# A Particle Filter based Multi-person Tracking with Occlusion Handling

Ruixing Yu[1], Bing Zhu[2], Wenfeng Li[3] and Xianglong Kong[3]
[1]*School of Astronautics, Northwestern Polytechnical University, No.127 Youyi West Road, Xian Shannxi, China*
[2]*School of Electronic Engineering, Xian Shiyou University, No.18 Dianzi er Road, Xian Shannxi, China*
[3]*Shanghai Institute of Satellite Engineering, No.3666 Yuanjiang Road, Shanghai, China*

Keywords: Multi-person Tracking, Occlusion, Reliability of Tracklets, Particle Filter.

Abstract: A multi-person tracking method is proposed concerning how to conquer the difficulties such as occlusion and changes in appearance which makes algorithm hard to get the correct positions of object. First, we indicate whether the target is blocked or not, through computing the Reliability of Tracklets (RT) based on the length of tracklets, appearance affinity and the size. Then, we propose a "correct" observation sample selection method and only update the weights of particle filter when the RT is high. Last, the greedy bipartite algorithm is used to realize data association. Experiments show that tracking can be successfully achieved even under severe occlusion.

## 1 INTRODUCTION

Multiple targets tracking play a key role in various applications, such as surveillance, robotics, human motion analysis and others. Tracking multiple objects in real time in an accurate way is to find all target trajectories in a given video scene while ensuring the target identities are correct. However, due to frequent occlusion by clutter or other objects, similar appearances of different objects, and other factors, target trajectories are fragmented. There are challenges made linking the fragmented trajectories up so difficult: such as targets often exit the field of view and enter back later on; and often become occluded by other targets or objects in the scene. These factors will get the appearance of target changed greatly, which make the target re-identify difficult. Thus, the tracking methods will suffer from track fragmentations and identity switches. In this paper, the Reliability of Tracklets (RT) is used to decrease the effect of occlusion, and only update the weights of particle filter when the RT is high.

The main contributions of this paper can be summarized as follows: (i) Selecting the correct object positions from the output set of the particle filter and detectors; (ii) An observing the selection process with RT is brought into the particle filter that could deal with partial object occlusion and generate reliable tracklets.

## 2 RELATED WORKS

Recently, with the big progress of object detection(Yang and Nevatia, 2012, Felzenszwalb et al., 2014, Dollar et al., 2012, Dollár et al., 2010), the detect-then-track approaches(Breitenstein et al., 2011, Brendel et al., 2011, Pellegrini et al., 2010) have become increasingly popular. The main idea of the detect-then-track approaches is that a detector is run on each frame to get the position and size of target, and then the data association is used to linked detections across multi-frames to obtain target trajectories and must not be assigned two different detections to the same target. Classical data association approaches include probability data association filter (PDA)(Bar-Shalom et al., 2010), joint probability data association filter (JPDA)(Fortmann et al., 1983), greedy matching (Pirsiavash et al., 2011), hungarian algorithm (H., 1955) or particle filters (Breitenstein et al., 2009), et al. To distinguish between each separate target and improve the accuracy of the data association, the appearance model (Xing et al., 2009, Li et al., 2008) is usually employed to associate the target and detections. And to conquer frequent



Yu, R., Zhu, B., Li, W. and Kong, X.
A Particle Filter based Multi-person Tracking with Occlusion Handling.
DOI: 10.5220/0005961602010207
In *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2016) - Volume 2*, pages 201-207
ISBN: 978-989-758-198-4

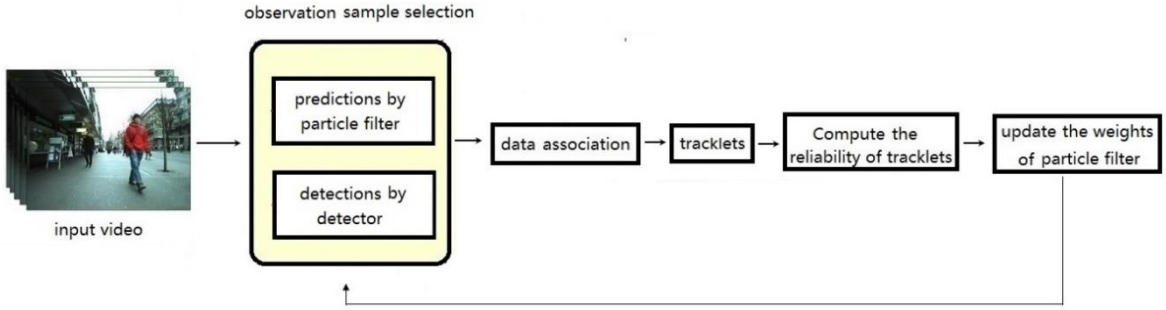



Figure 1: An Overview of our Proposed Multi-Object Tracking Framework.

prolonged occlusions and target interactions, the appearance should be updated. Online methods(Grabner et al., 2006, Collins et al., 2005, Babenko et al., 2009, Kalal et al., 2010) can be used to update the appearance of the object. Also to improve the accuracy of multi-object tracking, along with appearance models based on simple cues like color histograms, linear motion models help in maintaining track identity while linking tracklets by enforcing motion smoothness. Like, B. Wu (Wu and Nevatia, 2005) associates object hypotheses with detections by evaluating their affinities for appearances, positions, and sizes. X. Song(Song et al., 2008) associates object hypotheses with detections using three affinity terms: position, size, and the score of the person-specific classifier based on online learning. However, the above kind of detect-then-track approaches depends on the precise of the detectors. Unfortunately the detectors are often not perfect and can fail to detect the object of interest, or identify a false target position, which will accumulative error over time, resulting in tracking drift and failure. We proposed a correct observation sample selection method to compensate the error made by the detector. Through computing the RT of tracklets, we indicate whether the target is blocked or not. Our proposed framework is shown in figure 1.

## 3 SAMPLE SELECTION

The Particle filtering(Breitenstein et al., 2009) is a method for state estimation based on a Monte Carlo method and it handles nonlinear models with non-Gaussian noise. The particle filter approximates the state of object in these two steps by updating the weights of the particles. And it can be done by well-known two-step recursion procedure:

Predict:

$$(s_t^i|o_{1:t-1}^i) = \int p(s_t^i|s_{t-1}^i)p(s_{t-1}^i|o_{1:t-1}^i)\, ds_{t-1}^i \quad (1)$$

Update:

$$p(s_t^i|o_{1:t}^i) \propto p(o_t^i|s_t^i)p(s_t^i|o_{1:t-1}^i) \quad (2)$$

Where $s_t^i$ is the position and size of particular object *i* at frame *t*. $s_{1:t}^i$ is the states of the object *i* form the frame 1 to *t*. $o_{1:t}^i$ is the observations of the object *i* form the frame 1 to *t*.

The state of the object can be well updated, only if the system has a reliable observation model. However if some object(s) is blocked, we won't have the correct position of target. A lot of methods were proposed to solve this problem, like (Yang and Nevatia, 2012, Song et al., 2008). They collected *N* images or image patches with different locations and scales from a small neighbourhood around the current tracking location, which made the prediction method confued which one is the correct observation sample. So we only select one "correct" sample as observation sample. The selection process is illustrated in figure 2.

Since tracklets with low RT (Reliability of Tracklets) values are likely to be polluted by occlusion, we propose to infer occlusion information from the RT scores and eventually utilize only those with high RT. We assume that a target's appearance changes little in short time. So we use the previous n (n=5) frames to represent the appearance of object to be tracked. In order to represent the target more precisely, the cumulative histogram of the past n frames is used to represent the targets. The HOG (Histogram of Oriented Gradient)(Dalal and Triggs, 2005) and RGB Histogram techniques are used to generate the features. The benefits of our technique include: reducing the impact of the occlusion, by updating the observation template only with samples of RT values higher than 0.5.

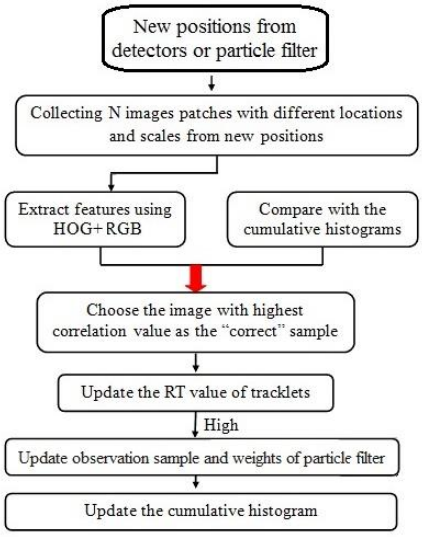



Figure 2: "Correct" observation sample selection process.

## 4 RELIABILITY CRITERION OF TRACKLETS

In order to get reliable trajectories, a criteria is proposed to judge the reliability of the target trajectory. In our algorithm, a tracklet which is considered as reliable needs to fulfill three requirements:

- It is longer than a certain threshold;
- A high affinity between a tracklet and an associated detection;
- The size of the detected object has not changed substantially. That means the target is not occluded and successfully segmented from the background.

According to these three constraints, we propose the criteria to calculate the reliability of the target trajectory. The formula:

$$RT(T_i)=\left(\frac{1}{L}\sum_{k=s}^{e}A(T_i,d_i^k)\right)\times\left(1-\frac{L-W}{L}\right)\times\left(1-\frac{\left|\#F(t)-\frac{1}{n}\sum_{x=1}^{n}\#F(t-x)\right|}{\frac{1}{n}\sum_{x=1}^{n}\#F(t-x)}\right) \quad (3)$$

where *RT* is an abbreviation for Reliability of Tracklets. $T_i$ is tracklet of target *i*; $d_i^k$ is detections. *s* and *e* are the time stamps of the start- and end-frame of the tracklet. *A* is used to calculate the affinity between trajectories and detections. $\#F(t)$ indicates the total number of pixels in detecting box at frame *t*; while $\sum_{x=1}^{n}\#F(t-x)\Big/n$ represents the average pixel numbers of *x* frames before frame *t* in their respective detecting box. *L* is the length of a tracklet, and *W* is the number of frames in which the object *i* is missing due to occlusion by other objects or unreliable detection, and is given as:

$$W=F_{end}^{i}-F_{begin}^{i}-L^{i}+1 \quad (4)$$

Where $F_{end}^{i}$ is the end of frame of the target *i*, and $F_{begin}^{i}$ is the start of frame of the target *i*, $L^i$ is the length of tracklets of target *i*.

From formula (3) we can see that the larger RT value is, the more reliable the trajectory is. So we use the formula (3) to judge the reliability of the trajectory.

The appearance, shape and motion attributes are calculated to compare the affinity between two tracklets. See more details in literature [24].

$$A(T^i,d_t^i)=A_a(T_i,d_t^i)\times A_m(T_i,d_t^i)\times A_s(T_i,d_t^i) \quad (5)$$

Appearance Model: We use HOG+RGB histograms as the appearance model of a tracklet.

$$A_a(T_i,T_j)=\exp(-d(T_i,T_j)) \quad (6)$$

In equation (6), *d* is the Bhattacharya distance.

Motion Model: We calculate both the forward velocity and backward velocity of the tracklet as its motion model. The forward velocity is calculated from the refined position of the tail response of the tracklet while the backward velocity is calculated from the refined position of the head response of the tracklet.

$$A_m(T_i,T_j)=G(p_i^{tail}+v_i^F\Delta t;\ p_j^{head},\sum_j^B)\bullet G(p_j^{tail}+v_j^B\Delta t;\ p_i^{tail},\sum_i^F) \quad (7)$$

In equation (7), motion model in forward direction is represented by a gaussian $\{x_i^F,\sum_i^F\}$, and in backward direction by a gaussian $\{x_i^B,\sum_i^B\}$. $p_i^{head}$ is the position of the head response of tracklet $T_i$ and $p_i^{tail}$ is the position of the tail response of tracklet $T_i$.

Size Model: we calculate the height *h* and width *w* of objects.

$$A_s(T_i,T_j)=\exp\left(-\left\{\frac{h_i-h_j}{h_i+h_j}+\frac{w_i-w_j}{w_i+w_j}\right\}\right) \quad (8)$$

Where, *h* is height and *w* is width.

# 5 DATA ASSOCIATION BASED ON GREEDY BIPARTITE GRAPH MATCH

The greedy bipartite graph match(Breitenstein et al., 2009, Birnbaum and Mathieu, 2008) is used to associate the detections with existing trajectories in every frame. First, the similarity is computed between tracklets and detections to get the matching pairs. Then, the pair with maximum similarity score is iteratively selected, and the rows and columns belonging to tracklets and detections are deleted. This is repeated until no further valid pair is found. Finally, in order to guarantee a selected detection is actually a good match to a target, we only save the pairs above the threshold we set.

The detections which are not associated with any existing trajectories are used to initialize a new potential trajectory. Once the length of a potential trajectory becomes longer than a threshold, it gets formally initialized. On the other hand, when a new detection is associated to a trajectory, we update all its state variables, namely, the position, the size, the velocity, the RT based on the new detection. However, due to occlusion or miss-detection, there are some fragmentations in trajectories, we use extrapolation and interpolation method to complement trajectory.

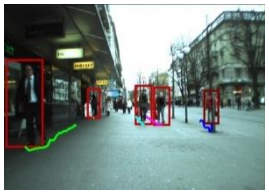
# 6 EXPERIMENTS AND ANALYSIS

## 6.1 Tracking Results

We tested our algorithm on several publicly challenging available video sequences, which are ETH BAHNHOF sequence, ETH SUNNY DAY sequence, ETH JELMOLI sequence. We use the ground truth annotations and automatic evaluation code provided by (Bing et al., 2014) for quantitative evaluation. In these provided annotations, “BAHNHOF” sequence contains 95 individuals over 399 frames. “SUNNY DAY” sequence contains 30 individuals over 354 frames. Below are the tracking results and the trajectories of targets.

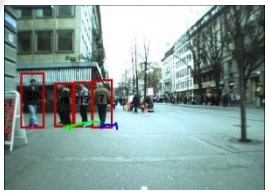
(a) Frame 11

(b) Frame 19(target 1 is blocked by target 3)



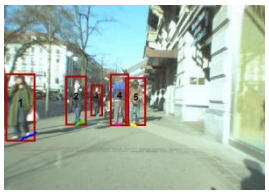
(c) Frame 71(From the left to right: the second target, labelled by 7, stopped moving in frame 71)

(d) Frame 186(Our algorithm correctly recaptured the target. Then until it disappeared from the visual field, the label of the target is still No.7)

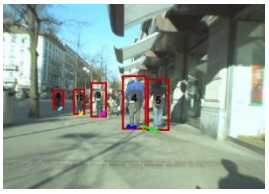


(e) Frame 301

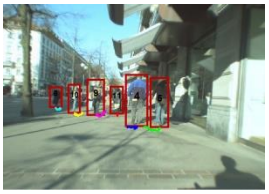
Figure 3: Tracking results on BAHNHOF scene.



(a) Frame 7

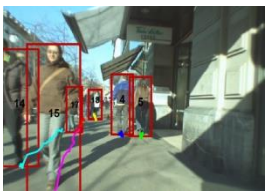


(b) Frame 88(target 11 and 13 are lost)



(c) Frame 94(target 11 is recaptured)

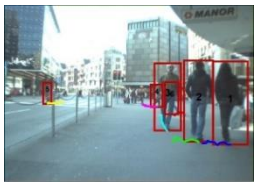
(d) Frame 118(target 13 is recaptured)



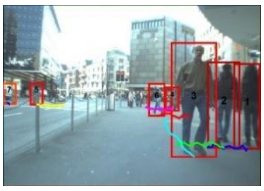
(e) Frame 196(target 17 is blocked by target 16, however our method still gets the correct trajectory)

(f) Frame 302

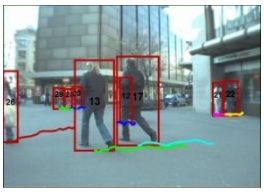
Figure 4: Tracking results on Sunny Day scene.



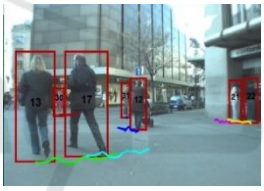
(a) Frame 17(target 4 and 6 are blocked by target 3, however our method still gets the correct trajectories, see in frame 27 below)



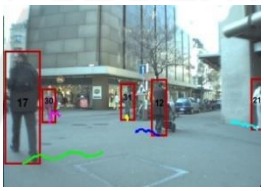
(b) Frame 27



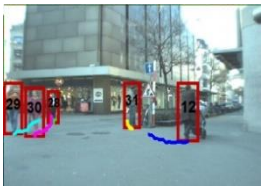
(c) Frame 271(The target 30 is blocked until frame 289.)



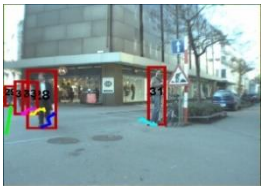
(d) Frame 289(Target 30 is recaptured)



(e) Frame 307(Target 30 is tracked correctly)



(e) Frame 323(Target 30 is tracked correctly)



(f) Frame 358

Figure 5: Tracking results on Jelmoli scene.

Table 1: ETH dataset tracking results comparison.

| Method | Recall | Precision | FAF | GT | MT | PT | ML | Frag | IDS |
|---|---|---|---|---|---|---|---|---|---|
| PRIMPT(Kuo and Nevatia, 2011) | 76.8% | 86.6% | 0.891 | 125 | 58.4% | 33.6% | 8.0% | 23 | 11 |
| Our method | 79.6% | 90.1% | 0.696 | 125 | 66.3% | 24.2% | 8.4% | 17 | 5 |

### 6.2 Evaluation Metric

Since it is difficult to use a single score to judge any tracking performance, several definitions are used as follows:

- Recall: correctly matched detections / total detections in ground truth.
- Precision: correctly matched detections / total detections in the tracking result.
- FAF: average false alarms per frame.
- GT: the number of trajectories in ground truth.
- MT: the ratio of mostly tracked trajectories, which are successfully tracked for more than 80%.
- ML: the ratio of mostly lost trajectories, which are successfully tracked for less than 20%.
- PT: the ratio of partially tracked trajectories, i.e., 1-MT-ML.
- Frag: fragments, the number of times the ground truth trajectory is interrupted.
- IDS: id switch, the number of times that a tracked trajectory changes its matched id.

Higher scores the recall, precision and MT are the better results of tracking algorithm are. While, lower scores the FAF, ML, PT, Frag and IDS are indicate the better results of the tracking method.

We evaluate our approach on two public sequences: ETH BAHNHOF sequence and ETH SUNNY DAY sequence. These two sequences are captured by a stereo pair of cameras mounted on a moving child stroller in a busy street scene. Because of the low view angle and forward moving cameras, occlusions and interactions of the targets frequently occur in these video sequences, which make the dataset rather challenging. For fair comparison, the two sequences are both from the left camera and also use the same ground truth as reference(Kuo and Nevatia, 2011). The tracking evaluation results are shown in Table 1.

Compared with (Kuo and Nevatia, 2011), the improvement is obvious for some metrics. Our approach achieves the highest recall. It also achieves the lowest Frag, ID switches. Meanwhile, our approach achieves competitive performance on precision, false alarms per frame compared with (Kuo and Nevatia, 2011).

## ACKNOWLEDGEMENTS

This work was supported, in part, by the National Natural Science Foundation of China (Grant No. 61101191), Aeronautical Science Foundation of China (Grant No. 20130153003), Science and technology research of Shaanxi Province(Grant No. 2013K09-18), and SAST Foundation (Grant No. SAST201342, No. SAST2015040)

## REFERENCES

Yang, B. & Nevatia, R. An Online Learned Crf Model For Multi-Target Tracking. In Cvpr, 2012. 2034-2041.

Felzenszwalb, P. F., Girshick, R. B., Mcallester, D. & Ramanan, D. 2014. Object Detection With Discriminatively Trained Part-Based Models. *Ieee Transactions On Pattern Analysis & Machine Intelligence* 32**,** 6-7.

Dollar, P., Wojek, C., Schiele, B. & Perona, P. 2012. Pedestrian Detection: An Evaluation Of The State Of The Art. *Pattern Analysis & Machine Intelligence Ieee Transactions On*, 34**,** 743-761.

Doll R, P., Belongie, S. & Perona, P. 2010. The Fastest Pedestrian Detector In The West. *Bmvc***,** 1-11.

Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E. & Van Gool, L. 2011. Online Multiperson Tracking-By-Detection From A Single, Uncalibrated Camera. *Ieee Transactions On Pattern Analysis & Machine Intelligence*, 33**,** 1820-1833.

Brendel, W., Amer, M. & Todorovic, S. Multiobject Tracking As Maximum Weight Independent Set. Computer Vision And Pattern Recognition (Cvpr), 2011 Ieee Conference On, 2011. 1273-1280.

Pellegrini, S., Ess, A. & Gool, L. V. 2010. Improving Data Association By Joint Modeling Of Pedestrian Trajectories And Groupings. *Lecture Notes In Computer Science*, 6311**,** 452-465.

Bar-Shalom, Y., Daum, F. & Huang, J. 2010. The Probabilistic Data Association Filter. *Ieee Control Systems*, 29**,** 82-100.

Fortmann, T. E., Bar-Shalom, Y. & Scheffe, M. 1983. Sonar Tracking Of Multiple Targets Using Joint Probabilistic Data Association. *Oceanic Engineering Ieee Journal Of*, 8**,** 173-184.

Pirsiavash, H., Ramanan, D. & Fowlkes, C. C. Globally-Optimal Greedy Algorithms For Tracking A Variable Number Of Objects. Proceedings Of The 2011 Ieee

Conference On Computer Vision And Pattern Recognition, 2011. 1201-1208.
H., W. 1955. The Hungarian Method For The Assignment Problem. *Naval Research Logistics Quarterly*, 2**,** 83-97.
Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E. & Van Gool, L. Robust Tracking-By-Detection Using A Detector Confidence Particle Filter. Computer Vision, 2009 Ieee 12th International Conference On, 2009. 1515-1522.
Xing, J., Ai, H. & Lao, S. Multi-Object Tracking Through Occlusions By Local Tracklets Filtering And Global Tracklets Association With Detection Responses. Ieee Conference On Computer Vision & Pattern Recognition, 2009. 1200-1207.
Li, Y., Ai, H., Yamashita, T., Lao, S. & Kawade, M. 2008. Tracking In Low Frame Rate Video: A Cascade Particle Filter With Discriminative Observers Of Different Life Spans. Ieee Trans Pattern Anal Mach Intell. *Ieee Transactions On Pattern Analysis & Machine Intelligence*, 30.
Grabner, H., Bischof, H. & Grabner, M. 2006. Real-Time Tracking Via On-Line Boosting. *Bmvc***,** 47-56.
Collins, R. T., Liu, Y. & Leordeanu, M. 2005. Online Selection Of Discriminative Tracking Features. *Ieee Transactions On Pattern Analysis & Machine Intelligence* 27**,** 1631-1643.
Babenko, B., Yang, M. H. & Belongie, S. 2009. Visual Tracking With Online Multiple Instance Learning. *Ieee Trans Pattern Anal Mach Intell***,** 983-990.
Kalal, Z., Matas, J. & Mikolajczyk, K. P-N Learning: Bootstrapping Binary Classifiers By Structural Constraints. Proceedings / Cvpr, Ieee Computer Society Conference On Computer Vision And Pattern Recognition. Ieee Computer Society Conference On Computer Vision And Pattern Recognition, 2010. 49-56.
Wu, B. & Nevatia, R. Detection Of Multiple, Partially Occluded Humans In A Single Image By Bayesian Combination Of Edgelet Part Detectors. Computer Vision, 2005. Iccv 2005. Tenth Ieee International Conference On, 2005. 90-97.
Song, X., Cui, J., Zha, H. & Zhao, H. 2008. *Vision-Based Multiple Interacting Targets Tracking Via On-Line Supervised Learning*, Springer Berlin Heidelberg.
Dalal, N. & Triggs, B. Histograms Of Oriented Gradients For Human Detection. Computer Vision And Pattern Recognition, 2005. Cvpr 2005. Ieee Computer Society Conference On, 2005. 886-893 Vol. 1.
Birnbaum, B. & Mathieu, C. 2008. On-Line Bipartite Matching Made Simple. *Acm Sigact News*, 39**,** 80-87.
Bing, W., Gang, W., Kap Luk, C. & Li, W. Tracklet Association With Online Target-Specific Metric Learning. Computer Vision And Pattern Recognition (Cvpr), 2014 Ieee Conference On, 23-28 June 2014 2014. 1234-1241.
Kuo, C. H. & Nevatia, R. How Does Person Identity Recognition Help Multi-Person Tracking? Proceedings /Cvpr, Ieee Computer Society Conference On Computer Vision And Pattern Recognition. Ieee Computer Society Conference On Computer Vision And Pattern Recognition, 2011. 1217-1224.