

Multi-pedestrian Tracking Based on Social Forces *

Hengle Ren, Fang Xu, Fengshan Zou, Kai Jia, Pei Di and Jie Kang

Abstract— Multi-pedestrian tracking based on video has always faced many problems. Tracking-by-detection paradigm is a popular method to solve these problems. For example, due to the influence of sensors, lighting, background, detection may result in some false detections and missed detections. In order to solve this problem, in this paper, we propose a new tracking method based on the social force model. Here, pedestrians are divided into two categories: candidate pedestrians and real pedestrians. The real pedestrians are the pedestrians we want to track. Both can be transformed into each other by their respective historical records. The social force model is used to predict the position of each person in the next frame, and the weighted distance between the detected pedestrian in the current frame and the detection in the next frame of image is calculated. According to the distance matrix, the Hungarian algorithm is used to assign identities so as to achieve the purpose of multi-pedestrian tracking. Our results were evaluated on the MOT challenges dataset and compared with existing advanced algorithms. The results show that this method outperforms traditional algorithms in the number of mostly tracked (MT), mostly lost (ML) and the number of frames processed per second (FPS). Including Particle filter, traditional social force model and Kalman filter algorithm tracking method.

I. INTRODUCTION

The task of multi-pedestrian tracking is to identify the same pedestrian in different frames by using a set of image sequences. And it is widely used in many aspect of our life, such as human-computer interaction, video surveillance, intelligent transportation, and autonomous navigation of robots [1]-[3]. In terms of monitoring, multi-pedestrian tracking technology can monitor the pedestrians on the road based on the videos taken in the streets or squares. This helps to measure the flow of pedestrians or further analyze the trajectories of pedestrians to ensure safety. In the application of robots, multi-pedestrian tracking allows the robot to walk safely and effectively in crowded places. Pedestrian tracking technology is even more important in certain special occasions, such as smart luggage and autonomous shopping carts. All of these require the robot to accurately track pedestrians in a complex environment and have high real-time requirements. However, there are still many challenges such as changing numbers of pedestrians, complex backgrounds, and mutual occlusion. In recent years, a large number of researchers have devoted themselves to the research of



Figure 1. The result of false positives and false negatives.

multi-pedestrian tracking and have also made remarkable achievements [4].

In many ways, tracking-by-detection has gradually become one of the most popular methods [5]. In this paradigm, the pedestrians of each frame are first detected by the object detector, and then the pedestrians in all the image sequences are correlated by the data association algorithm. However, these detections will appear false positive and missed detections. Fig. 1 shows an example of two cases. In the left figure, red ellipses indicate false negative results. In the right figure, red ellipses indicate false positive results. For these problems, we must minimize its impact on the tracking effect. A good tracker should have sufficient robustness to deal with this situation.

Pedestrian trajectory prediction is a key part of multi-pedestrian tracking. Many Kalman filter and particle filter algorithms are currently the most studied algorithms. The Kalman filter works better for the Gaussian linear model. A multi-target tracking method proposed by Sahbani B et al. based on Kalman filter and iterative-hungarian algorithm [6]. For a non-Gaussian and non-linear model, the particle filter solves better. An efficient object tracking method presented by Sangale S P et al. in video sequences using multiple features by embedding mean shift into particle filters [7]. However, these two types of algorithms also have significant limitations. Both assume that the number being tracked is known and fixed. The actual environment is complex and difficult to meet such harsh conditions. In particular, the robot is in a more complex environment. Once there are a varying number of target pedestrians in the environment, neither of these methods can effectively track the target. Therefore, it is difficult for us to use these two methods for robotics.

Actually, pedestrian behavior will be more complicated. They will be affected by the surrounding environment. Interested places will attract pedestrians. People will always maintain a safe and comfortable distance with other people. If there is a close distance between two pedestrians, there will be a repulsive force between them. There are many models that describe the interaction between the environment and pedestrians. Include the social force model, cellular automaton model, discrete choice model, queuing network model and so on. Compared to several other models, the

* Research supported by the National Key Research and Development Program of China 2017YFC0806700.

Hengle Ren and Jie Kang are with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: renhengle & kangjie@ sia.cn).

Fang Xu, Fengshan Zou, Kai Jia, Pei Di, are with Shenyang SIASUN Robot & Automation Co., LTD., China, Shenyang 110168, China. (e-mail: xufang & zoufengshan & jiakai & dipei@siasun.com).

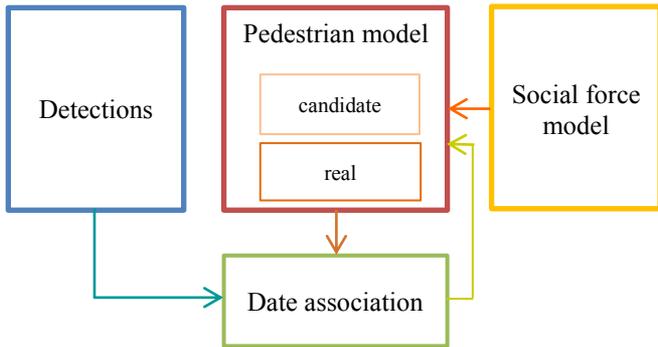


Figure 2. Tracking flowchart

social movement model proposed by Helbing D et al. is more visual and effective. The method proposed by Meyers R A also assumes that there is interaction between pedestrians [8]. He believes that the interaction between pedestrians is not only related to the absolute position of pedestrians, but also related to the velocity of pedestrians. Even if the model does not take into account the relative speed of movement between pedestrians, it can also calculate some of the characteristics of pedestrian movements, even in high-density people. Pellegrini et al. took into account a pedestrian's prediction goal and the influence of pedestrian collision avoidance, and proposed a multi-pedestrian tracking algorithm [9]. This method also introduces pedestrian speed when calculating the interaction between pedestrians. This is done by estimating the closest future distance in the pedestrian trajectory and calculating the distance in the potential. Y Wang et al combined particle filter and social force model to perform pedestrian tracking in unknown environments [10]. This method uses particle filtering to predict the distance between pedestrians and robots, relying on social forces to calculate the pedestrian's expected location and control robots to track pedestrians. Zhang X et al. considered all the people in the surrounding environment and proposed a new multi-pedestrian tracking algorithm in conjunction with social force model and mean shift [11]. In this method, authors utilize Mean shift which uses the color information of the image to estimate the position of pedestrians. In addition, in order to explain how the environment impacts pedestrian movements from the viewpoint of force, the author's proposed expansion of social forces is used to reflect the interaction of pedestrians.

Because the factors considered by the social force model are closer to the actual situation of pedestrians, they have obvious advantages over other algorithms. However, the above algorithm basically assumes that people in the scene are accurately detected in advance. In fact, there is no detection algorithm that can detect people in the video completely and correctly. As shown in Fig. 1, it is very common to get false negative and false positive results. Therefore, tracking under such conditions is prone to problems with wrong and missing target.

In this paper, we propose a new tracking-by-detection method based on the social force model to solve this problem. The overall framework of the method is shown in Fig. 2. For each frame, we use the detections obtained by the pedestrian detection algorithm as input, and these detections also have problems of misdetection and missed detection. However, we don't use the position of each detection directly as the final

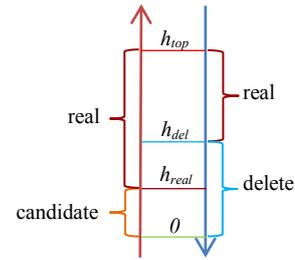


Figure 3. An history-based pedestrian model

output destination. We divided the detected pedestrians into two categories: candidate pedestrian and real pedestrian. The real pedestrian is the pedestrian we really want to track. Both identities can be converted according to their respective historical records. The social force model is used to predict the estimated position of each pedestrian in the next frame. The distance between the pedestrian detected in the current frame and the detection in the next frame is calculated, which is referred to as the weighted distance. Different identities have different weights. Finally, data association based on Hungarian algorithm is used to assign identities so as to achieve the purpose of multi-human tracking.

II. MATH

A. Pedestrian Model

Here, we propose a new pedestrian model. We divide pedestrians into two types: candidate pedestrians and real pedestrians. The real pedestrians are the pedestrians we want to track. All pedestrians have ID, history, position, desired position, and velocity attributes. The ID represents the identity of the current pedestrian, history records the past few frames of the pedestrian's position, and the desired position is the position of the pedestrian in the next frame estimated by the social force model. Between the two types of pedestrians can be converted to each other based on their respective historical records. Fig. 3 shows the conversion of the pedestrian model. Both types of pedestrians in the current frame need to match the detections in the next frame. For each successful match, if the pedestrian's historical record does not reach the maximum value of the real historical record h_{max} , the historical record is incremented by one. If the historical record of the pedestrian reaches the minimum value of the real pedestrian historical record h_{real} , the pedestrian is a real pedestrian. If no match is found, the historical record is decremented by one. If the historical record is smaller than h_{real} , the pedestrian is assigned to the candidate pedestrian category. If the historical record is less than the delete history record h_{del} , we believe that there is indeed no pedestrian in this location and delete the pedestrian. Finally, we output our tracking results for each real pedestrian's ID and location. Using this model can greatly improve the effectiveness of multi-pedestrian tracking.

B. Date Association

Pedestrian tracking is to associate the same pedestrian in different frames. For data association, the Hungarian algorithm is a valid algorithm [12]. The Hungarian algorithm was named by Edmonds, a Hungarian mathematician, in 1965. Hungarian algorithms are often used to match bipartite graphs. The core of the algorithm is to find an optimal matching way



Figure 4. The result of overlap detection.

to minimize the overall cost. Many researchers use this algorithm to solve the problem of pedestrian tracking. A fast tracking algorithm proposed by Bochinski E et al. does not require image information based on the Hungarian algorithm [13]. This is also an online method. There are three steps in our proposed data association method. The first step is to calculate the distance between the pedestrian in the current frame and the detection in the next frame. The distance is defined as follows:

$$d = w \cdot (p_{ped} - p_{det}) \quad (1)$$

where d is the distance between the current frame pedestrian and the next frame of detection while w is the weight of the distance. Different types of pedestrians have different weights. p_{ped} and p_{det} are the positions of the pedestrians in the current frame and detection in the next frame, respectively.

The second step, we use the Hungarian algorithm based the distance matrix and get a preliminary association. But the Hungarian algorithm does not distinguish the candidate from the real pedestrian.

Finally, we deal with the result of the second step. When the detection is associated with a candidate pedestrian and its distance from the real pedestrian is less than the maximum distance d_{max} , we preferentially associate it with the real pedestrian. If a detection is not associated with all pedestrians we add it as a candidate pedestrian. However, when a real pedestrian has been associated with a detection and the distance between another detection and this pedestrian is less than the minimum distance d_{min} , the detection can't be added as the candidate pedestrian. It can be seen from Fig. 4 that the two detections in the red ellipse are particularly close together.

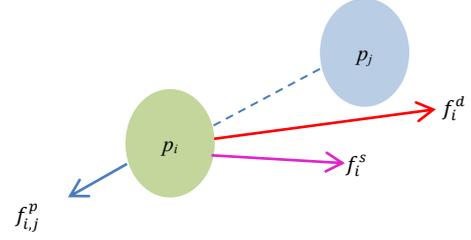


Figure 5. The force analysis of social force model.

This is the result of repeated detection on the same pedestrian. If we do not have this step, the program will consider the second detector as a new pedestrian and assign it an ID. We know it's just a misdetection. And this step of ours sets a detection area for each pedestrian, which can effectively deal with this problem.

C. Social Force Model

When pedestrians walk on the road, they are often affected by the external environment, such as other pedestrians, vehicles, beautiful scenery and so on. Here, we only discuss the impact from other pedestrians. The social force model can accurately describe the impact. Fig. 5 shows the basic principle of this model. The red arrow represents the driving force of pedestrian i 's own goal. The blue arrow represents the repulsive force of others and pink represents social force. The social force model is given by driving force and repulsive force from other pedestrians. The corresponding expressions are shown as follows:

$$\mathbf{f}_i^s = \mathbf{f}_i^d + \sum_{i \neq j} \mathbf{f}_{i,j}^p \quad (2)$$

where \mathbf{f} is the magnitude of the vector quantity, \mathbf{f}_i^s is the social force of pedestrian i . \mathbf{f}_i^d is the driving force of pedestrian i . $\mathbf{f}_{i,j}^p$ is the repulsive force from pedestrian j .

We can calculate a desired velocity based on the history of pedestrians. The relationship between pedestrian real velocity and desired velocity can be described by the driving force. It can be represented as follows:

$$\mathbf{f}_i^d = m_i(\mathbf{v}_i^d - \mathbf{v}_i) \quad (3)$$

where m_i is the mass of pedestrian i . \mathbf{v}_i^d is the desired velocity, and \mathbf{v}_i is the real velocity of pedestrian i . Here, we do not consider the sudden change in subjective consciousness of pedestrians.

The main component of social force model is the repulsive force from others:

TABLE I. TRACKING RESULTS ON THE MOT CHALLENGE TRAINING DATASET

Method	Indicator								
	recall \uparrow	precision \uparrow	MT \uparrow	ML \downarrow	FP% \downarrow	FN% \downarrow	IDs \downarrow	MOTA \uparrow	MOTP \uparrow
Kalman-HA	28.5	79.0	32	334	7.6	71.4	685	19.2	69.9
Kalman-HA2*	28.3	83.4	39	354	5.6	71.7	105	22.4	69.4
JPDA _m *	30.6	81.7	38	348	6.8	69.4	109	23.5	69.0
RNN-LSTM	37.1	73.5	50	260	13.3	62.9	572	22	69.0
Ours	47.6	64.8	68	231	25.7	52.4	870	19.6	71.6

a. * Denotes offline post-processing

b. The best value in each category is boldface.

TABLE II. TRACKING RESULTS ON THE MOT CHALLENGE TEST DATASET

Method	Indicator							
	MOTA% ↑	MOTP% ↑	MT% ↑	ML% ↓	FP% ↓	FN% ↓	IDs ↓	FPS ↑
TC_ODAL	15.1	70.5	3.2	55.8	21.1	62.7	637	1.7
JPDA _m *	23.8	68.2	5.0	58.1	10.4	65.2	365	32.6
GSCR	15.8	69.4	1.8	61.0	12.4	71.0	514	28.1
Ours	15.2	70.5	5.1	46.51	21.7	60.3	1,702	38.2

a. * Denotes offline post-processing

b. The best value in each category is boldface.

$$f_{i,j}^p = A \cdot \omega(\theta_{i,j}) \frac{v_i}{t_i} e^{-\frac{d_{i,j}}{B}} \quad (4)$$

where A denotes the magnitude of the repulsive force and B scales the range of the force, $d_{i,j}$ is the magnitude of distance vector between the pedestrian i . The term v_i/t_i assumes that pedestrians want to stop at t_i . In addition, due to the limited field of view of pedestrian, the repulsive force might not be isotropic. The term $\omega(\theta_{i,j})$ is the anisotropic factor of the social force model.

$$\omega(\theta_{i,j}) = \left(\lambda + (1 + \lambda) \frac{1 + \cos(\theta_{i,j})}{2} \right) \quad (5)$$

where λ scales the range of the anisotropic factor and $\theta_{i,j}$ describes the angle between velocity v_i and distance $d_{i,j}$.

III. EXPERIMENTS

A. Environment

Our work is done in C++, instead of in Python. Because C++ is expected to result in faster run time compared to languages such as Python. The experiments are, unless otherwise stated, run on a 2012 laptop, with a 2.4 GHz Intel Core i3-3110M, and 8 GB of RAM.

B. Parameter

Firstly, we set parameters for the pedestrian model. The maximum value of the real historical record h_{max} is 12 and the minimum h_{real} is 2. The delete history record h_{del} is set to 4.

Next, we set parameters for data association process. Different types of pedestrians have different weights. We assume that the weight of the distance between the current frame candidate pedestrian and the next frame of detection is 2. When the current pedestrian is real, the weight is 1. The maximum distance d_{max} is set to 45 and the minimum distance d_{min} is set to 30.

Finally, we initialize the social force model. We assume that everyone has a mass of 70 kg. The magnitude of the repulsive force is $A = 60 N$ and the range of the force is $B = 0.4 m$. The anisotropic factor is set to $\lambda = 0.5$.

C. Benchmark Evaluation

We will test the performance of our proposed method at the MOT Challenge 2015 benchmark. This data set is mainly used for multi-target tracking and multi-pedestrian tracking testing. This dataset has a total of two training and testing sets and each set has 11 video sequences. These 22 videos are taken from real cameras and contain a variety of complex scenes. For example, an unspecified number of people, moving cameras, dim lighting, people of different densities,

etc. All these have made it more difficult for pedestrians to track them. Because the true value of the test set is not public, everyone uploads their own results to the official website for evaluation. And the official website will give some performance indicators so that everyone can measure the merits of their own methods. Here we show a comparison of some of these indicators, including precision and recall. FP and FN represent the number of false positives and false negatives in the trace results, respectively. In addition, the MT indicates that the pedestrian trajectory can be recovered by more than 80% of the pedestrians and the ML indicates that the number is below 20%. IDs represent the number of identity switches. The FPS reflects the speed with the number of image frames that can be processed per second. Finally, MOTA and MOTP represent the accuracy and precision of multi-target tracking, respectively [14]. The arrows next to each indicator are used to indicate the optimal direction. Upward (↑) means that the bigger the number, the better. On the contrary, downwards (↓) means that the smaller the better.

Firstly, we compared our method with four baselines on the MOT Challenge training dataset. The performance comparison of the results is presented in Table 1. The first baseline (Kalman-HA) combines the Kalman filter with the Hungarian algorithm for tracking. Each unassigned measurement will begin tracking and will stop tracking once the measurement is missed. Secondly, Kalman-HA2 is similar to the first one. But it uses a set of heuristics later to remove the wrong trajectory. Next, JPDA_m recently proposed by Rezatofighi et al. solves these problems from approximate examples, using recurrent neural networks (RNNs). Finally, the approach proposed by Milan A et al. also uses neural network methods to solve data association and trajectory prediction problems [15]. Our method outperforms other algorithms in terms of recall, MT, ML, FN, and MOTP. Although Kalman-HA2 is superior to us in precision, FP, and IDs, this method is an off-line method and the whole video sequence information needs to be used in tracking. However, robotic applications cannot provide global video sequences but only online methods. So this method has great limitations. JPDA_m is also an offline method.

Next, we compared our method with three baselines on the MOT Challenge testing dataset. The accuracy of detectors can greatly affect the results of pedestrian tracking. Many of the well-known methods disclosed on this dataset use their own detectors, and we cannot compare the effects of these algorithms. So for the sake of fairness, we have chosen some algorithms that use the dataset exposed detectors to compare the results. All the results are compared in Table 2.



Figure 6. Our tracking results on the MOT Challenge sequence.

TC_ODAL is a multi-target tracking method proposed by Bae S H et al. in 2014. This method utilizes the tracklet confidence and the appearances of objects to improve tracking performance [16]. TC_ODAL is much better than us in IDs, but this method can only perform 1.7 times per second, which is not enough for robotic application. GSCR proposed by Fagot-Bouquet L et al. uses the global sparse collaborative representations to achieve the goal of tracking multiple targets [17]. In the MOTA, FP, IDs, GSCR is better than us. Since we have fully considered the requirements of real-time, we need to complete the tracking in the shortest possible time. Therefore, the above aspects will be slightly worse. So our method is less than us in terms of MOTP, ML, and FN, and our method performs faster. The method proposed by us is superior to other methods in terms of MOTP, MT, ML, FN, and FPS. The remaining three indicators JPDA_m are optimal, but because it is off-line, we cannot use them in robots.

Fig. 6 shows the results of our method on PETS09-S2L1. This video sequence contains obstacles such as mutual obstruction between pedestrians, sudden changes in pedestrian trajectories, and pedestrians walking backwards. The images from left to right show the tracking results at frame $f = 285, 290, 295$ and 300 . Different color rectangles represent the position of different pedestrians. History trajectory is shown as small dots behind every pedestrian. The first number above the pedestrian represents the ID of the pedestrian and the second represents the history record. It can be seen from Figure 6 that pedestrians with ID 7 and 10 are represented by purple and green rectangles, respectively. The two people had mutual obstruction due to staggered walking. If you do not predict the trajectory of pedestrians, relying solely on the detector's position information, the identities of the two are prone to errors. And we use the social force model to make predictions on the trajectories of pedestrians. At the 295th frame, the two pedestrians overlap each other in the occlusion position. We predict that pedestrians with an ID of 7 will then walk to the upper right, while pedestrians with an ID of 10 will walk to the left. So at the 300th frame, we associate the left detector with the pedestrian with ID 10, and the detector in the upper right is associated with the pedestrian with ID 7. Therefore, the two pedestrians are accurately tracked in the case of mutual obstruction.

As the number of pedestrians tracked increases, the dimensions of the matrix solved by the Hungarian algorithm also continue to increase. The corresponding calculation will greatly increase. In order to ensure the real-time calculation, the maximum number of pedestrians in the field of vision should not exceed 20 at the maximum.

Overall, although the accuracy of our method tracking is not the best, it is an online tracking method that can run in real time.

IV. CONCLUSION

In this paper, we propose a new tracking method based on social force model. First, we constructed a pedestrian model to improve the robustness of the algorithm. Then, in terms of data association, we supplemented the Hungarian algorithm to make it adaptable to our pedestrian model. Finally, using the social force model to predict pedestrian trajectories makes the data association more accurate.

On the one hand, this algorithm can effectively solve the problem that multiple pedestrians are mismatched due to mutual obstruction. On the other hand, for the problem of false detection or missed detection of the detection algorithm itself, our algorithm can mitigate the impact of this problem on the final tracking of pedestrians.

From the results of our comparison with other algorithms on the MOT challenge dataset, the overall effect of our method tracking is superior to other algorithms compared. Although the accuracy of this method is not the best, its speed is very fast. So it is suitable for robots, security, traffic monitoring and other areas with high real-time requirements.

In the future, at the expense of a small amount of operating speed, we will consider combining position information and image information to improve accuracy. In addition, the social force model can also be further optimized to achieve pedestrian tracking in crowded scenarios.

REFERENCES

- [1] Khan, Zia, T. Balch, and F. Dellaert. "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 27.11(2005):1805-1819.
- [2] Rao, Yunbo. *Automatic vehicle recognition in multiple cameras for video surveillance*. Springer-Verlag New York, Inc. 2015.
- [3] Naqvi, Syed Mohsen, M. Yu, and J. A. Chambers. "A Multimodal Approach to Blind Source Separation of Moving Sources." *IEEE Journal of Selected Topics in Signal Processing* 4.5(2010):895-910.
- [4] Chen, Jiahui, et al. "Enhancing Detection Model for Multiple Hypothesis Tracking." *Computer Vision and Pattern Recognition Workshops IEEE, 2017:2143-2152*.
- [5] Murray, Samuel. "Real-Time Multiple Object Tracking - A Study on the Importance of Speed." (2017).
- [6] Sahbani, Bima, and W. Adiprawita. "Kalman filter and iterative-hungarian algorithm implementation for low complexity point tracking as part of fast multiple object tracking system." *International Conference on Frontiers of Information Technology IEEE, 2016:109-115*.

- [7] Sangale, S. P., and S. B. Rahane. "Live object monitoring, detection and tracking using mean shift and particle filters." *International Conference on Inventive Computation Technologies IEEE*, 2017:1-6.
- [8] Meyers, Robert A. "Encyclopedia of Complexity and Systems Science." *Encyclopedia of Complexity & Systems Science* (2008):1035-1081.
- [9] Pellegrini, S, et al. "You'll never walk alone: Modeling social behavior for multi-target tracking." *IEEE, International Conference on Computer Vision IEEE*, 2009:261-268.
- [10] Wang, Yang, et al. *People Tracking in Unknown Environment Based on Particle Filter and Social Force Model. Intelligent Robotics and Applications*. 2017.
- [11] Zhang, Xuguang, et al. "Extended social force model-based mean shift for pedestrian tracking under obstacle avoidance." *Iet Computer Vision* 11.1(2017):1-9.
- [12] Kuhn, H. W. "The Hungarian method for the assignment problem." *Naval Research Logistics* 52.1(2005):7-21.
- [13] Bochinski, Erik, V. Eiselein, and T. Sikora. "High-Speed tracking-by-detection without using image information." *IEEE International Conference on Advanced Video and Signal Based Surveillance IEEE*, 2017.
- [14] Bernardin, Keni, and R. Stiefelhagen. "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics." *Eurasip Journal on Image & Video Processing* 2008.1(2008):246309.
- [15] Milan, Anton, et al. "Online Multi-Target Tracking Using Recurrent Neural Networks." (2016).
- [16] Bae, Seung Hwan, and K. J. Yoon. "Robust Online Multi-object Tracking Based on Tracklet Confidence and Online Discriminative Appearance Learning." *IEEE Conference on Computer Vision and Pattern Recognition IEEE Computer Society*, 2014:1218-1225.
- [17] Fagot-Bouquet, Low, et al. "Online multi-person tracking based on global sparse collaborative representations." *IEEE International Conference on Image Processing IEEE*, 2015:2414-2418.