Real Time Body Orientation Recognition for Customer Pose Orientation

Nethravathi P. S. [1] & P. S. Aithal [2]

[1] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore
Orcid ID: 0000-0001-5447-8673; Email ID: nethrakumar590@gmail.com

[2] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore
Orcid ID: 0000-0002-4691-8736; Email ID: psaithal@gmail.com

Area/Section: Computer Science.
Type of the Paper: Research Article.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed in: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.6556192
Google Scholar Citation: IJMTS

How to Cite this Paper:
Nethravathi, P. S., & Aithal, P. S., (2022). Real Time Body Orientation Recognition for Customer Pose Orientation. *International Journal of Management, Technology, and Social Sciences (IJMTS), 7*(1), 390-399. DOI: https://doi.org/10.5281/zenodo.6556192

International Journal of Management, Technology, and Social Sciences (IJMTS)
A Refereed International Journal of Srinivas University, India.

CrossRef DOI: https://doi.org/10.47992/IJMTS.2581.6012.0197

Received on: 26/03/2022
Published on: 18/05/2022

Real Time Body Orientation Recognition for Customer Pose Orientation

Nethravathi P. S. [1] & P. S. Aithal [2]

[1] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore
Orcid ID: 0000-0001-5447-8673; Email ID: nethrakumar590@gmail.com

[2] Professor, Institute of Computer and Information Sciences, Srinivas University, Mangalore
Orcid ID: 0000-0002-4691-8736; Email ID: psaithal@gmail.com

ABSTRACT

Background/Purpose: *One of the most significant areas in marketing is consumer position analysis. Retailers can assess the extent of client interest in the goods based on customer pose data. Due to occlusion and left-right similarity difficulties, pose estimation is problematic. We describe a CNN-based solution that includes the body orientation and visibility mask to overcome these two challenges. It provides global information about posture configuration using simple gaits in a retail setting. When a person looks to the right, for example, the left side of his or her body is hidden by the body orientation. In the same way, the person faces the camera, the right shoulder will most likely be on the image's left side. A novel Deep Neural Network design is used to merge body orientation and local joint connections. Second, the visibility mask simulates each joint's occlusion state. Because body orientation is the major source of self-occlusion, it is inextricably tied to it. Detecting an occluding object (such as a shopping cart in a retail setting) might provide give visibility mask prediction clues. Global body position, local joint connections, client mobility, and occluding obstructions are all taken into account in the final advised method. Finally, we run a number of comparison tests to see how effective our technique is.*

Objective: *This work presents customer posture estimation, and a visibility mask to build a prototype for inner and self-occlusion. It also concentrates on local joint connections, global body orientation, and customer mobility.*

Methodology: *The suggested technique is depicted in its entirety in Figure-2. To figure out which portion of the human picture is concealed, we employ stance markers to identify the viewable areas. The landmarks in the occlusion zone have a lower confidence score when we extract posture landmarks. As a result, it is possible to obtain visible masks incorporating occlusion information. We employ visible signs to assist the hidden individual in three ways. To begin, visible masks are utilized to detect viewable parts and to construct spatial masks that filter noise caused by occlusions.*

Findings/Results: *The proposed method outperforms to overcome left-right similarity difficulties, the network incorporates body orientation information, and the visibility mark layers are introduced into the network to enhance the efficiency of occluded joints.*

Conclusion: *For customer pose estimation, A novel architecture using the concepts of Deep Learning is proposed and the occluding object detection clearly provides inter-occlusion by object cues. As a result, local joint connection, the global body orientation, and occluding object and human motion.*

Paper Type: *Research article.*

Keywords: Customer monitoring, Deep neural network, Pose estimation, Surveillance camera, Visibility mask, Convolutional neural network, Body orientation

1. INTRODUCTION :

Customer posture analysis is critical for retail businesses on a commercial level. Traditionally, retailers have analysed customer shopping behaviour using cash register or credit card records. The black box, on the other hand, still contains the customer's shopping experience at the store. Client positions such as stretching an arm, standing or leaning over may provide some information about the consumer's behaviour and level of interest in the product in practise. It is essential to estimate the customer postures from the security camera. As a result, scholars are growing more interested in using surveillance cameras to estimate client posture.

Another challenge is self-occlusion and inter-occlusion by other objects. If a person stands to the right of the camera, the left side of his or her body will most likely be hidden. Inter-occlusions, on the other hand, can be caused by a variety of factors. In a retail establishment, shopping baskets are the most typical source of inter-occlusions. The problem of left-right comparation, which comes from the proportion of the human body, is another stumbling block in pose estimation. A person's left shoulder, for example, seems to be almost similar to right one when viewed from behind.

Body orientation is a type of feature which enhances the knowledge about a person's overall postural arrangement. The eight distinct ways the body can be orientated are depicted in Figure 1. We may deduce that if a person is facing right, the body orientation is obstructing his or her left body. We may determine that if a person is facing the camera, his or her right shoulder is on the image's left side. We propose a unique deep neural network design to overcome this.

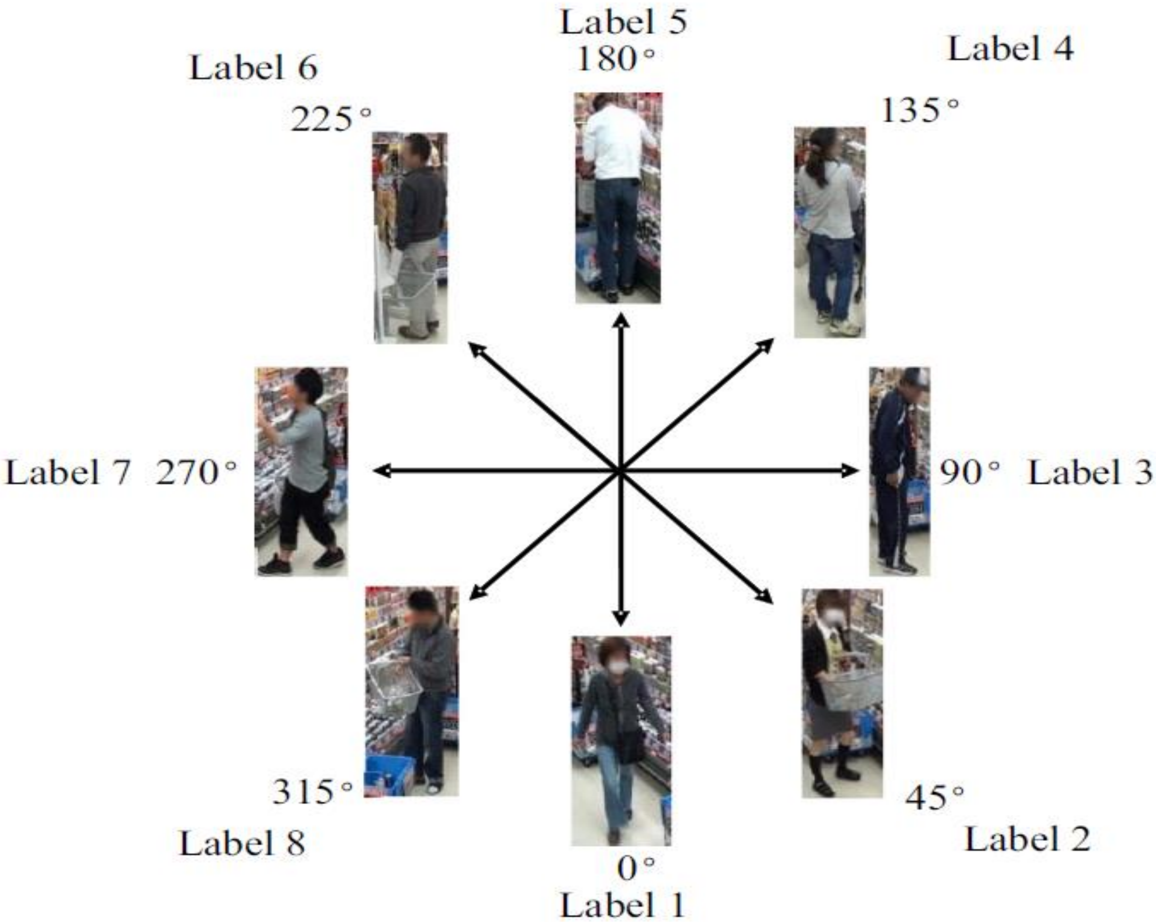

Fig. 1: The orientation of the body which is divided into 8 different direction and the change is for every 45^0 (Compiled by the authors)

Each element of the visibility mask indicates whether or not each joint is visible. The word "visibility mask" was invented by Haque et al. Only in the top view images did they utilize the visibility mask for self-occlusions. More adjustable view angles, as well as self-occlusion and object inter-occlusion, are now available when using the visibility mask. In other cases, as mentioned in a binary vector was used to express the occlusion. Regardless of how different systems represented occlusions, none of them took into account the link between joint occlusion and body position. Body orientation, we believe, is an

important part of the occlusion model. Our proposed technique may additionally pass diverse occlusion information according to different body orientations, in addition to solving the left-right similarity problem. Furthermore, in our system, we incorporate the detection of occluding objects. For visibility mask prediction, it appears to be an inter-occlusion clue. According to our observations, the occluding objects are primarily shopping baskets at retail outlets. (The main occluding part of the shopping cart is also the basket on the cart.)
Three elements can be said about our work's contributions. To show self-occlusion and tackle the left-right similarity problem,

- First to introduce body orientation in customer posture estimate. Novel deep learning architecture combines this with local pair-wise joint connection.
- Second, we use the visibility mask to model both inter and self-occlusion. Body orientation is also linked to the visibility mask.
- Third, local joint connections, global body orientation, customer mobility, and occluding objects are all taken into account in this approach.

The next parts, which are organized as follows, give further information. The literature evaluation of existing latest and robust methods that are relevant to the proposed methodology is summarized in Section II. Section III goes through the recommended strategy in detail. Section IV contains the results, analyses, disputes, and conclusions. Section V concludes the paper by laying out some of the project's future directions.

2. OBJECTIVES OF THE STUDY :

(1) To introduce body orientation in customer posture estimate. Novel deep learning architecture combines this with local pair-wise joint connection.
(2) To check the use of the visibility mask to model both inter and self-occlusion. Body orientation is also linked to the visibility mask.
(3) To local joint connections, global body orientation, customer mobility, and occluding objects are all taken into account in this approach.

3. LITERATURE SURVEY :

Researchers developed a variety of methods for estimating human pose based on various data. One of these methods attempted to estimate the pose using global silhouette information [1][18]. Other studies looked at individual human body elements to determine the overall stance. Pictorial Structures [2], for example, recognised where biological components were located. To mimic the interactions between surrounding joints, Deformable Part Models (DPM) were utilised [3] [20]. Pishchulin et al. [4] creates a better and robust model. The Histogram and other custom attributes were used in all of these solutions [5].
Convolutional Neural Networks (CNN) extensively used in number of computer vision applications and the performance is better than any machine learning methods [6]. Based on micro-videos, CNNs have been utilised in application industries to estimate popularity [7] and venue category. Toshev et al. [8] reported a novel study in the pose estimation task. They proposed the DeepPose, which uses several cascaded CNNs to estimate human pose. The methods for estimating pose with CNN may be loosely classified into two groups. The joint positions were immediately output by the first group [9]. However, because there is only one prediction per image, this type of technique was difficult to optimize. To enhance the number of candidates, the second group of approaches created probabilistic heatmaps of joints.
Furthermore, flexible optimizations, such as the Deformable Parts Model (DPM) [10], might be applied based on the joint heatmaps. Many DPM variations of the multiple stream Message Passing (MP) methodology on top of an FCN were suggested in [11] [19]. The MP technique from [11] is used in this study to apply the recommended model, that integrates the pairwise joint model with orientation of the body. Learning the occlusion structure for the obstructed joints was one technique to deal with the occlusion problem [12]. A binary mask was also used to mimic self-sealing [13] and object inter-occlusion [14]. Despite the fact that coined the term "visibility mask," it was only used in the paper to describe the occlusion component from a top view photo.

4. RESEARCH GAP :

Predicting the customer behaviour and understanding them to improve business is necessary to determine the customer needs and wants in framing market strategies. This understanding and behaviour are to be analysed on the basis of the information given by the customer. The personalized, individualized, and relevant information of the customers are required for business intelligence appraisal. Study on online customer behaviour and their purchase pattern has been experimented under various context over these years. There are only few studies directly investigated the relation between psychology and online customer behaviour. Humans share emotions which are exhibited through facial expressions and their movements. Detection, extraction, and evaluation of the expressions will allow for automatic recognition of human emotion in images. Hence this work analyses the psychological factors of the customer based on Emotional Recognition Technique. There hasn't been a work published that models the relationship between occlusion, body orientation, and consumer happiness

This work contains occluding object identification for modelling inter-occlusion, as well as customer satisfaction classification, in addition to body orientation for modelling self-occlusion. A CNN-based architecture model is suggested for analyzing client pose orientation in real-time using body orientation recognition.

5. METHODOLOGY :

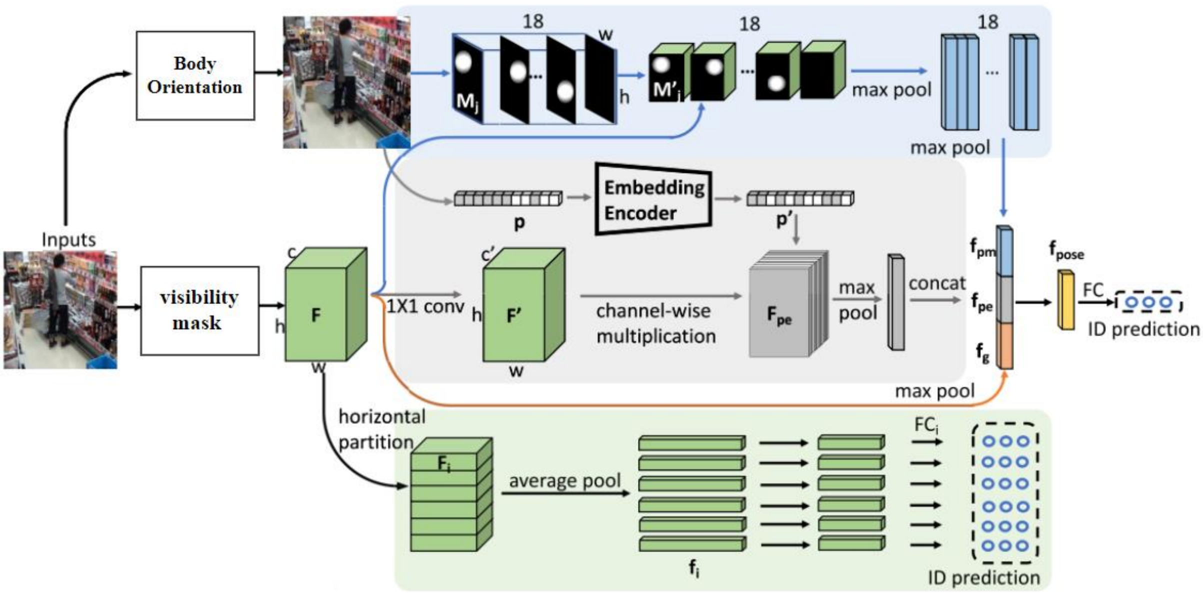

Fig. 2: The complete Flow of the proposed methodology (Compiled by the authors)

The suggested technique is depicted in its entirety in Figure 2. In fining the portion of the human picture is concealed, stance markers utilized to identify the viewable areas. The landmarks in the occlusion zone got low confidence score when we extract posture landmarks. As a result, it is possible to obtain visible masks incorporating occlusion information. We employ visible signs to assist the hidden individual in three ways. To begin, visible masks are utilized to detect viewable parts and to construct spatial masks that filter noise caused by occlusions. This is where the pose orientation characteristic is found. Second, visible landmarks include occlusion information, which indicates which sections are obscured and which are visible. As a consequence, the visible mask is used to build a posture embedding. Pose embeddings efficiently used to adjust channel-wise feature responses, choosing visible channel features while suppressing non-visible ones. The pose-embedded feature branch is what it's called. The visible masks, on the other hand, reveal which regions of the image are masked during testing. The query and gallery photographs are compared using the most common visual component features, which are organized into parts on the global feature map.

ResNet50 [15], which, like previous state-of-the-art systems [16], excludes the final average pooling and completely related layers, forms the backbone of our method. When given a human picture I with a size of H*W, the original ResNet50 generates a feature map with a spatial dimension of H/32W/32. We increase the spatial dimension of the obtained feature map to H/16W/16 and alter the stride of conv4 1 to 1 to extract a more relevant feature map [16]. A broader spatial feature map makes it simpler to distinguish the targeted individual. The recovered feature map is represented by the letter F.

We generate spatial masks using the position information of visible landmarks in the pose-masked feature branch. S_j^{conf} and lambda, respectively, signify the confidence score and the threshold, denote the positions of the posture markers. P_j denotes the position of the j^{th} landmark, whereas cx_j and cy_j give the coordinates of the j^{th} landmark, j = 1,..., N. We may get visible landmarks using spatial position information P. For creating the posture embedding that comprises the occlusion information in the pose-embedded feature branch, a visible landmark vector p belongs to $\{0, 1\}^N$ is produced.

$$\mathbf{P}_j = \begin{cases} (cx_j, cy_j) & \text{if } S_j^{\text{conf}} \geq \gamma \\ 0 & \text{else} \end{cases} \quad (j = 1, \ldots, N) \qquad \text{-----------} \quad (1)$$

S_j^{conf} and the threshold, respectively, signify the confidence score along with the threshold. The visible landmark vector p has a p_j element that specifies whether or not the j^{th} landmark is hidden. As a consequence, the visible landmark vector p encodes the occlusion and creates posture embedding.

$$\mathbf{p}_j = \begin{cases} 1 & \text{if } S_j^{\text{conf}} \geq \gamma \\ 0 & \text{else} \end{cases} \qquad \text{----------} \quad (2)$$

Locations of visible landmarks P are employed to create spatial masks. $P_j = (cx_j, cy_j)$ is the resultant posture mask, which is a Gaussian heatmap with the centre at the observable pose markers (cx_j, cy_j). With $P_j = 0$, the spatial posture will be set to zero for the unseen landmarks. Each posture mask is denoted by M_j,j = 1,..., N. The posture masks M are downscaled to the spatial size of F via bilinear interpolation. To generate N feature maps M_j,j = 1,..., N, each pose mask M_j is multiplied by the feature map F. The pose-masked feature maps reduce the occlusion information by focusing on the observable body components and occlusion portions.

The observable feature matrix p is utilized to create the posture embedding in the pose-embedded characteristic branch. In the global feature, the posture encoding is employed as channel gates. It's difficult to encode the pose embedding appropriately because p has a tiny dimension while the global feature map F has a huge dimension (2048). As a result, we use a one-layer convolution layers to reduce F's channel dimension and construct a new convolutional feature F with a developmental continuum of 1024. To create the posture embedding p, we utilize two totally linked regions and a sigmoid activation layer finally. By multiplying F channel-wise and utilizing the pose encoding p as channel gates, the body position feature map F_{pe} is generated. The pose-embedded feature f_{pe} is built using a max-pooling layer. The occlusion information is encoded implicitly via the pose-embedded feature f_{pe}. PCB [16] is similar to our limited feature branch. In contrast to PCB, our system examines visual mask features in frequently visited places and incorporates occlusion information via observable landmarks. The feature comparison in the visible limited area reduces the occlusion component in the occlusion picture.

6. RESULTS AND ANALYSIS :

We used two types of datasets in the studies. The customers pose dataset (Items of Customer's Interest (ICI) dataset), which is supplemented by the Leeds Sports Pose (LSP) dataset [17]. The ICI and LSP datasets are used to supplement to increase the robustness of the posture data's human variability. The proposed deep learning architecture was built using the Keras library with Tensor Flow as the back-end. The LSP dataset was considered for both the training and testing of the network [14].

The LSP dataset is a well-known standard for human posture estimation. They're employed to broaden the customer dataset's diversity. 1000 persons were used for training and 1000 people were used for testing in the LSP dataset. Each picture in LSP has 14 joint locations determined and a visibility mask, as mentioned above. The LSP dataset, on the other hand, contains no orientation labels. In fact, this collection contains a large number of complicated sports photos that cannot be simply oriented. As a result, we manually select a subset from LSP. The body orientations of these subgroups can be classified into one of eight ways. It means that the feet are usually at the bottom and the heads are usually at the top of the photos in the selected subsets, similar to how customers are placed in a store. Poses with the following conditions in the considered dataset are not permitted:

(1) The heads are lower than the waists; (2) The feet are higher than the waists; (3) The entire body is horizontal. This is referred to as Orientational LSP in this case (OLSP). The backgrounds of OLSP are removed using FCN-8 s person segmentation [28]. The customer place is then dilated by a circular construction with a radius of 25 pixels. Figure 1 shows some OLSP background subtraction examples.
Customer behaviour in a real-life buying setting must be examined in order to detect customer interest. There is a shopping situation as well as typical situations such as walking, looking around stuff, and talking to each other in a real-life store. However, we need settings that represent people and products in order to accurately study customer behaviours when purchasing. We used the ICI (item of customer interest) dataset for this, which included surveillance recordings of various customer actions such as pausing in front of a product and gazing at it. The videos in the dataset were found on YouTube, where they searched in multiple languages to increase the variety and quantity of videos. A total of 72 videos were picked by average span of 9.7 seconds. The videos have a fixed frame rate of 20 fps and a resolution of 720 x 480 pixels, respectively. The recordings taken by real-world CCTV surveillance cameras, with varying perspectives, lighting circumstances, and resolution quality, make up the dataset.
The Percentage of Correct Key points (PCK) is a commonly used indicator for evaluating posture estimate. The anticipated joint that is closer to the true joint than a specific distance is the appropriate key point. On the other hand, our technique not only generates joint locations but also prominence masks. To assess the outcomes of both the visibility mask and joint locations, we used the Visibility sensitive Percentage of Correct Keypoints (VPC) [3]. The following is how the VPCK is defined: The number of invisible joints is P_{inv}, the number of visible joints is P_v, and the number of visible joints is PC_v. $Accuracy_{inv}$ is the rate at which an unseen joint may be detected. The visibility threshold for all of the joints is set to 0.5 since we let the 0 be completely occluded and the 1 be completely visible.

$$VP\ C= (PC_v * P_v) + (Accuracy_{inv} * P_{inv}) / (P_v + P_{inv}) \qquad \text{------} \quad (3)$$

Table 1: Results Performed in the Entire Dataset Using the VPC (%) (Compiled by the authors)

Method	Overall	Head	Shoulder	Elbow	Waist	Hip	Knee	Ankle
HOG + MP	33.01	42.11	36.23	33.21	28.23	21.24	38.87	31.22
Inception V4	47.18	46.23	44.84	39.23	51.44	52.29	55.12	41.11
Xception	75.63	82.29	80.12	78.21	71.23	77.23	71.23	69.12
Cascaded CNN	81.31	84.11	86.23	89.11	76.11	81.23	74.21	78.23
ResNet	86.19	94.23	94.11	81.23	83.10	87.23	81.11	82.38
Proposed Method	90.20	99.23	94.23	85.12	83.18	88.11	89.23	92.32

The above-mentioned Table-4 gives testing values after applying the dataset. The total outcome curves at various thresholds are depicted in Figure 4. The visibility mask is not formed by the technique in experiments (1) and (2). (2). As a result, the values for experiments (1) and (2) in the table are the standard PCK values. In Table 1, the VPCK (or PCK) threshold is 0.09 of body height. Figure 3 depicts the comparison instances.
These findings reveal a number of intriguing facts.

(1) To reach the best results, the offered strategies aggregate the most data.
(2) Handcrafted features methods perform worse than deep learning methods.
(3) To improve the VPCK in general (or PCK), orientation information is playing a vital role.

ResNet, on the other hand, has just a slight improvement in orientation. Because the ResNet can only make one prediction per image, this is the case. If no other possibilities are available, the orientation data is unlikely to correct the forecast. (4) Orientational message passing layers, such as rows 1 and 4 in Fig. 3, can help with the left-right similarity problem. The right (blue/cyan) and left (red/purple) limbs are rectified from the other side. (5) As seen in the data in Fig. 3 at row 4, columns 1 and 4, the basket heatmap can aid in the identification of occluded joints. (6) The optical flow can help estimate moving joints like the wrists in rows 1 and 3, columns 1 and 4, in Fig. 3. Incorporating the optical flow or basket heatmap into ResNet techniques, on the other hand, is difficult. For each frame the results are obtained from the methodology in 0.018 seconds making it a real-time application.

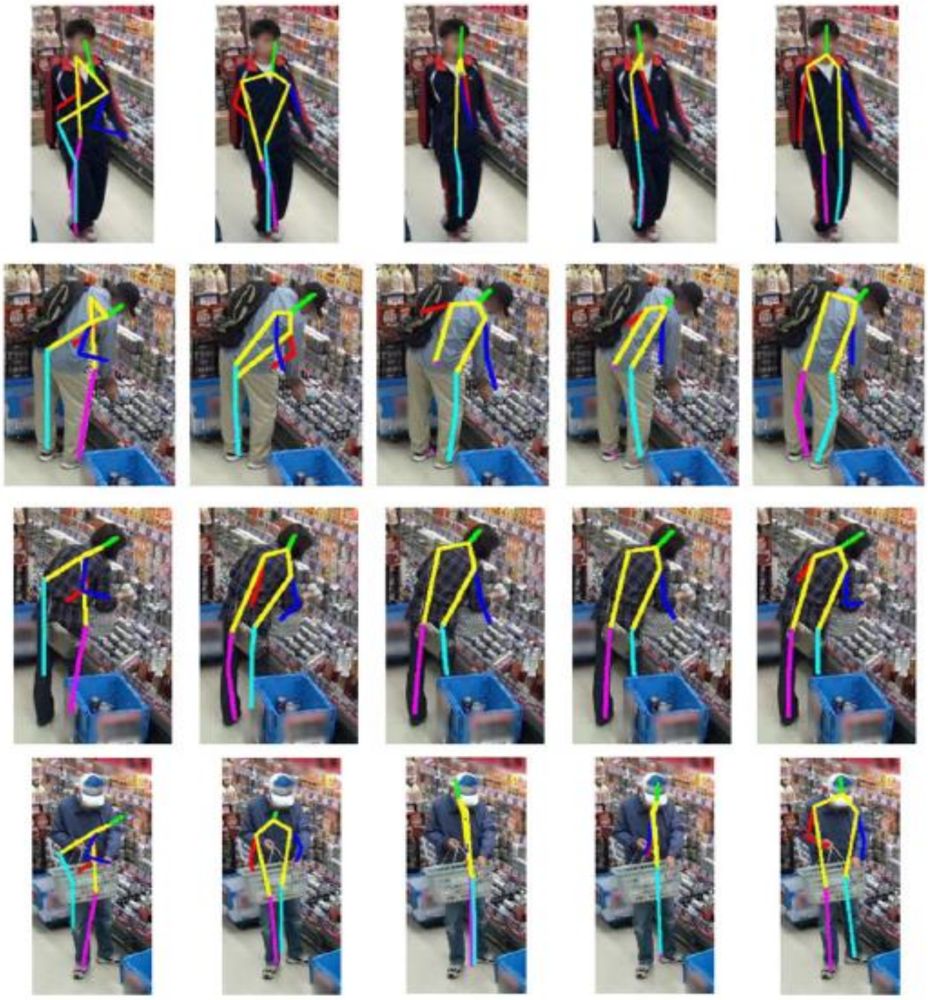

Fig. 3: Train and test curves of unagligned face images (Compiled by the authors)

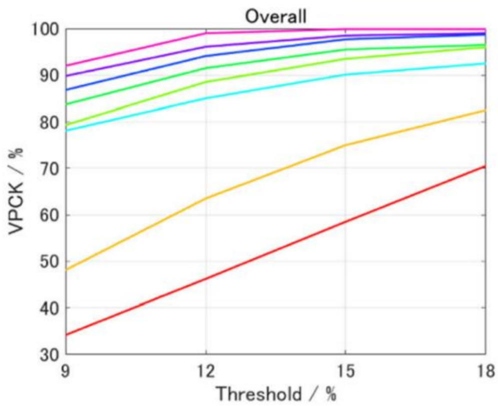

Fig. 4: Overall VPCK Threshold

(1) To overcome left-right similarity difficulties, the network incorporates body orientation information.
(2) The visibility mark layers are introduced into the network to enhance the efficiency of occluded joints.
(3) The proposed architecture incorporates a number of best practices from related works, including optical flow, posture heatmap, and visible landmark vector.

7. CONCLUSION AND FUTURE DIRECTIONS :

For customer pose estimation, a novel architecture using the concepts of deep learning is proposed. Furthermore, the occluding object detection clearly provides inter-occlusion by object cues. As a result, local joint connections, the global body orientation, occluding object, and human motion to address the occlusion problems and the left-right similarity, the proposed network combines the visibility mask and the body orientation. Information is all taken into account as part of a single deep neural network architecture. In future work, we intend to address customer pose estimates in numerous consumers' inter-occlusion. Further optimizations can be done to improve the effectiveness of the system. The robustness of the systems needs to be explored with data from various regions. The impact of the VPCK on customer satisfaction in different shopping sectors needs to be explored.

REFERENCES :

[1] Sminchisescu, C., & Telea, A. C. (2002). Human pose estimation from silhouettes. a consistent approach using distance level sets. In *10th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG'02)*, *10*(1), 1-10. Google Scholar

[2] Dantone, M., Gall, J., Leistner, C., & Van Gool, L. (2013). Human pose estimation using body parts dependent joint regressors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, *1*(1), 3041-3048. Google Scholar Crossref

[3] Eichner, M., & Ferrari, V. (2012). Appearance sharing for collective human pose estimation. In *Asian Conference on Computer Vision*, 138-151. Springer, Berlin, Heidelberg. Google Scholar Crossref

[4] Pishchulin, L., Andriluka, M., Gehler, P., & Schiele, B. (2013). Poselet conditioned pictorial structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 588-595. Google Scholar Crossref

[5] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, *1*(1), 886-893. Google Scholar Crossref

[6] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*(1), 396-404. Google Scholar

[7] Zhang, J., Nie, L., Wang, X., He, X., Huang, X., & Chua, T. S. (2016). Shorter-is-better: Venue category estimation from micro-video. In *Proceedings of the 24th ACM international conference on Multimedia*, 1415-1424. Google Scholar Crossref

[8] Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1653-1660. Google Scholar Crossref

[9] Carreira, J., Agrawal, P., Fragkiadaki, K., & Malik, J. (2016). Human pose estimation with iterative error feedback. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4733-4742. Google Scholar Crossref

[10] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), 1627-1645. Google Scholar Crossref

[11] Chu, X., Ouyang, W., Li, H., & Wang, X. (2016). Structured feature learning for pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4715-4723. Google Scholar Crossref

[12] Chen, X., & Yuille, A. L. (2015). Parsing occluded people by flexible compositions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3945-3954. Google Scholar Crossref

[13] Haque, A., Peng, B., Luo, Z., Alahi, A., Yeung, S., & Fei-Fei, L. (2016, October). Towards viewpoint invariant 3d human pose estimation. In *European Conference on Computer Vision*, 160-177. Springer, Cham. Google Scholar Crossref

[14] Rafi, U., Gall, J., & Leibe, B. (2015). A semantic occlusion model for human pose estimation from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 67-74. Google Scholar↗ Crossref↗

[15] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., & Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2758-2766. Google Scholar↗ Crossref↗

[16] Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 3686-3693. Google Scholar↗ Crossref↗

[17] Johnson, S., & Everingham, M. (2010). Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. In *bmvc*, *2*(4), 1-5. Google Scholar↗ Crossref↗

[18] Tafazzoli, F., & Safabakhsh, R. (2010). Model-based human gait recognition using leg and arm movements. *Engineering applications of artificial intelligence*, *23*(8), 1237-1246. Google Scholar↗ Crossref↗

[19] Yang, W., Ouyang, W., Li, H., & Wang, X. (2016). End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3073-3082. Google Scholar↗ Crossref↗

[20] Yang, Y., & Ramanan, D. (2012). Articulated human detection with flexible mixtures of parts. *IEEE transactions on pattern analysis and machine intelligence*, *35*(12), 2878-2890. Google Scholar↗ Crossref↗

****