max}\big(\min \big(\mathcal{G}(k,i), \mathcal{G}(k,j)\big)\big) \quad (2)$$

where c_i and c_j are two CN codes and $\mathcal{G}$ is the $2^{15} \times 11$ matrix describing the mapping from $32 \times 32 \times 32$ quantized RGB space to 11-valued CN space provided by [31]. In order to compare object color distributions, we use Earth mover distance (EMD) [24]. EMD computes the minimum paid cost to transfer one histogram to another histogram. We use color correlation matrix $\mathcal{W}_{CN}$ as transfer cost. See Figure 3 for more details. Use of cross-bin color histogram distance computation using EMD scheme and color correlation matrix $\mathcal{W}$ described above decreases sensitivity to color changes caused by illumination variations and improves tracking results compared to bin-to-bin color histogram comparison.

3. Experimental Results

We have tested and evaluated our MOT tracker on UA-DETRAC-test multi-object tracking benchmark dataset [33] consisting of 40 challenging real-world traffic videos. We have evaluated the performance of our tracker using the UA-DETRAC evaluation toolkit. The evaluation process includes the following metrics described in [33]: mostly track (**PR-MT**), mostly lost (**PR-ML**), identity switches (**PR-IDS**), track fragmentation (**PR-FRAG**), false positives (**PR-FP**), false negatives (**PR-FN**), multi-object tracking precision (**PR-MOTP**), and multi-object tracking accuracy (**PR-MOTA**). Accuracy of the obtained tracks are best reflected by combination of the metrics **PR-IDS** and **PR-FRAG**. Final evaluation measures are computed by applying the trackers on a set of object detection results obtained by varying detection threshold, corresponding to different detection precision and recall levels.

We have combined our tracker on four different state-of-the-art object detectors, CompACT [7], R-CNN [13], ACF [11], and DPM [12]; and compared our tracking results to six state-of-the-art MOT trackers, including GOG [21], CEM [2], DCT [3], IHTLS [10], H^2T [34], and CMOT [5]. Table 2 summarizes the tracking performances of our tracker and other state-of-the-art trackers (described in [33]). Our tracker outperforms the other trackers in term of **PR-MOTA**, **PR-MOTP**, **PR-MT**, **PR-ML**, and **PR-FN**; produces comparable results for **PR-FRAG**, **PR-IDS**, and results in second highest score in terms of **PR-MOTA**, **PR-ML**, and **PR-FP**.

We have analyzed the contribution of different components to the overall performance of our tracker by systematically enabling/disabling different components (Figure 4). Our tests show that all components contribute to better accumulated performance. For example, in $Experiment-E$, disabling the CN-to-CN color correlation weight matrix $\mathcal{W}_{CN}$ and assuming uncorrelated color codes, results in a drop in PR-MOTA metric from 13.1 to 12.7. Figure 5 shows performance improvement in the tracker by addition of each component.

Table 3 summarizes frame rates (in terms of frames per seconds) for our tracker and other state-of-the-art trackers provided in [33]. Frame rate for our tracker has been obtained averaging frame rates on 40 sequences of UA-DETRAC-test set. During the tests, CompACT object de-

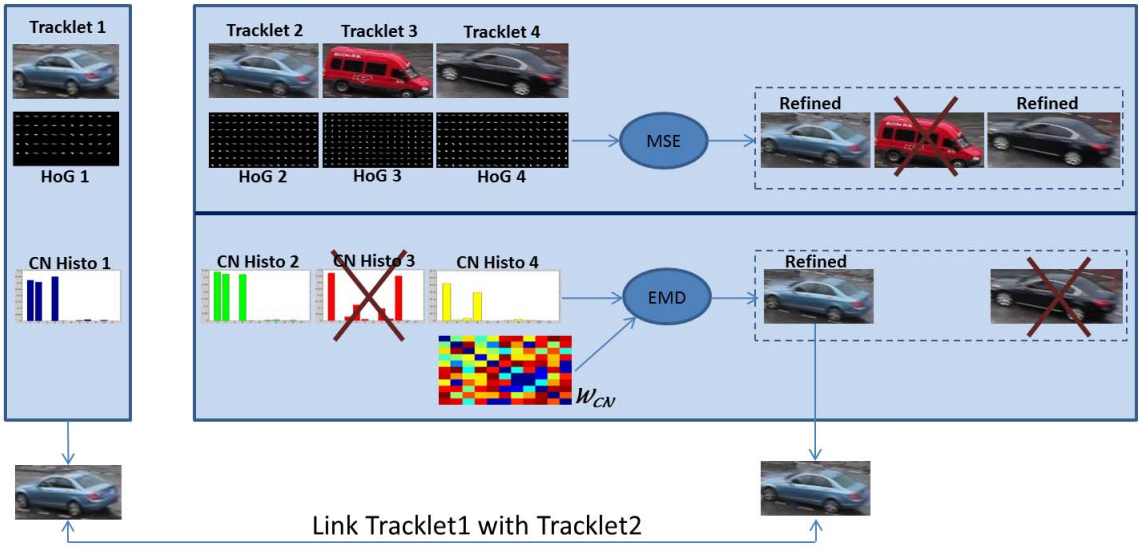

Figure 3. Apperance model refinment for global assignment. Tracklet_1 need to be assigned to one of the candidates ($\text{Tracklet}_{2,3,4}$) that born after Tracklet_1. In the first step, Tracklet_3 has been excluded by HoG constrain. Then, Tracklet_4 has been excluded by color constrain. Finally, After refinement Tracklet_1 is linked with Tracklet_2

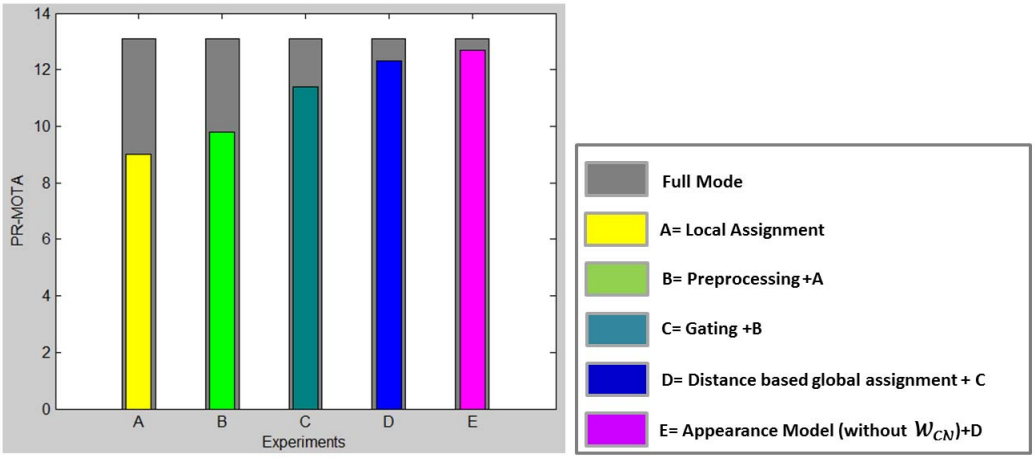

Figure 4. Contribution of each tracker component. Left: PR-MOTA metric for each experiment compared to the Full Mode (higher is better). Right: description of each experiment.

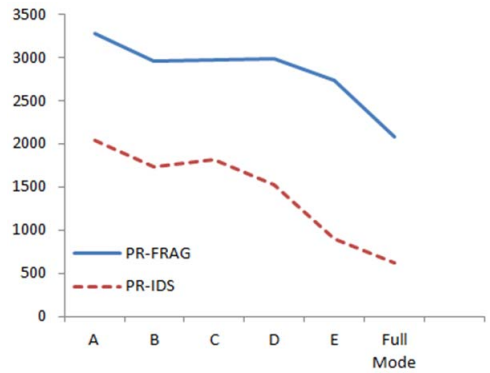

Figure 5. Contribution of each tracker component on PR-IDS and PR-FRAG metrics (lower is better).

tection results have been used as input. Our simple feature set combined with two-step distance-only local and distance appearance combined global data association scheme allows high frame rates while still preserving reliability and accuracy of our tracker.

4. Conclusions

In this paper, we proposed a detection-based multi-object tracking system that uses a two-step data association scheme to ensure time efficiency while preserving tracking accuracy. Low cost local data association operates at object level on consecutive frames relying on only spatial distance, while higher cost global data association operates on track-

Table 2. Tracker performance comparison using four state-of-the-art object detectors as input. Last raw indicates whether the metric for the specific column is better when high or low, + and − respectively. Best performance for each metric are marked in bold. We have two entries for our tracker based on how the performances scores are averaged. Challenge reports results in two groups corresponding to Easy versus (Medium + Hard) videos. Proposed-A computes average score as $(Score_{Easy} + Score_{Med+Hard})/2$. Proposed-B computes average score as $(0.25 \times Score_{Easy} + 0.75 \times Score_{Med+Hard})$ based on number of video sequences in each group. Scores are computed and averaged for all four object detectors according to [33].

Trackers	PR-MOTA	PR-MOTAP	PR-MT	PR-ML	PR-IDS	PR-FRAG	PR-FP	PR-FN
GOG [21]	10.1	35.3	10.9	22.5	4248.5	4137.3	43657.6	199926.8
CMOT [5]	7.0	34.6	12.8	**21.2**	414.3	1817.5	79577.3	187508.8
H^{2}T [34]	7.8	34.6	11.2	22.2	1298.9	1477.4	65275.3	196107.0
DCT [3]	8.3	35.6	5.5	32.3	**270.1**	**261.4**	**17658.2**	241590.8
IHTLS [10]	5.8	35.1	9.6	23.4	1329.2	4597.1	68635.0	205649.3
CEM [2]	3.9	33.6	2.4	36.1	394.3	529.0	19044.8	267699
SCTrack(Ours)-A	**12.8**	**37.7**	**12.9**	21.9	494.9	1408.0	27581.0	**99642.3**
SCTrack(Ours)-B	9.8	35.3	10.96875	22.637	656.48	1408.08	35945.59	130496.6
Better	**+**	**+**	**+**	**-**	**-**	**-**	**-**	**-**

let level integrating spatial, temporal, and appearance features. Our proposed tracker also includes a robust but discriminative object appearance model that combines shape and texture information using HoG descriptor with object color attributes using our novel color correlation cost matrix. Experiments on UA-DETRAC dataset shows promising results against state-of-the-art trackers. We are in the process of extending our appearance model for further improved results.

References

[1] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[2] A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1265–1272, 2011.

[3] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1926–1933, 2012.

[4] N. T. L. Anh, F. Bremond, and J. Trojanova. Multi-object tracking of pedestrian driven by context. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016.

Table 3. Frame rate comparison.

Trackers	Code	Frame Rate (fps)
CEM	Matlab	4.62
GOG	Matlab	**389.51**
DCT	Matlab, C++	2.19
IHTL	Matlab	19.79
H2T	C++	3.02
CMOT	Matlab	3.79
SCTrack(Ours)	Matlab	362.00

[5] S.-H. Bae and K.-J. Yoon. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1218–1225, 2014.

[6] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1820–1833, 2011.

[7] Z. Cai, M. Saberian, and N. Vasconcelos. Learning complexity-aware cascades for deep pedestrian detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3361–3369, 2015.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005.

[9] C. R. del Blanco, F. Jaureguizar, and N. Garcia. An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications. *IEEE Transactions on Consumer Electronics*, 58(3), 2012.

[10] C. Dicle, O. I. Camps, and M. Sznaier. The way they move: Tracking multiple targets with similar appearance. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2304–2311, 2013.

[11] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545, 2014.

[12] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Ttransactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on*

computer vision and pattern recognition, pages 580–587, 2014.

[14] H. Kieritz, S. Becker, W. Hübner, and M. Arens. Online multi-person tracking using integral channel features. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 122–130, 2016.

[15] C.-H. Kuo and R. Nevatia. How does person identity recognition help multi-person tracking? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1217–1224, 2011.

[16] 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 *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages III–III, 2002.

[17] D. Mikami, K. Otsuka, and J. Yamato. Memory-based particle filter for face pose tracking robust under complex dynamics. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 999–1006, 2009.

[18] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.

[19] F. Panoff-Eliet, B. Berlin, and P. Kay. Basic color terms. their universality and evolution, 1971.

[20] O. Pele and M. Werman. Improving perceptual color difference using basic color terms. *arXiv preprint arXiv:1211.5556*, 2012.

[21] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1201–1208, 2011.

[22] J. Prokaj and G. Medioni. Using 3d scene structure to improve tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1337–1344, 2011.

[23] V. Reilly, H. Idrees, and M. Shah. Detection and tracking of large number of targets in wide area surveillance. *Computer Vision–ECCV 2010*, pages 186–199, 2010.

[24] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[25] S. Shantaiya, K. Verma, and K. Mehta. Multiple object tracking using kalman filter and optical flow. *European Journal of Advances in Engineering and Technology*, 2(2):34–39, 2015.

[26] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005.

[27] X. Shi, P. Li, H. Ling, W. Hu, and E. Blasch. Using maximum consistency context for multiple target association in wide area traffic scenes. In *IEEE International Conference on Speech and Signal Processing (ICASSP)*, pages 2188–2192, 2013.

[28] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah. Part-based multiple-person tracking with partial occlusion handling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1815–1821, 2012.

[29] F. Solera, S. Calderara, and R. Cucchiara. Learning to divide and conquer for online multi-target tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4373–4381, 2015.

[30] M. Ullah, F. A. Cheikh, and A. S. Imran. Hog based real-time multi-target tracking in bayesian framework. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 416–422, 2016.

[31] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1523, 2009.

[32] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan. Static and moving object detection using flux tensor with split gaussian models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 414–418, 2014.

[33] L. Wen, D. Du, Z. Cai, Z. Lei, M. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *arXiv CoRR*, abs/1511.04136, 2015.

[34] L. Wen, W. Li, J. Yan, Z. Lei, D. Yi, and S. Z. Li. Multiple target tracking based on undirected hierarchical relation hypergraph. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1282–1289, 2014.

[35] B. Wu and R. Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International Journal of Computer Vision*, 75(2):247–266, 2007.

[36] J. Xing, H. Ai, and S. Lao. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In *IEEE Conferefnce on Computer Vision and Pattern Recognition (CVPR)*, pages 1200–1207, 2009.

[37] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1345–1352, 2011.

[38] B. Yang and R. Nevatia. Multi-target tracking by online learning a crf model of appearance and motion patterns. *International Journal of Computer Vision*, 107(2):203–217, 2014.

[39] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.