2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Skeleton model based behavior recognition for pedestrians and cyclists from vehicle scene camera

Qiwen Deng, Renran Tian, Yaobin Chen and Kang Li

Abstract—**With the significant advances in computer vision research, skeleton model based human pose recognition has become more accurate and time-efficient, although most of the applications are limited in laboratory environment or on surveillance videos. This paper proposes a pose tracking and behavior recognition method from in-vehicle scene camera. It will not only detect pedestrians on the road, but also generate their skeleton models describing head, limb, and trunk movements. Based on these more detailed movements of body parts, the proposed method is designed to track poses of pedestrians and cyclists with the potentials to enable automated pedestrian gesture reading and non-verbal interactions between autonomous vehicles and pedestrians. The proposed algorithm has been tested on different databases including TASI 110-car naturalistic driving database and Joint Attention for Autonomous Driving (JAAD) database. Results show that key frames describing different pedestrian and cyclist negotiation gestures are detected from the raw video streams using the proposed method. These results will improve our understanding of pedestrian and cyclist's intentions and can be further used for autonomous vehicle control algorithm development.**

Index Terms – pose tracking, behavior recognition, pedestrian and cyclists, human-autonomous vehicle interaction

I. Introduction

Autonomous driving research has been among hottest research topics for the last several years. Extensive research has already been conducted to develop automated driving systems for cars without drivers involved on highways. They can actively perceive surrounding environment and react correspondingly, such as avoid obstacles or change lanes. Nonetheless, situation in urban area is more complex than highway. To further develop algorithms in urban area, more road users should be considered such as pedestrians and cyclists. Existing algorithms focus more on safety and consider less making vehicles more intelligent and act as a human driver.

Pedestrian detection and tracking algorithms have been widely developed and utilized in vehicle pre-crash systems to issue warnings and perform automated braking. However, these safety-oriented features are not sufficient to read pedestrian gestures and estimate pedestrian intentions, which are critical functionalities to autonomous driving in mixed traffic situations with pedestrians and cyclists. Other than tracking surrounding vehicles' behavior in highway, tracking behavior of pedestrians and cyclists in urban area is much more challenging since pose features are involved.

Qiwen Deng, Renran Tian, Yaobin Chen are with Transportation Active Safety Institute and Department of Electrical and Computer Engineering, Indiana University Purdue University Indianapolis, Indianapolis, IN, 46202, USA (e-mail: {qiwdeng, rtianu, ychen}@iupui.edu).

Kang Li is with Department of Industrial and Systems Engineering, School of Engineering, Rutgers University, New Brunswick, NJ, 08901, USA (e-mail: kl419@soe.rutgers.edu).

To understand these features and further to use them for intention analysis, more precise detection and tracking algorithms should be developed. Tracking pedestrians and cyclists with bounding box is prevalent in previous research to understand movement of pedestrians and cyclists. Most of these research works start with full-body detection of pedestrians and cyclists. Histogram of Oriented Gradients (HOG) [1] and part-based model [2] are previously used for human detection purpose. For tracking of pedestrians and cyclists after detection, many strategies have been applied such as Mean-shift [3], globally-optimal greedy algorithm [4] and continuous energy minimization [5]. Tracking pedestrians has also been applied by using Markov decision processes [6]. However, tracking with bounding box will only give position and speed of pedestrians and cyclists. We will know little about the intention of the pedestrians and cyclists largely carried by their pose features.

Furthermore, most of research focus on tracking road users from fixed surveillance camera [7]. While few research has been done from the perspective of moving scene camera, especially when it is still challenging to track poses of pedestrian and cyclist from moving camera behind the windshield of vehicles. To our best knowledge, our presented paper is one of the first such studies to propose a method to track and recognize full-body pose of pedestrians and cyclists from a moving in-vehicle camera based on their skeleton model.

II. Previous Work

Previous work with pedestrian and cyclist tracking are mostly based on bounding box. Different object detection methods have been applied for pedestrian and cyclist detection. P. Viola *et al.* proposed a motion and appearance based detector in pedestrian detection [8]. A HOG detector for pedestrian was further introduced in 2005 by N. Dalal *et al.* [9]. Further extension of the HOG method called Deformable parts model (DPM) [10] by P. Felzenszwalb *et al.* in 2008 has been proposed to improve deformation effect. Cyclist detection research, on the other hand, are less and mostly based on HOG, SVM and part-based model [11] [12]. Recently developed pedestrian and cyclist detection algorithms have been improved by applying Faster R-CNN [13] and Mask R-CNN [14] through deep regional CNN. The Concurrent detection algorithm by using upper-body detection with bounding box and Fast R-CNN has been proposed to simultaneously detect pedestrians and cyclists from moving camera [15]. Z. Jiang *et al.* introduced a fixed-camera-based tracking algorithm by integrating multiple models with an ellipse bounding box [24]. However, precise pose information cannot be extracted from these published studies.

Skeleton-based pedestrian and cyclist detection and tracking on the other hand can help with pose recognition and analysis. One common approach for human skeleton estimation is a top-down method, which starts with a single person detection followed with a pose estimation. The pose estimations usually rely on local body parts [16] [17] for articulated model estimation. Although, this method has several drawbacks. Failure of person detection will result in pose estimation loss. Also, computational time is proportional to the number of people in each frame which may cost hours to finish processing a single video with crowds. On the other hand, another method has been proposed called bottom-up method. This method could potentially decouple runtime problem from number of people involved in each video. L. Pishchulin et al. proposed a method that labels candidates of detected body part and connect them to individuals using integer linear programming [18]. Based on this work, E. Insafutdinovv et al. further proposed a ResNet based body part detector, image-dependent piecewise score terms, and better optimization strategy [19] which speed up the overall performance to several minutes per frame. Recently, Z. Cao et al. suggested a Part Affinity Fields based pose estimation which can detect multi-person pose in real time [20]. The pose detection method used in this paper was based on that as presented in [20].

Skeleton model based tracking algorithms have also been discussed. E. Insafutdinov et al. developed a tracking algorithm that simplified body-part relationship graph and applied a feed-forward convolutional architecture to associate parts even in clutter [21]. In another study, an algorithm called PoseTrack has been proposed. A graph with both spatial and temporal edges for detection is built. Then it simultaneously associates body parts within each single frame and each person over different frames by integer linear programming [22]. These methods are designed for more general-purpose pose tracking and are still not time efficient for pedestrian and cyclist tracking and pose recognition.

III. Overall Method

The overall procedure for the proposed pose tracking and recognition method in this work, as shown in Figure 1, is composed of four (4) steps:

1. Cut frames from raw videos
2. Generate skeleton model based on generated frames
3. Apply tracking algorithm towards each pedestrian and cyclists along frames
4. Store data of each pedestrian and cyclist in separate files for pose recognition and intention analysis.

This method was tested based on raw videos from two databases. One is the TASI-110 naturalistic driving database developed by the Transportation Active Safety Institute (TASI) at IUPUI, and another one is Joint Attention for Autonomous Driving (JAAD) by York University [23]. TASI naturalistic driving database is a large scale naturalistic database collected by Transportation Active Safety Institute at Indiana University-Purdue University Indianapolis from 2012 to 2013. It focuses on understanding road user's behavior pattern during pre-crash as well as crash scenario [25]- [27]. The database includes driving data of 110 drivers with their cars running for one-year duration. A camera with speedometer and GPS coordinate was provided by TASI and mounted on the windshield behind the rear mirror of each vehicle for video recording. Drivers were recruited from the greater Indianapolis area in Indiana, USA. Most of the driving data are within 30 miles radius of Indianapolis area. As Indianapolis is a typical American city with various road conditions, most of data collected have a diverse sample of urban street, suburban, highway with a variety of light and weather conditions [26]. Thus, there are plenty of pedestrian and cyclist data with different scenarios that can be used for posture tracking and behavior recognition purposes. We randomly select several videos of pedestrians and cyclists from the database for validation. Besides TASI naturalistic database, JAAD database is chosen for our pose recognition and tracking algorithm analysis since they are labeled data with pedestrian and cyclist action recorded. They have 346 videos with most of them having pedestrians and cyclists involved.

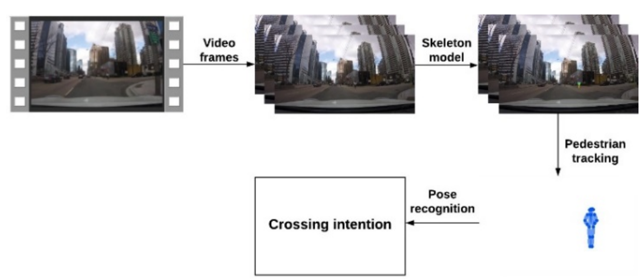

Fig. 1. Overall of pose tracking and recognition method

IV. Pose Estimation And Tracking Algorithm

A. Pose estimation

Our pose estimation is based on work of [20]. The skeleton model defined in this paper for pose estimation contains 18 points to present human body as shown in Figure 2. Points with number 1 and 2 represent head and neck respectively; the tracking algorithm in this study is based on coordinates of head and neck. Points with number 15 and 16 represent eyes of a human while points with number 17 and 18 present both ears of a person. Points 3 to 5 and 6 to 8 represent left and right arms and points 9 to 11 and 12 to 14 represent left and right legs. Every pedestrian or cyclist detected by this method is based on this skeleton model. More precise detection can better support behavior recognition, especially with eyes and ears detected.

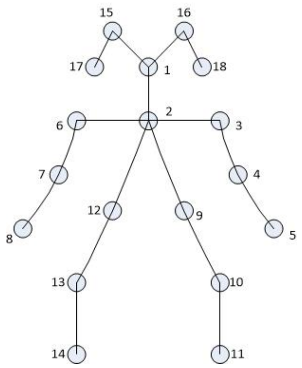

Fig. 2. Skeleton model for pose estimation

To estimate pose of pedestrian and cyclist, the heatmap of each body part candidate is generated by convolution neural network from each frame. This algorithm is using bottom up algorithm and no person is detected before parts connection. Body parts are then connected based on part affinity field. Redundant parts will be deleted from the graph. Body parts will be connected through predetermined sequence so that the skeleton model of pedestrian or cyclists can be detected. The computational speed of the pose estimation method is fast and both pedestrian and cyclist can be detected simultaneously in one frame, as shown in following Figure 3.

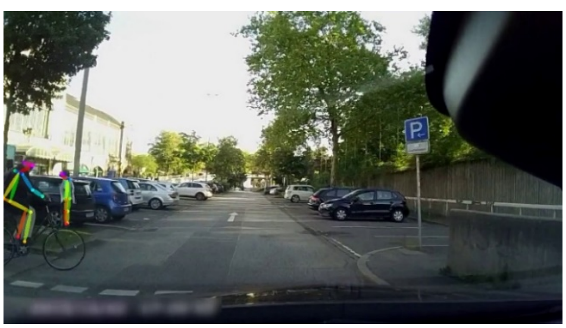

Fig. 3. Pedestrian and cyclists skeleton model for pose estimation

B. Pose tracking

After extracting skeleton model of pedestrians and cyclists from each frame, each detected person should be associated across frames so that their pose can be tracked and recognized for next step. Although existing tracking algorithms using skeleton model for human pose tracking from raw video, pedestrian and cyclist tracking could be handled with simpler algorithm since they are heading straight towards the same direction throughout the entire video in rather slow speed with less occlusion happened most of the time.

Our tracking method is based on the positions of head and neck from the skeleton model. One main reason to use these two points is because that pose estimation of pedestrians and cyclists is not robust enough for full-body pose tracking in some scenario. With light or weather condition varies, skeleton models of pedestrians and cyclists can be incomplete, which makes it difficult to associate people from different frames. However, head and neck can be continuously captured in majority of frames; they are more robust compared to other parts of body in skeleton model generated. Also, the movement of limbs for each pedestrian and cyclist is drastic compared to movement of head and neck thus they are hard to be chosen as reference for tracking. By only tracking the upper body of each person, we will be able to separate people from each other effectively and assign them with a unique ID. Movement of head and neck of single pedestrian can be shown as below.

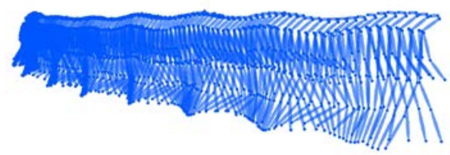

Fig. 4. Movement of head and neck of single pedestrian

The output from a pose estimation algorithm need to be cleaned before applying the tracking algorithm. There are two types of common errors in the skeleton model detection process: (1) random noises that do not belong to any pedestrian or cyclist skeleton models, and (2) false-positive human model detection (usually shown above horizontal line and hanging in the sky due to shadows in the tree or dark light conditions). To prevent fault data restored in tracking output, our algorithm neglect head and neck captured on the upper one third height of each frame and will only restore skeleton data corresponding to head and neck. Pseudo-code of data cleaning is expressed below.

if *head and neck detected* **then**
  // check if they are in top one third area of image;
  if y_{head} *or* $y_{neck} > image\ height*2/3$ **then**
    neglect data points;
  else
    restore data in candidate pool;
  end
end

Algorithm 1: Data cleaning for pose tracking

Since pedestrians and cyclists are moving within limited distances between adjacent frame, calculating distance of head and neck between each frame could be an effective way to distinguish one person from another. We record head and neck in certain structure of each person and compare the translational movements of their locations as a group crossing different frames. The data structure of each person's head and neck location is recorded as follow:

[person_ID, (x_head, y_head), (x_neck, y_neck), point_ID]

If x and y coordinates of head and neck of each person are within certain range, it will be classified as same person. The core tracking algorithm can be demonstrated by the following Pseudo-code.

```
for i ← 1 to candid_len do
    Head_neck_candid ← candid_pool[i][: 2] ;
    Head_neck_pre ← pre_pool[i][: 2] ;
    if abs[head_candid − head_pre] < γ then
        if abs[neck_candid − neck_pre] < β then
            ID_candid ← ID_pre;
        else
            swap candidate sequence;
        end
    else
        swap candidate sequence;
    end
end
```

Algorithm 2: pose tracking algorithm

After matching coordinates head and neck, data of each person are saved to separate files. All 18 points of skeleton data are stored along with frame ID and person ID in a row for each frame. If there is no data output from pose estimation algorithm, it will be recorded as -1.

It could be possible that multiple pedestrians and/or cyclists keep cutting in and/or cutting out in video. To track posture of each person precisely, we will need to augment the matrix when new people appear in video and stop recording skeleton data after they cut out. To augment the matrix, we compare the candidate number with maximum candidate number stored. If candidate number is larger than maximum number, we augmented the matrix by adding -1 to previous frames and record data starting from current frame. If one person disappears in one video, we will need to add a penalty value to make sure there will be no data recorded afterwards.

V. Behavior Recognition

The purpose of pose tracking is to extract posture information of each pedestrian and cyclist from scene video such that we can further analyze how pedestrians and cyclists behave after seeing a vehicle before or during their crossing. The goal of understanding these behaviors is to know how pedestrians and cyclists communicate with driver. To understand more detailed pose intention of a pedestrian and/or cyclist, we first need to extract critical pose from video and then analyze corresponding crossing intentions.

A. *Pose extraction*

Pose extraction is based on files recorded from pose tracking. For file of each person, we extract their points of body parts and reconnect them by the same sequence. In this way, we can separate pedestrian and cyclist pose from background. The results of the pose extraction can be shown as below.

Fig. 5. Pose extraction and recognition

B. *Crossing intention*

With pose data extracted from videos, the next step is to read pedestrian or bicyclist gestures and estimate intention. As automatic intention estimation is beyond the scope of this study, we adopt manual process to evaluate the crossing intention based on the extracted pose information. Relying on a previous study [23], the videos have been manually processed with different movement and gesture labels. Key frames of automatically tracked poses are extracted to compare with the manual labels for intention estimation. The main goal is to evaluate the feasibility to use pose detection and tracking data for studying road user intentions.

One example is described below. We extract pose of one pedestrian crossing the road while vehicle is turning right. When the vehicle is slowing down to let pedestrian cross, the pedestrian is waving hands to let the vehicle pass. There is an implicit waving action in one of the arm which is captured by our method, as illustrated in Figure 6.

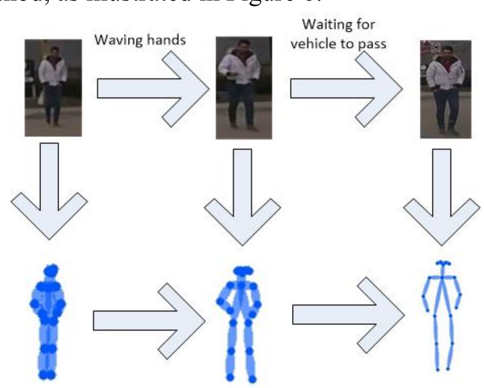

Fig. 6. Waving hand example of pose extraction and recognition

After detailed poses are extracted from each person along the video, their intention of crossing can be predicted based on their posture. It is quite helpful to know crossing intention of pedestrian/cyclist since it will make the control algorithms for autonomous vehicle more intelligent and less conservative comparing to safety oriented design concepts.

VI. Results and Analysis

We tested 346 videos from JAAD database [23] and some selective pedestrian videos from TASI naturalistic database in different scenarios [25]- [27]. We analyze data manually by separating videos into different categories: clear path, crossing, waving hand, looking, moving fast/slow, nod, slow down/speed up, stopped.

A. *Clear path*

Clearing path is the situation with a pedestrian or cyclist stepping away from the lane where both the pedestrian/cyclist and the vehicle occupied. It could happen when they are heading towards each other or going along the same way. In Figure 7 we see the clearing path action of a pedestrian or cyclist.

Fig. 7. Clearing path of pedestrian

Orange marked skeleton model represents the clearing path of pedestrian. It happened when the pedestrian saw the vehicle, and then walked towards the side lane to let the vehicle pass. The density of the postures can represent the change of relative moving speeds of the pedestrian from the vehicle perspective. The speed change can also represent the likelihood for the pedestrian to give up the right of way. By understanding the intention of pedestrians or cyclists, autonomous vehicle can be more intelligent and act more like a human driver by adjusting driving speed and directions correspondingly.

B. *Waving hands*

Waving hand(s) could happen when a pedestrian or cyclist would like to let vehicle go first, as shown in Figure 6. It could also happen when the pedestrian or cyclist would like to thank vehicle for letting them go first. This is a typical interaction between pedestrians /cyclists and drivers.

Figure 8 shows the cyclist wave hands towards the driver when the driver is slowing down to let cyclist go first. It happened during a very short time and it is implicit. By applying our algorithm, this waving posture is successfully captured and shown in the figure. Once vehicle could understand this from their movement, it will be able to communicate with pedestrian or cyclist better and improve overall traffic efficiency. Also, we can see the bicyclist reduced the moving speed before the hand-waving gestures, increased the speed during/after the gestures, and then reduced the moving speeds again (based on the density of the postures illustrated as the sample rate remains constant). This speed-change aligns well with the negotiation process, which shows good potentials to fully model the interaction process based on the pose detection and tracking output.

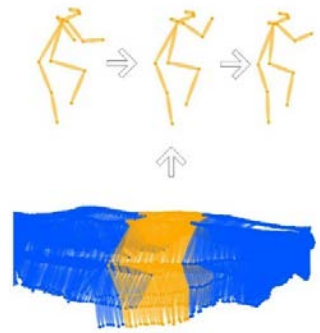

Fig. 8. Waving hands of cyclist to driver

C. *Looking*

Looking happens commonly when a pedestrian or cyclist is crossing the road. Understanding if the pedestrian or cyclist is looking at the car is very important for estimating their situational awareness and decision making. Some of pedestrians/cyclists will keep on checking if a vehicle is within safe distance between the pedestrians/cyclists and the vehicle when they are crossing or before crossing; some others will look at vehicle to make sure they already stopped at a stop sign so they can cross safely. Also, some of pedestrians or cyclists may not look at vehicles at all during their crossing, as shown in Figure 4.

Similarly, looking is also hard to capture from conventional full-body motion tracking methods. By defining the head model with five points, our pose tracking output could recognize eyes and ears of pedestrians and their movements, as illustrated in Figure 9. We can tell from the pose tracking results that the pedestrian is looking at the vehicle before crossing; he/she starts to cross the street after making sure the vehicle has slowed down its speed.

VII. Conclusion and Future Work

This paper has proposed a skeleton model based method for pose recognition and tracking from vehicle scene camera. Based on the pose estimation algorithm developed by authors in [20] and other post-processing methods such as data cleaning and pose tracking, the results by using proposed method have illustrated that it could be effective to extract

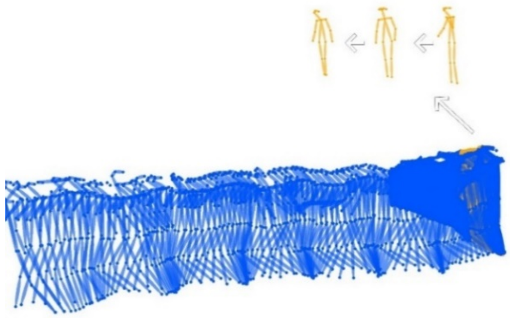

Fig. 9. Crossing pedestrian looking at driver

pedestrian and bicyclist poses from driving scene videos, and the skeleton model of each pedestrian/cyclist can be used for gesture recognition and intention estimation. This technique can provide more detailed information compared to traditional full-body motion tracking methods applied in current safety-oriented vehicle systems.

Further improvement of this method will be helpful to develop more intelligent control algorithms for future autonomous vehicles. Our current pose-tracking algorithm is designed for videos with less amount of people and less occlusion happened during crossing. One limitation of the top-down algorithm is the significantly increased runtime for larger number of people involved in one frame. Therefore, this kind of algorithms could be improved with attention weight added into consideration. The improved algorithms will only focus on crossing pedestrian or cyclists that could potentially be dangerous to driver and neglect those people only walking on sidewalk. Corresponding learning techniques need to be introduced.

Based on the results of this study, we have proved the feasibility to read pedestrian and bicyclist gestures and estimate the corresponding intentions for right-of-way negotiation. We are currently working on automatic intention estimation using both spatial and temporal pose tracking results.

REFERENCES

[1] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Proc. CVPR, vol. 1. Jun. 2005, pp. 886893.

[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 16271645, Sep. 2010.

[3] D.Comaniciu, V.Ramesh and P.Meer, Real-time tracking of Non-rigid Object Using Mean-shift, Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.142-149, 2000

[4] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, Globally-optimal greedy algorithms for tracking a variable number of objects, in Proc. CVPR, 2011, pp. 12011208

[5] A. Milan, S. Roth, and K. Schindler, Continuous energy minimization for multitarget tracking, IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 1, pp. 5872, Jan. 2014.

[6] Y. Xiang, A. Alahi and S. Savarese, ”Learning to Track: Online Multi-object Tracking by Decision Making,” 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 4705-4713.

[7] Yasuhide Hyodo, Shinya Yuasa, Kaichi Fujimura, Takeshi Naito and Shunsuke Kamijo, ”Pedestrian tracking through camera network for wide area surveillance,” 2008 IEEE International Conference on Systems, Man and Cybernetics, Singapore, 2008, pp. 656-661.

[8] P. Viola, M. J. Jones, and D. Snow, Detecting pedestrians using patterns of motion and appearance, in Proc. IEEE ICCV, 2003, pp. 734741.

[9] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Proc. IEEE CVPR, 2005, pp. 886893.

[10]P. Felzenszwalb, D. McAllester and D. Ramanan, ”A discriminatively trained, multiscale, deformable part model,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8.

[11]T. Li, X. Cao, and Y. Xu, An effective crossing cyclist detection on a moving vehicle, in Proc. IEEE WCICA, 2010, pp. 368372

[12]H. Cho, P. E. Rybski, and W. Zhang, Vision-based bicyclist detection and tracking for intelligent vehicles, in Proc. 4th IEEE, 2010, pp. 454461.

[13]S. Ren, K. He, R. Girshick and J. Sun, ”Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, June 1 2017.

[14]K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask R-CNN. arXiv:1703.06870, 2017

[15]X. Li et al., ”A Unified Framework for Concurrent Pedestrian and Cyclist Detection,” in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 2, pp. 269-281, Feb. 2017.

[16]S. E. Wei, V. Ramakrishna, T. Kanade and Y. Sheikh, ”Convolutional Pose Machines,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 4724-4732.

[17]F. Wang and Y. Li, ”Beyond Physical Connections: Tree Models in Human Pose Estimation,” 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, pp. 596-603.

[18]L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. An-driluka, P. Gehler, and B. Schiele. Deepcut: Joint subsetpartition and labeling for multi person pose estimation. InCVPR, 2016

[19]E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, andB. Schiele. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. InECCV, 2016.

[20]Z. Cao, T. Simon, S. E. Wei and Y. Sheikh, ”Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1302-1310.

[21]E. Insafutdinov et al., ”ArtTrack: Articulated Multi-Person Tracking in the Wild,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1293-1301.

[22]U. Iqbal, A. Milan and J. Gall, ”PoseTrack: Joint Multi-person Pose Estimation and Tracking,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4654-4663.

[23]A. Rasouli, I. Kotseruba and J. K. Tsotsos, ”Agreeing to cross: How drivers and pedestrians communicate,” 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, 2017, pp. 264-269.

[24]Z. Jiang and D. Q. Huynh, ”Multiple Pedestrian Tracking From Monocular Videos in an Interacting Multiple Model Framework,” in IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1361-1375, March 2018.

[25]R. Tian, L. Li, K. Yang, F. Jiang, Y. Chen, R. Sherony, “Single-Variable Scenario Analysis of Vehicle-Pedestrian Potential Crash Based on Video Analysis Results of Large-Scale Naturalistic Driving Data”, Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management: Ergonomics and Health, Volume 9185 of the series Lecture Notes in Computer Science, pp: 295-304, 2015.

[26]R. Tian, L. Li, K.Yang, S. Chien, Y. Chen, and R. Sherony, “Estimation of the Vehicle-Pedestrian Encounter/Conflict Risk on the Road based on TASI 110-Car Naturalistic Driving Data Collection”, In IEEE Intelligent Vehicles Symposium, June 8-11 2014, pp: 623 - 629, Dearborn, MI, USA.

[27]R. Sherony, R. Tian, S. Chien, L. Fu, et al., "Pedestrian/Bicyclist Limb Motion Analysis from 110-Car TASI Video Data for Autonomous Emergency Braking Testing Surrogate Development," SAE Int. J. Trans. Safety 4(1):113-120, 2016,