Soft Biometrics Integrated Multi-target Tracking

Xiaojing Chen
Center for Research in Intelligent Systems
University of California, Riverside
Riverside, California 92521, USA
Email: xchen010@ucr.edu

Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside
Riverside, California 92521, USA
Email: bhanu@cris.ucr.edu

Abstract—In this paper, we present a soft biometrics based appearance model for multi-target tracking in a single camera. Tracklets, the short-term tracking results, are generated by linking detections in consecutive frames based on conservative constraints. Our goal is to "re-stitching" the adjacent tracklets that contain the same target so that robust long-term tracking results can be achieved. As the appearance of the same target may change greatly due to heavy occlusion, pose variations and changing lighting conditions, a discriminative appearance model is crucial for association-based tracking. Unlike most previous methods which simply use the similarity of color histograms or other low level features to construct the appearance model, we propose to use the fusion of soft biometrics generated from sub-tracklets to learn a discriminative appearance model in an *online* manner. Compared to low level features, soft biometrics are robust against appearance variation. The experimental results demonstrate that our method is robust and greatly improves the tracking performance over the state-of-the-art method.

I. INTRODUCTION

Multi-target tracking in real-world scenes has been extensively studied for many years, as it is an important prerequisite for many industrial applications, such as video surveillance, human-computer interaction, behavior analysis, and anomaly detection. The goal of multi-target tracking is to provide the trajectories of all targets while maintaining the consistency of identity labels. There are many challenges in multi-target tracking, such as illumination and appearance variation, occlusion, and sudden change in motion [1][2].

As great progress has been achieved in object detection, association-based tracking has become more popular recently [3][4]. In the association-based tracking framework, an object detector is first applied to all image frames, and conservative constraints are used to associate consecutive detection responses into tracklets. Thus, each tracklet is a sequence of short but reliable estimations of the same target. Longer tracks of different targets are generated by associating tracklets under certain spatial-temporal constraints according to their affinities. In most association-based tracking approaches, an affinity model is constructed which combines time, appearance, and motion information to measure the affinity between two tracklets. Many previous methods use color histograms only to represent the appearance of a target [5][6][7]. However, the color information is not always consistent and reliable, as the appearance of a target may change drastically even within a short time period due to occlusion, pose and illumination variations, as shown in Fig. 1.

In this paper, to enhance the appearance model for tracking, in addition to color histogram, we use soft biometrics which

Fig. 1. Examples of drastic appearance variation within a short period of time. Bounding boxes with the same color indicates the same target. Images in the first row, with two seconds time gap, illustrate appearance change caused by occlusion and pose variation; images in the second row, with one second time gap, show the impact of illumination variation on appearance.

are invariant to view and illumination changes to learn a discriminative appearance model in an online manner. Soft biometrics are characteristics that can be used to describe a person [8], for instance, height, weight, gender, hair color and clothes color. Although each one of them is not discriminative enough to uniquely identify an individual, when bundled as a whole they can provide a coarse representation of a target. Because soft biometrics can be directly acquired from surveillance videos without any subject's cooperation, they are suitable for constructing appearance models of tracked targets. Soft biometrics have been widely used for retrieval and recognition tasks on image datasets [8][9][10], target identification in surveillance video data [11], and person re-identification across cameras [12]. However, to the best of the authors' knowledge, soft biometrics have not been used to improve tracking performance.

We generate the appearance model that can differentiate the tracklets containing different targets using AdaBoost [13]. With the online learning process, the generated appearance model is specifically designed for tracklets within current time interval, thus good discriminality can be achieved. For each frame, a set of soft biometric features is extracted from every detection response. As the appearance of target in a tracklet
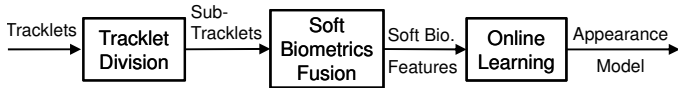
Fig. 2. An overview of the proposed soft biometrics based appearance model.

may vary even after a short time period because of illumination change and AdaBoost is sensitive to outliers and noisy data, we divide each tracklet into several sub-tracklets and apply a fusion method to aggregate a number of soft biometric feature sets into a single one. It is demonstrated by the experiments that the fusion of soft biometrics provides a concise and reliable representation for tracked targets and the use of sub-tracklets greatly improves the efficiency and effectiveness of the proposed model. Besides the appearance constraint, we also use spatial and temporal constraints to deal with ambiguities in tracklet association.

An overview of our appearance model is shown in Fig. 2. The contributions of this paper are the following:

- We integrate soft biometrics information into appearance model to mitigate the confusion caused by appearance variation.

- Tracklets are divided into sub-tracklets based on time constraint and appearance similarity. The use of sub-tracklet instead of detection responses greatly reduces online training time, and alleviates the negative impact of incorrectly extracted soft biometrics.

- Real-world surveillance data are used in the experiments and the effectiveness of the proposed method is validated.

The rest of the paper is organized as follows: Section II discusses related work; an outline of our tracking method and our proposed soft biometrics based appearance model is described in Section III; experiments are presented in Section IV; and Section V concludes this paper.

II. RELATED WORK

Multi-target tracking has been extensively explored in the past. A general categorization divides multi-target tracking approaches into two groups. In the first group, tracking is considered in a time critical scenario in which only the past and current frames are used for tracking [14][15][16]. On the other hand, less time critical applications take future frames also into consideration [3][4][17][18][19][20].

For time critical tracking, usually particle filters are used. A multi-target tracking framework is proposed in [14] which uses both continuous pedestrian detector confidence and online instance-specific classifier. Hue et al. [15] extend the classical particle filter in order to deal with multiple targets and multiple receivers in the regime of sequential monte carlo methods. Recently, an efficient particle filter is proposed using an independence approximation for multi-target tracking [16]. However, the particle filtering based methods are not robust against identity switches caused by occlusion, abrupt motion, appearance variation, etc.

For less time critical tracking, *data association based tracking* (DAT) has become a major researched area recently. With the detection responses ready in hand, the focus is shifted to robust tracklet association using global optimization [3][4]. To achieve robust association, reliable affinity scores between tracklets are essential. Such scores are generally extracted from appearance information and motion information. To enhance the appearance model, multiple features including color histogram, HOG features, covariance matrices are used in a boosting framework [20]. To mitigate occlusions, part-based appearance models have been applied in multi-target tracking [17][18]. Non-linear motion is proposed in [19] to better model direction changes for robust motion affinity. As a drawback, the appearance or motion affinity is prone to error since these low-level features are not invariant against complex and noisy scenes.

Compared to low-level features such as color histogram which may drift significantly over time, soft biometrics, such as gender, height, are rather stable with respect to changes in appearance, time, and motion. Soft biometrics contain high level semantic information, which has been used for recognition or retrieval tasks [9][11][12]. In [9], facial marks such as freckles and scars are used for improved face recognition. Reid et al. [11] use soft biometrics to retrieve specified subjects in surveillance video data. Most recently, soft biometrics have been used to improve the person re-identification accuracy across non-overlapping surveillance cameras [12].

III. TECHNICAL APPROACH

A. Tracking Framework

Given a set of tracklets, the goal of association-based tracking is to find the best set of tracklet associations with the maximum linking probability. The global optimal solution is often obtained by solving the maximum *a posteriori* problem (MAP) using various optimization algorithms. As linking all the tracklets in an entire video directly has a high computational cost, it is a common routine to divide the video using a time sliding window, and carry out tracklet association independently in each time sliding window.

We define a set of tracklets generated from a time interval $[0, T]$ as $T = \{T_1, ..., T_n\}$. Each tracklet is a consecutive sequence of detection responses that contain the same target. Let association ψ_{ij} define the hypothesis that tracklet T_i and T_j contain the same target, with T_i occurring before T_j. A valid association matrix ψ is defined as follows:

$$\psi = \{\psi_{ij}\}, \psi_{ij} = \begin{cases} 1 & \text{if } T_i \text{ is associated to } T_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$
$$\text{s.t.} \sum_i \psi_{ij} = 1 \text{ and } \sum_j \psi_{ij} = 1$$

The two constraints for matrix ψ indicates that each tracklet can only be associated to and associated by one other tracklet. To handle the case of initial and terminating tracklet, an augmented matrix similar to [21] is used.

We define A as the pairwise transition score matrix, and A_{ij} as the cost for associating T_i and T_j based on the affinity
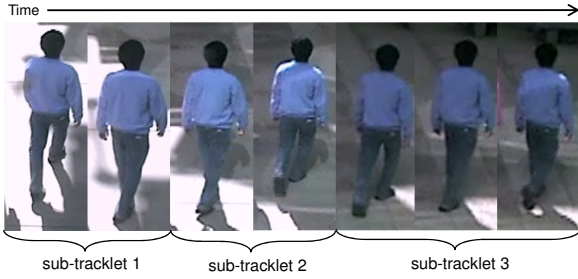
Fig. 3. An example of tracklet division (only detections on key frames are shown). Detections in the same sub-tracklet have higher appearance similarities when compared to detections in other sub-tracklets.

Name	Value	Type
Gender	Male, Female	Symbolic
HairColor	Light, Dark	Symbolic
SkinColor	Caucasian, Non_Caucasian	Symbolic
Height	Centimeters	Scalar
Weight	Kilograms	Scalar
BodyColor	1-D probability distribution	Vector
TorsoColor	1-D probability distribution	Vector
LegsColor	1-D probability distribution	Vector

TABLE I.   SOFT BIOMETRICS EXTRACTED FROM DETECTION RESPONSES.

model. Let Ψ be the set of all possible association matrices, we formulate the multi-target tracking as the following optimization problem:

$$\psi^* = \arg\min_{\psi \in \Psi} \sum_{ij} A_{ij} \psi_{ij} \qquad (2)$$

The Hungarian algorithm is used to solve this assignment problem in polynomial time.

The affinity model is constructed by integrating multiple cues, as show in Eq. (3):

$$A_{ij} = Appr(T_i, T_j) + Time(T_i, T_j) + \gamma Motion(T_i, T_j) \quad (3)$$

where T_i, T_j are two different tracklets, γ is a weighting parameter, and $Appr(\cdot)$, $Time(\cdot)$, and $Motion(\cdot)$ are the appearance, time and motion models respectively. In appearance and motion models, the cost for associating T_i and T_j is defined as their negative loglikelihood. The time model is defined as the following:

$$Time(T_i, T_j) = \begin{cases} 0 & \text{if } 0 < Gap_{ij} < GAP \\ \infty & \text{otherwise} \end{cases} \qquad (4)$$

where Gap_{ij} is the frame difference between T_i and T_j, and only when Gap_{ij} is within a pre-defined maximum allowed gap, GAP, the two tracklets can be linked. A linear motion model which is widely used for multi-target tracking [6][21] is employed in our approach.

B. The Discriminative Appearance Model

1) Tracklet division: In a tracklet, the appearance of a target may vary with time (see Fig. 3), but the detection responses that are obtained in consecutive frames often possess high visual similarity. For efficient computation and to create concise representations of a tracklet, we further divide each tracklet into several sub-tracklets and consider every sub-tracklet as an appearance instance of a target. An example of tracklet division is shown in Fig. 3

We assume that a target cannot have large pose variation in a very short period of time Δt (0.5s in our experiments). Tracklet division starts from the beginning of a tracklet and

sub-tracklets are generated one by one. Let len be the number of frames included in Δt (about 10 in our experiments). The first detection of a tracklet is the appearance reference to form the current sub-tracklet. Specifically, with respect to the detection in the first frame as the reference detection, the detections from the following frames within Δt are compared to the reference detection. As long as the detection similarities are above a pre-defined threshold (0.9 in our experiments), the corresponding frames are kept in this sub-tracklet. Thus, the number of detections in a sub-tracklet is smaller or equal to len. Once a detection's similarity to the reference detection is below the threshold or the number of detections in the current sub-tracklet exceeds len, this detection becomes a new reference detection and detections in latter frames are compared to this reference detection to form a new sub-tracklet. Here color histogram is used to measure the detection similarity.

2) Fusion of soft biometrics: The soft biometric features extracted from a detection response are shown in Table I, the potential values for each feature are also listed. These soft biometric features are extracted using state-of-the-art techniques [8][22], and can be categorized into three types: symbolic, scalar-valued, and vector-valued. A confidence level which scales from 0 to 1 is associated with each feature to indicate the prediction confidence. Our goal is to design a fusion method that can combine common soft biometric features extracted from several detections into a single one. In the remainder of this paper, fn represents the feature name, $fval$ is the feature value, and fc is the confidence level.

For binary symbolic features, the sum of confidence levels of all potential values is equal to 1, thus, given the confidence level of one potential value, the confidence level for the other potential value can be inferred. When fusing symbolic features, the averaged confidence level for each potential value is computed and the one with the highest score is selected as the fused confidence level, and the corresponding value is the fused feature value. For scalar-valued and vector-valued features, the fused value and fused confidence level are defined as follows:

$$fval_{\_fused} = \frac{1}{\sum fc_i} \sum_{i=1}^{n} fval_i \times fc_i \qquad (5)$$

$$fc_{\_fused} = \frac{1}{n} \sum_{i=1}^{n} fc_i \qquad (6)$$

where n is the number of features used for fusion.

3) Similarity of soft biometric features: After soft biometrics fusion, each sub-tracklet is represented by a single set of soft biometric features. Then the similarity between common features for each feature type (symbolic, scalar, vector) is computed. For the symbolic features (Gender, HairColor, SkinColor), if the symbolic value of the two features are the same then the similarity is the average of the two confidence levels. If the symbolic values are dissimilar, then the similarity is the maximum confidence level, as defined in Eq. (7).

$$sim_1(fval_1, fval_2) = max(fc_1 \times (1 - fc_2), (1 - fc_1) \times fc_2) \tag{7}$$

For the scalar-valued features (Height and Weight), we assume that the feature values are from a normal distribution with parameters μ and σ^2. As the height accuracy is $\pm 12.7cm$ and the weight accuracy is $\pm 9kg$ (learned by analyzing soft biometrics extracted from previous data), we define the standard deviation so that for the height $P([fval-12.7, fval+12.7]) = 80\%$ and for the weight $P([fval-9, fval+9]) = 80\%$. For the accumulated probability to be equal to 80% the range should be $(\mu - 1.28\sigma, \mu + 1.28\sigma)$. Thus, the standard deviations are equal to $1.28\sigma = 12.7$ for height and $1.28\sigma = 9$ for weight. The similarity score sim_2 is defined as:

$$sim_2(fval_1, fval_2) = 1 - \sqrt{1 - e^{\frac{-(fval_1 - fval_2)^2}{8\sigma^2}}} \tag{8}$$

For the vector-valued features (BodyColor, TorsoColor, LegsColor), the Bhattacharyya Coefficient is used to measure the similarity sim_3, which approximates the amount of overlap between two probability distributions, as given in Eq. (9):

$$sim_3(fval_1, fval_2) = \sum_{i=1}^{n} \sqrt{fval_{1i} \times fval_{2i}} \tag{9}$$

4) Learning of the appearance model: Our goal is to design a discriminative appearance model that gives high similarity for a pair of tracklets that contain the same target (positive pairs) while assigning low similarity for two tracklets that contain different targets (negative pairs). We use AdaBoost to learn the appearance model [20]. AdaBoost consists of a number of weak classifiers and adaptively learns a strong classifier that is a linear combination of all weak classifiers and minimizes the overall error. In our appearance model, the similarity computed from each soft biometric feature is a weak classifier, and AdaBoost assigns a weighting parameter for each feature during the online learning process. We formulate the learned appearance model as follows:

$$H(T_i, T_j) = \sum_{t=1}^{T} \omega_t h_t(T_i, T_j) \tag{10}$$

where t indicates the iteration index, ω_t is the weighting parameter and $h_t(T_i, T_j)$ is a weak classifier based on one of the soft biometric features extracted from tracklets T_i and T_j.

As the appearance of the same target may vary over time, each tracklet is represented by the average of all its sub-tracklets. The appearance affinity for a pair of tracklets T_i and T_j is measured according to their corresponding averaged features. In case the training of AdaBoost cannot be carried out because no negative tracklet pairs are found or the training set size is too small, we directly compute the pair-wise similarities of all tracklets using the soft biometric feature similarities. The similarity of two sets of soft biometrics is defined as the weighted sum of all feature similarities. The weights are assigned according to confidence levels so that similarities calculated from feature values with less confidence levels have less effect, and the summation of all weights equals to 1.

5) Online training sample collection: In order to collect positive and negative training samples online, we adopt the selection criteria similar to [20]. Note that our training set is constructed on sub-tracklets level, while in [20] the detection responses are used directly.

Two spatial-temporal constraints are used to identify negative tracklet pairs (two tracklets that contain different targets). Two tracklets form a negative pair, if one of the following constraints is satisfied:

- the two tracklets have large overlap in time;
- the two tracklets have a small frame gap but large spatial difference.

The first constraint is based on the observation that the same target cannot be simultaneously tracked in two different tracklets; and the second constraint is based on the observation that one target cannot make a significant distance move within a short time period with limited velocity.

A pair of sub-tracklets (t_x, t_y) is a positive sample if $t_x, t_y \in T_i$, and $t_x \neq t_y$, and it is a negative sample when $t_x \in T_i$, $t_y \in T_j$ and (T_i, T_j) is a negative tracklet pair. The value of a training sample is an 8-dimensional vector, where each element of the vector is the similarity between two common soft biometric features in the two sub-tracklets.

IV. EXPERIMENTAL RESULTS

In order to evaluate the proposed model on a dataset that contains many different kinds of appearance variations, eight real-world videos of different scenes (five indoor and three outdoor) are collected in a real-world surveillance camera network. Each video is about 9 minutes at the frame rate of 20fps with a resolution of 704×480. Sample frames are shown in Fig. 4. The number of participants involved in each video ranges from 9 to 24. We plan to make this data set publicly available in the future.

This dataset is very challenging for the following reasons: (1) it covers multiple different scenes (both indoor and outdoor), while most public tracking datasets only have a single scene or multiple scenes that are quite similar; (2) there are intense illumination variations for the outdoor videos (see Fig. 4), which may cause an abrupt and drastic change to the appearance of a target. (3) even for the indoor videos there still exist noticeable illumination changes, for instance, in video 3 the girl in purple is walking from a bright region to a dark region. Therefore, traditional appearance models that

Fig. 4. Sample frames from each video (the first five videos are indoor and the last three videos are outdoor). Obvious illumination variation exists for both indoor and outdoor videos. Best viewed in color.
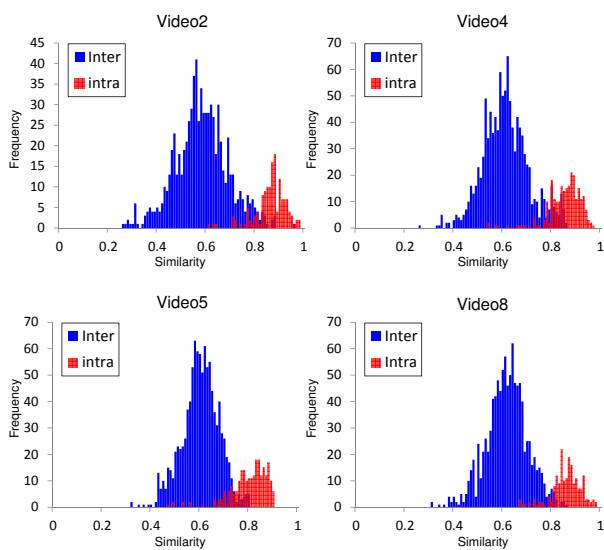

Fig. 5. Sample histograms of inter-class and intra-class tracklets similarities from four testing videos. Best viewed in color.

rely mainly on color information are much less reliable on this dataset.

A. Soft biometrics verification

As we use soft biometrics to build the appearance model, the quality of soft biometrics and the soft biometrics similarity measurement are crucial for the performance of tracklet association. To validate the robustness of the soft biometrics, for each video we compute the similarity between all pairs of tracklets based on the proposed soft biometrics fusion method, and these similarities are categorized into intra-class (tracklets from the same target) and inter-class (tracklets from different targets). The histograms for each category are plotted. Four sample histograms are shown in Fig. 5. The histograms suggest that most intra-class similarities are larger than most inter-class similarities and with a single threshold these two classes can be coarsely separated. Therefore, the soft biometrics extracted from the same target have high degree of consistency.

B. Tracklet association results

In this evaluation, the main goal is to reduce the number of fragments. The proposed model uses soft biometrics as features and generates training samples using sub-tracklet pairs. Two baseline models are presented for comparison. In **Baseline Model I**, only soft biometrics are used without sub-tracklet. Specifically, training samples are collected directly from detection pairs. In **Baseline Model II**, sub-tracklets are used but no soft biometrics are used except color histograms. We also compare to the appearance model in [20] which uses detection pairs and color histograms are adopted as extracted features.

The experimental results for all the videos are shown in Table II. Two sets of parameters are used, one is applied to all indoor videos and the other is applied to all outdoor videos. It is obvious that the association time is reduced roughly by one order of magnitude for models that use sub-tracklets, as the size of training set for AdaBoost is much smaller. The reduction rate of Baseline Model II is comparable to the proposed model, however, the number of association errors for each video is greater than the proposed model. This demonstrates that the other soft biometric features (Gender, HairColor, SkinColor, Height, Weight) are complementary to color histograms, as they capture the appearance information that is overlooked by color histograms. When using Baseline Model I, although the number of errors is small, fewer fragments are detected compared to the proposed model. Because detection pairs are directly used to generate the training set, two detections from the same tracklet but with quite different appearance (*e.g.*, the appearance of targets in outdoor videos is greatly affected by lighting) are also collected as positive samples. This would provide bad training data for AdaBoost and makes the learned appearance model less discriminative.

1) Comparison with state-of-the-art: The model in [20] also uses detection pairs for training set and only color histograms are used as extracted features. Comparison between Baseline model I and the model in [20] indicates that fusion of soft biometrics in a sub-tracklet is essential, as for a single detection the other soft biometric features (Gender, HairColor, SkinColor, Height, Weight) are more prone to errors than color histograms. Based on the results, the proposed model still

	FB	The Proposed Model (SoftBio.+sub-tracklet)				Baseline Model I (SoftBio.+detection)				Baseline Model II (Color+sub-tracklet)				Model in [20] (Color+detection)			
		FA	AE	RR	Time	FA	AE	RR	Time	FA	AE	RR	Time	FA	AE	RR	Time
Video1	9	2	0	78%	77s	4	1	56%	607s	2	3	78%	86s	2	1	78%	595s
Video2	6	1	0	83%	73s	1	0	83%	514s	1	2	83%	69s	1	0	83%	517s
Video3	3	0	0	100%	34s	1	0	67%	187s	1	1	67%	35s	1	0	67%	187s
Video4	12	2	1	83%	49s	4	1	67%	260s	2	2	83%	44s	3	2	75%	261s
Video5	3	1	0	67%	90s	2	1	33%	560s	1	0	67%	77s	1	0	67%	559s
Video6	11	2	1	82%	99s	5	1	55%	774s	3	5	73%	84s	3	2	73%	760s
Video7	0	0	0	-	8s	0	0	-	70s	0	1	-	8s	0	1	-	70s
Video8	23	7	1	70%	116s	10	1	57%	821s	9	8	61%	112s	9	2	61%	818s

TABLE II.   TRACKLET ASSOCIATION RESULTS. "FB" AND "FA" ARE THE NUMBER OF FRAGMENTS BEFORE AND AFTER TRACKLET ASSOCIATION, "AE" IS THE NUMBER OF ASSOCIATION ERRORS, "RR" IS THE FRAGMENT REDUCED RATE AFTER TRACKLET ASSOCIATION, "TIME" IS THE ASSOCIATION TIME FOR A CORRESPONDING VIDEO. DURATION OF EACH VIDEO IS ˜9MIN.

outperforms the model in [20] with a higher fragment reduction rate and less association errors, and the running time is much less due to the employment of sub-tracklets. Also, it is worth noting that when no fragment exists in the input tracklets (*e.g.*, Video 7), our model does not mistakenly associate any tracklet pairs, which further demonstrates the robustness of our model.

V. CONCLUSIONS

In this paper, we proposed a single camera multi-target tracking method by integrating soft biometrics into the appearance model. Compared to previous methods that only use low level features such as color histogram for data association, soft biometrics, as the high level semantic information, are more invariant against changes in pose, illumination, etc. Fusion of soft biometrics alleviates the negative impact of incorrectly extracted soft biometrics features and by using sub-tracklets the computational cost is significantly reduced. Compared to the baseline models and state-of-the-art method [20], the soft biometrics integrated tracking model has fewer fragments, higher fragment reduction rates, and significantly lower computation time.

REFERENCES

[1] B. Song, T.-Y. Jeng, E. Staudt, and A. K. Roy-Chowdhury, "A stochastic graph evolution framework for robust multi-target tracking," in *Proceedings of the 11th European Conference on Computer Vision (ECCV): Part I*, 2010, pp. 605–619.

[2] H. Liu and F. Sun, "Visual tracking using sparsity induced similarity," in *20th International Conference on Pattern Recognition (ICPR)*, 2010, pp. 1702–1705.

[3] J. Berclaz, F. Fleuret, and P. Fua, "Robust people tracking with global trajectory optimization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2006, pp. 744–750.

[4] Z. Wu, T. Kunz, and M. Betke, "Efficient track linking methods for track graphs using network-flow and set-cover techniques," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1185–1192.

[5] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1822–1829.

[6] C. Huang, B. Wu, and R. Nevatia, "Robust object tracking by hierarchical association of detection responses," in *Proceedings of the 10th European Conference on Computer Vision (ECCV): Part II*, 2008, pp. 788–801.

[7] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1200–1207.

[8] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739–777, 2011.

[9] A. Jain and U. Park, "Facial marks: Soft biometric for face recognition," in *16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 37–40.

[10] G. Guo, G. Mu, and K. Ricanek, "Cross-age face recognition on a very large database: The performance versus age intervals and improvement using soft biometric traits," in *20th International Conference on Pattern Recognition (ICPR)*, 2010, pp. 3392–3395.

[11] D. A. Reid and M. Nixon, "Using comparative human descriptions for soft biometrics," in *International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–6.

[12] L. An, X. Chen, M. Kafai, S. Yang, and B. Bhanu, "Improving person re-identification by soft biometrics based reranking," in *ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2013, pp. 1–6.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[14] M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *IEEE 12th International Conference on Computer Vision (ICCV)*, 2009, pp. 1515–1522.

[15] C. Hue, J.-P. Le Cadre, and P. Perez, "Sequential monte carlo methods for multiple target tracking and data fusion," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 309–325, 2002.

[16] W. Yi, M. Morelande, L. Kong, and J. Yang, "A computationally efficient particle filter for multitarget tracking using an independence approximation," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 843–856, 2013.

[17] B. Yang and R. Nevatia, "Online learned discriminative part-based appearance models for multi-human tracking," in *Proceedings of the 12th European Conference on Computer Vision (ECCV) - Volume Part I*, 2012, pp. 484–498.

[18] 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)*, 2012, pp. 1815–1821.

[19] B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1918–1925.

[20] C.-H. Kuo, C. Huang, and R. Nevatia, "Multi-target tracking by online learned discriminative appearance models," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 685–692.

[21] Z. Qin and C. Shelton, "Improving multi-target tracking via social grouping," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1972–1978.

[22] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross, "Soft biometrics for surveillance: an overview," in *Machine Learning: Theory and Applications*.   Elsevier, 2013, pp. 327–352.