Tracking Vehicle and Faces: Towards Socialistic Assessment of Human Behaviour

Vaibhav Malviya
Robotics & Machine Intelligence Laboratory
Indian Institute of Information Technology
Allahabad, India
vaibsidea@gmail.com

Rahul Kala
Robotics & Machine Intelligence Laboratory
Indian Institute of Information Technology
Allahabad, India
rkala001@gmail.com

Citation: V. Malviya and R. Kala, "Tracking Vehicle and Faces: Towards Socialistic Assessment of Human Behaviour," 2018 Conference on Information and Communication Technology (CICT), Jabalpur, India, 2018, pp. 1-6.
Full Text Available At: https://ieeexplore.ieee.org/document/8722427

Abstract— The modern day robots can do a variety of tasks with great efficiency, however their utility is limited due to the non-social behaviour of the robots. For the same it is important to assess the human behaviour in diverse conditions to as to eventually make robots socialistic in nature. Object & people tracking is an excellent field of computer vision in which we have tried to detect and track multiple vehicles and people at outdoor traffic environment and indoor office environment respectively. A background subtraction algorithm is applied for vehicle detection. Kalman filter is used to predict the estimated position of every vehicle in the next frame and updating of new track. Some vehicles are also detected in cluttered scenes. Every moving vehicle is counted in video frames. Multiple face detection and tracking is an also attractive field of computer vision. The faces are behaviourally very different to vehicles and indoor scenarios are also very different to outdoor scenarios. Hence a different methodology to track people is used. In this paper, we have tried to detect and track face of multiple people on two different datasets with different height of camera. Point feature is extracted and compared it in the successive frames to track face of multiple persons. Every face is bounded by rectangular shape with unique identity. We have also counted total number of faces in the frame sequences. Results on both vehicle and face datasets show promising results and the proposed methodology can accurately track the trajectories. The output of the research is a good dataset to assess human behaviour to be used in social robotic applications of the future.

Keywords—object tracking; multi face tracking, Kalman filter, computer vision, human robot interaction;

I. Introduction

Human motion tracking is getting a lot of attention in the computer vision domain. Here detection of human behaviour is the most challenging issue in the domain of computer vision. Many researchers have proposed various algorithms to track human motion in order to understand their behaviour. It has a wide spectrum of application such as surveillance, ability to automatically monitor human activity, behavioural analysis, video conferencing, etc. These systems are applicable in security sensitive areas such as border, airports, lobbies of building, and area of military and police. Tracking of human motion can be performed in indoor and outdoor environment. Human motion tracking has become a very challenging task because-people get occluded by others, different lighting condition, complex environment, cluttered background, colour distribution of the target people and background colour model are different in dark and bright environment. At the initial stage, to track a human motion we can utilize a single and stationary camera based on colour information.

Object tracking is a salient task in the particular branch of computer vision. The field of computer vision includes the problems that involve interfacing computer with environment via a visual means. Due to incredibility of high-power computer and accessibility of high quality and inexpensive video cameras expand application of the computer vision. There are three major steps in the video analysis which are: Finding of interesting moving object, Tracking of object from one frame to another, and Analysis of object tracking

The modern day robots can perform a variety of tasks including going to a point of interest, getting to the master of the robot, picking and carrying items, etc. The modern day robots are equipped with vision, mapping, localization, planning and control modules that facilitate a flawless navigation of mobile robots. However, even after much advancements into robotics by decades of research, the actual motion of the robots is usually unacceptable by humans. The robots tend to navigate by objective based and experience based motion planning algorithms that makes them take trajectories that do not adhere to the social norms. As an example it may not be wise to overtake a human, even if that is computationally more efficient, the robot must maintain a distance from the person that is neither too large to look un-social, and not too small to invade the personal space, etc. In order to introduce the social behaviour into mechanical robots, it is first important to study the human behaviour and then to teach the same to the robots. The paper solves the first part of the problem only, and the specific sub-problem in the overall project to track human trajectories so as to be able to assess human behaviour.

The humans primarily operate in two very distinct types of environments. The first is highly structured environments consisting of queues, a fixed path of motion and rule based display of behaviours. The best example is the traffic scenario operating outdoors. The second is the highly unstructured environments wherein a human operates in wide open spaces with little rules. The best example is navigating in wide open office spaces. To get the best understanding of the diverse human behaviours the paper targets both the extreme cases of a structured outdoor traffic environment and an unstructured indoor office environment. The aspect of assessment of behaviour and transferring it to a robot is not dealt with in this paper and is a part of future work.

Tracking has been widely performed for different applications, primarily using Particle Filters or Extended Kalman Filters. Particle filtering method is investigated in the area of computer vision since it has ability to handle non-Gaussian and non-linear problem. Measurement model (likelihood function) is often non-linear due to noise or clutter in background [1], causing the posterior distribution of system state being non-linear. Effectiveness of particle filter is very good due to two factors, one is likelihood function which is used to extract visual information from images and another one is proposal distribution, which is a set of discrete sample that will be drawn. Li et al. [2] introduced a proposal for distribution based on Kalman filter that depends on a learned motion model.

The tracking performance can be significantly increased if the object to be tracked is known a priori, in which case object recognition techniques can be used to recognize the object and its pose to make the observer for tracking. People tracking system are able to recognize and find every person in frames of a video. Object recognition is the most challenging issue that is helpful for navigation of autonomous robot and scene understanding. It is a determining component for way finding and obstacle avoidance. Any detected object can be assumed as obstacle or waypoint to guide blind passengers to a destination.

Object recognition and tracking is also based on local feature extraction method, like Scale Invariant & Feature Transform (SIFT) and Speed Up Robust Features (SURF) algorithm to extract the feature points. SIFT is computed over local region, usually centred on feature points. The first stage of this algorithm is to search overall scales image locations. To identify potential feature point, a difference of Gaussian function is implemented which is invariant to orientation and scale. To find the location of each candidate, a detail model is fit. All the feature operations are performed on the image data which has been transformed relative to assign location, scale and orientation for each feature. At the last step a descriptor vector is found which is composed of histogram and computed from orientation and gradient magnitude of neighbour point in the window around every key point.

Herbert Bay et al. [3] proposed a SURF algorithm which is also based on SIFT algorithm but some steps are different. First step of SURF algorithm is the fixing of reproducible orientation which is based on information from a circular region around interest point. To extract a descriptor, construct a square region centered around interest point and oriented along orientation selected [3].

In this paper we intend to track people and therefore the sub-problem of object recognition is replaced with face recognition. Although various progress has been developed in the face recognition over the past decades. Face recognition application includes smart surveillance system where faces may be occluded by another face or object under the crowded scene, in these situation only partial faced can be obtained. CNN approaches [4], [5], [6], [7] use the whole face images for recognition and it is not applicable for partial face recognition. In [8], a TPGM (topology preserving graph matching) method has been proposed for partial face recognition and TPSM (topology preserving structural matching) to exploit higher order structural information.

II. RELATED WORK

Many tracking algorithms have been developed on the background subtraction using one or several stationary cameras. In these systems, tracking of people or object is performed which is based on extracted component belong to the foreground. In general, object tracking can be categories as: Active contour based tracking, Model based tracking, Feature based tracking, and Region based tracking.

In the active counter based tracking, a bounding contour is used to represent outline of the object that is updated dynamically in consecutive frames. For more extension, active contour can also be used as object descriptor to increase the efficiency and reduce the computational complexity. These algorithms can track object in a situation of partial occlusion. Developments of 2D or 3D model are required for tracking in the model based approach [9]. It is a robust method for motion tracking and gives better result in the case of occlusion. Kalman filter can also be applied for object tracking, it can predict the behaviour of moving object categorised by a predefined model [10]. However, in the real time we often face a situation where predefined model fall apart. Region based tracking is performed on the variation of image region in motion. In this feature extraction and whole object segmentation is not required. Instead, a person can be modelled as one or more small region like torso, head and four limbs. By using this strategy a human is tracked with small region. However this approach suffers more computational complexity because it matches a window with each candidate window in next frame. It is not reliable to handle occlusion between objects. In the feature based tracking, object features are extracted for matching in the order of frame. Here, some features of object are used in a feature vector, such as velocity, position, size, coordinates of bounding box, orientation, histogram of RGB colour component, etc. These feature vectors are compared with another method like correlation-based approach [11], Euclidean distance [12].

In the intelligent transportation system, vehicle tracking is modern research work in recent years. Since there is a very complex traffic environment, so due to this it is very difficult to track multiple vehicles robustly and accurately. To overcome this challenges two classifier and improved ViBe algorithm [13] have been used because ViBe algorithm is difficult to differentiate multiple moving vehicle. This algorithm removes the shadows and object's residual shadows and restrains the dynamic noise. This method has time efficiency advantage of SVM and accuracy advantage of the CNN.

Vehicle tracking has one more excellent application parking scheme. To develop this, VeTrack (real time vehicle tracking system) has been proposed [14]. This system utilizes the inertial sensors in smart phone that gives exact location of vehicle. VeTrack tracks a location of vehicle in real time and keeps its final parking location, and it is not depend on WiFi or GPS signals. Only inertial data are required to track by VeTrack, and all computing or sensing has happen locally on phone. To convert movements of smart phone to vehicle novel shadow trajectory tracking method has used.

A probabilistic algorithm [15] has also been developed to track a multiple vehicles in a wide range area which is based on vehicle model that includes the stuff of surrounding traffic environment. These stuffs are built into timing profile, velocity, and acceleration by using the historical data collected.

Face tracking is the most extensive work in the area of computer vision. Face recognition includes three main step face detection, face recognition, and face extraction. In a video sequence a face is detected by multiple method like Viola Jones Face Detection Algorithm [16], Local Binary Pattern (LBP) [17], AdaBoost Algorithm for Face Detection [18]. Viola Jones object detection strategy [16] is the first method for object detection in the real time. Even though to detect different class of object it can be trained. In general, this method is mainly used for face detection problem. In the real time it is most admired method. In this, constructing a cascade of classifiers which totally reduces computation time while improving the detection accuracy. But it has very long training time. Local Binary Pattern is most effective method to explore texture feature of image. In this method, every pixel has assigned a texture value, which is naturally combined with target for tracking of thermo graphic and monochromatic video. The major uniform LBP patterns are used to recognize key points in target region and then form a mask for joint colour-texture feature selection. It is also applicable for image segmentation, image retrieval, and texture analysis. It is not sufficient for illumination changes. It is applicable only for binary and grey images. AdaBoost Algorithm is an another approach based on idea of creating based on the idea of highly accurate prediction rule by combining many relatively weak and incorrect rules to machine learning. Here a classifier is trained which is capable to process images rapidly while it have very fast detection rates. This algorithm does not require any prior knowledge but it is very sensitive to noisy data.

In the recent development of computer vision application, human face tracking has become a hot research topic. Here face tracking is considered as optimization problem, so to resolve the problem of face tracking a meta-heuristic optimization algorithm and differential harmony search (DHS) has been introduced [19]. This algorithm is compared with the meanshift, particle filter, HIS [20] and it found that DHS based tracker gives more accurate result.

An approach has also applicable for face tracking which depends on threshold determination of positive face sequence. In this approach [21], every faces have different angle (initially) that are find and tracked, angle and position information of the face being noted. The next step of this method, distinct faces are matched to the face in frame. Here the positive faces that are very close to the person are detected. At the last step, result attained by threshold determination method. The experiment analysis of this method shows this method has good recognition in case of moderate flow density. This method is also capable for real-time system.

Aghaei et al. [22] have also presented a multi face tracking approach that generates a tracklets group and takes the advantage of tracklets redundancy to handle with unreliable ones. Here same tracklets are integrated into the bag of tracklets which aimed to correspond to particular person. This method is tested on large dataset of egocentric photo stream and found its robustness and efficiency compared to another method.

III. FACE TRACKING

The first aspect of the work is to track faces in the indoor environments. A difference from the standard literature is that only the faces and being tracked and not the complete human body. This is because in high congestion situations representing interesting behaviours, while most of the human body is occluded, the face of the person is reasonably clear. Further, we are only interested in the trajectories of the humans when walking and no other dynamic obstacles. Full body motion trackers can seldom track other dynamic obstacles like moving chairs, people moving while rolling on chairs, etc. For the specific application a face based tracking is the most robust technique of work. Similarly, the specific work can incorporate multi-face tracking, unlike many other popular approaches that are not scalable to multiple humans with part occlusions.

In the video sequence we have detected and tracked multiple faces using Kanade-Lukas-Tomasi (KLT) algorithm. A point tracker is used to extract point feature of the face. Face is bounded by a bounding box and a box contains only single face, every bounding box has its own identity. In first frame a face is detected and a rectangular box is drawn, then a point feature of face is matched in the remaining set of continuous frames. If any another face occurs in the frames then we have also detected and tracked, new bounding box associated with it and point feature is calculated for new face. We assign new or next box identity for this face.

Point feature for this face is also compared for continuity of face tracking in remaining set of successive frames. If the scope of face is over after some time then we delete bounding box and its identity. Whenever a new face is detected then we associate new bounding box and unique box identity. Here we determine point feature of the object, initial in a video frame first we predict the human face and then we track it in successive frame. Multiple faces are also predicted by this method, we draw a bounding box around the face and every box has its own identity (that is count of face). Every point feature has a point identity by which we track and match the face in video. An algorithm for face tracking is as follow:

I. Read a video file.
II. Detect the human face and find the point feature of the face.
III. Draw a bounding box around the face.
IV. Assign identity or box number to every bounding box so that we can predict number of faces in a video sequence.
V. Match the point feature vector in the successive frames of the video.
VI. If point feature match then track a face and identity of bounding box identity will remain same.
VII. If point features do not match then
    a. Scope of the face does not exist in the frames.
    b. If another face found, compute new point feature and track it.

IV. VEHICLE TRACKING

The second extreme case of tracking is encountered in tracking of moving vehicles. In conception it may appear that the same tracking methodology as indoor person tracking can be used in this case as well. However this is not true since the environment noise is much larger for outdoor scenarios as compared to indoor scenarios, the resolution of operation is much poorer in outdoor scenarios. Further, there is no distinct feature like face in a traffic scenario, each vehicle in a traffic scenario can vary largely by size and shape. This is specially true with the Indian context where vehicles can vary from a bicycle to a big bus. Hence a new technique is designed to combat these limitations. Correspondingly, the method proposed here is not ideal for tracking of faces as the method does not benefit from the generic heuristics of a well-placed indoor camera and the shape of the human face.

In this we detect moving vehicles in a video sequence and also associate detection corresponding to same object over the time. Here we have applied background subtraction method in which first frame considered as a background frame and remaining frame as foreground. Initially moving object or person is detected with a noise. So we have applied morphological operation such as opening, closing and filling to remove the noise in the resultant foreground mask. For every moving object, blob analysis is applied to detect the set of connected pixel.

For the updating of new track Kalman filter is applied. Kalman filter is used to estimate the motion of every track and find the track location in every frame. If new object or person is detected in a video sequence the new track is created and scope of this track exist whenever this object or person exist in the video, simultaneously if new object or person are detected then another new track is created. Every track has its own unique identity. Moving person or object is not available after some frames then we delete old track.

To track moving object, we have created rectangular shape as a bounding box. Every box has its own identity. We have also count total number of moving object in video sequence with box identity. Algorithm for object tracking is as following:

I. Read video frame.
II. Consider first frame as a background frame.
III. Find location of the object in every frame using kalman filter.
IV. Draw a bounding box around the moving object and associate box identity.
V. If scope of object does not exist then delete the tracks and bounding box.
VI. If new object has been found in the middle of video then add new track and draw bounding box around the object and associate box identity.

V. DATASET

To implement this method we have recorded two distinct datasets. For vehicle tracking we have recorded traffic road video from outside IIIT Allahabad by webcam with fifth floor height. The placement of the camera is strategically done. If the camera is located at smaller heights, the resolution is very good for accurate tracking, however the special coverage is highly compromised. Similarly if the camera is too high, the special coverage is very high, however the resolution is compromised. The intent is to get the best bird's eye view of the vehicles in order to be able to easily identify behaviours. Once of the videos presented for testing features a person and vehicle both moving and we have tried to track both. The duration of this video is four minutes and two seconds.

We have also record video inside the Robotics and Machine Intelligence Laboratory at IIIT Allahabad for face tracking. The methodology here is inverted from an attempt of getting high special coverage bird's eye view to getting a low coverage view with a high resolution. There are two reasons for this. First, that a bird's eye view gets the head of the person that has lesser features and hence more prone to errors. Such a system was also not pre-installed and installation at enough heights got a lot of obstruction from a variety of objects and hence was not a feasible option, unlike traffic where aerial view has less occlusions. Second, eventually the camera will be mounted on a robot that can self-localize itself. Hence the relative motion of people with robot can be easily studied. This means that special coverage can be increased by making the robot move with time. This also facilitated future direction wherein the camera will be mounted on the robot for socialistic navigation. With the same future aim in mind, the camera was taken at the height of the robot. The camera is mounted on top of the Pioneer LX robot, which is the robot that will eventually be used for the socialistic navigation.

Two of the videos are explained. In the first clip two persons are moving with each other like friends moving into the lab. In another video four people are moving as one group. The length of the first video is 17 seconds and the second video is 48 seconds. In these dataset we have tried to track the face of all the people.

VI. RESULT

First, the results of tracking of vehicles are presented. We have applied background subtraction algorithm to track object and for updating of new track kalman filter has applied. An object has its own identity and bounding box. Due to space constrains, only one of the test cases is discussed that represents enough number of moving vehicles along with some static pedestrians. The results are shown in Fig. 1 (a).

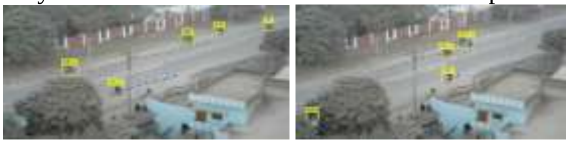
(a) (b)

Fig 1:(a). Tracked person (b). Tracked person in cluttered environment

Here we have detected multiple vehicles at the road, some object are also tracked in the cluttered environment as shown in Fig 1 (b). We have not detected nonmoving pedestrians as shown in Fig 1 (a) and Fig 1 (b). They are static person and not walking and standing whole time in the complete video sequences.

For the face detection we have applied Kanade-Lucas-Tomasi (KLT) algorithm. We have applied this algorithm on our dataset. Again, only two videos are discussed that have small and large number of people moving. In the first video two people are moving. The height of the camera from floor is 1.2 meter, distance between person and camera is 5.4 meter and we have detected face as shown in Fig 2 (a). After some time distance between person and camera is 2.1 meter and face is detected continuously as shown in Fig 2 (b). We have applied same algorithm on another dataset in which four persons are moving and we have detected all faces as represented in Fig 3. In this dataset, camera is mounted on head of the pioneer robot.

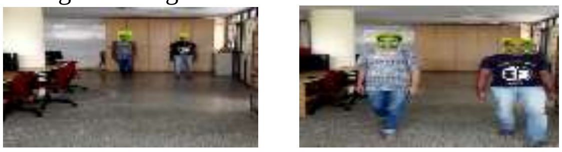
(a) (b)

Fig 2: Tracked face of moving two people

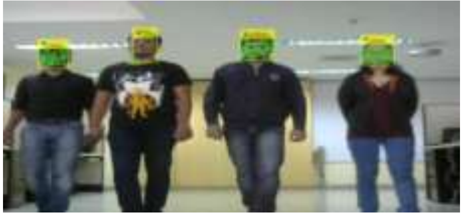
Fig 3: Tracked face

VII. CONCLUSION

In this paper we performed vehicle tracking on traffic environment. Face detection and tracking is also performed on indoor environments. We applied background subtraction algorithm to track vehicles in traffic environments. In this background subtraction algorithm position of every vehicle is estimated by using a kalman filter. Every object is bounded by rectangular box with unique identity. Here some object is also detected in cluttered environment. We have counted total number of moving object in video sequence. At another phase of this paper, we performed multiple face detection and tracking. To detect and track multiple faces a point feature of face is extracted and compared it in the remaining frames. Every face is bounded by rectangular box with unique identity. So we can also count total number of faces in the video sequence. Here face detection and its tracking is performed at indoor environment.

The experiments are done over numerous videos ranging from spars to high density in both indoor and outdoor environments. Experimental results suggest that the algorithm can accurately track people and vehicles, while the individual people and vehicles may be operative representing diverse behaviours and dynamics.

The paper presents a tool to track people and vehicles in both highly structured traffic scenarios as well as in highly unstructured indoor scenarios. Visually, there were eminent commonalities and typicality in both the two extreme scenarios. The task of learning and classification of human behaviour and the variance of behaviour in different contexts, scenarios and with a difference of people needs to be studied. Further, the assessment of such trajectories to get social etiquettes and to thereafter inject the same in the motion behaviour of the robot is the open topic of research for the future. The vision and tracking is one of the major aspects of the project and results suggest that the algorithm developed can be used for tracking people and vehicles so as to enable the robots learn from the human masters.

REFERENCE

[1] Weng, Shiuh-Ku, Chung-Ming Kuo, and Shu-Kang Tu. "Video object tracking using adaptive Kalman filter." *Journal of Visual Communication and Image Representation* 17.6 (2006): 1190-1208.

[2] Li, Stan Z., and ZhenQiu Zhang. "Floatboost learning and statistical face detection." *IEEE Transactions on pattern analysis and machine intelligence* 26.9 (2004): 1112-1123.

[3] Bay, Herbert, et al. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110.3 (2008): 346-359.

[4] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." *BMVC*. Vol. 1. No. 3. 2015.

[5] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[6] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

[7] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

[8] Duan, Yueqi, et al. "Topology Preserving Structural Matching for Automatic Partial Face Recognition." *IEEE Transactions on Information Forensics and Security* (2018).

[9] Lou, Jianguang, et al. "3-D model-based vehicle tracking." *IEEE Transactions on image processing* 14.10 (2005): 1561-1569.

[10] Chen, Xi, Xiao Wang, and Jianhua Xuan. "Tracking Multiple Moving Objects Using Unscented Kalman Filtering Techniques." *arXiv preprint arXiv:1802.01235* (2018).

[11] Haritaoglu, I.; Harwood, D.; Davis, L. S.; “Hydra: multiple people detection and tracking using silhouettes”, International Conference on Image Analysis and Processing, 27-29 Sept. Page(s):280 – 285

[12] Xu, L.; Landabaso, J. L.; Lei, B.; “Segmentation and tracking of multiple moving objects for intelligent video analysis”, BT Technology Journal, Vol 22, No 3, July 2004.

[13] Min, Weidong, et al. "A New Approach to Track Multiple Vehicles With the Combination of Robust Detection and Two Classifiers." *IEEE Transactions on Intelligent Transportation Systems* 19.1 (2018): 174-186.

[14] Gao, Ruipeng, et al. "Smartphone-Based Real-Time Vehicle Tracking in Indoor Parking Structures." *Smartphone-Based Indoor Map Construction*. Springer, Singapore, 2018. 81-109.

[15] Shan, Mao, et al. "Using delayed observations for long-term vehicle tracking in large environments." *IEEE Transactions on Intelligent Transportation Systems* 15.3 (2014): 967-981.

[16] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, 2001.

[17] Ahonen, Timo, Abdenour Hadid, and Matti Pietikäinen. "Face recognition with local binary patterns." *European conference on computer vision*. Springer, Berlin, Heidelberg, 2004.

[18] Meir, Ron, and Gunnar Rätsch. "An introduction to boosting and leveraging." *Advanced lectures on machine learning*. Springer, Berlin, Heidelberg, 2003. 118-183.

[19] Gao, Ming-Liang, et al. "Face tracking based on differential harmony search." *IET Computer Vision* 9.1 (2014): 98-109.

[20] Fourie, Jaco, Steven Mills, and Richard Green. "Harmony filter: a robust visual tracking system using the improved harmony search algorithm." *Image and Vision Computing*28.12 (2010): 1702-1716.

[21] Shi, Xin, Jian Wu, Xin Ling, Qiao-li Zheng, Xuan-qi Pan, and Zhi-qiang Zhao. "Real-Time Face Recognition Method Based on the Threshold Determination of the Positive Face Sequence." In Proceedings of the 22nd International Conference on Industrial Engineering and Engineering Management 2015, pp. 125-136. Atlantis Press, 2016.

[22] Aghaei, Maedeh, Mariella Dimiccoli, and Petia Radeva. "Multi-face tracking by extended bag-of-tracklets in egocentric photostreams." Computer Vision and Image Understanding 149 (2016): 146-156.