HYBRIDIZING BIOMETRICS BASED VISUAL TRACKING WITH FACE AND GESTURES USING SVM AND WAVELETS

Dr. Soni Chaturvedi; Ms. Samiksha Khandekar

1. Associate Professor & Head ECE, P.I.E.T. Nagpur, contact: soni2569@gmail.com
2. M.tech Scholar ECE, P.I.E.T. Nagpur. Contact: samikshakhandekar2412@gmail.com

Abstract: Gender and gesture recognition has been a topic of research for many researchers for more than a decade. Researchers have worked on gender recognition and gesture recognition as two separate entities, and have achieved optimum results in both the domains. But, very few have worked on improving the quality of gender and gesture detection together in videos. In this paper, we propose a framework for improving accuracy of gender and gesture recognition, which can be used in visual surveillance systems as a tool to measure the irregularities in user behaviour of a particular gender. We have used daubichies based wavelets & spiking neural network (SNN) for gesture recognition and viola-jones cascade object detection combined with support vector machine (SVM) and facial geometry features in order to improve the quality of gender detection from faces. Our results when compared with k-Nearest Neighbour and Hidden Markov Models (HMMs) provides a 10% improvement in accuracy of gender recognition and 15% improvement of accuracy in gesture recognition from standard and real time videos. We plan to further enhance this work using machine learning in order to reduce the delay of processing for the system.

Keywords: Gender, Gesture, SVM, SNN, Wavelet, facial geometry

1. Introduction

Gender recognition from images basically means to categorize the user's input image into male, female or other based on certain distinctive facial features. These features can be evaluated on the basis of color, edge, geometry of the face or any other distinguishable set of features. The color based gender recognition is the most inaccurate, while the facial geometry based recognition is the most accurate, due to the fact that the color features only show color variation in the faces, which can vary for all gender types, while the geometry based features change for genders linearly, for example in most of the cases the eye to eye distance in males is more than that of females for a given dataset. Observations like these help in evaluating the gender from the given facial features.

Gesture recognition involves evaluating the actions performed by the user under test, and then concluding from those actions as to what the user is doing. This is a very vast field, and has been researched for more than a decade now. Researchers conclude that the window based technique which uses differential feature comparison is the most effective while evaluating gestures from a given set of frames for a given user under test.
In recent years the popularity of the gender from face pictures has attracted interest in each elementary and applied analysis. From the elemental purpose of read it's terribly intriguing to know however for personalities gender recognition is a simple operation that is finished terribly apace, except for a laptop vision algorithmic rule the task may be terribly difficult. The difficulties emerge from the doable variations of a face captured by a camera [1], that rely upon the image acquisition method (pose of the face, image illumination and distinction, background), the intrinsic variations between people’s faces (expression, age, race), in addition because the occlusions (sunglasses, scarves, hats). From the applied analysis purpose of read, there's an advert interest to own systems that may mechanically acknowledge the gender from face pictures. Examples embrace police work systems that may assist to limit areas to 1 gender solely quicker process in statistics systems that believe face recognition, custom user interfaces reckoning on the gender of the person interacting with them, good billboards designed to draw in the eye of male or feminine audience, and systems for the gathering of knowledge in support of marketing research. From these pictures, one might observe variations within the intensity distribution particularly within the hair and eyes regions. Supported the observation of facial pictures from normal datasets, several researchers use the pel intensity values of the faces to coach a binary classifier for gender recognition [3, 4, 5]. more variations are often ascertained in terms of texture. this might ensue to the softer countenance of ladies and additional pronounced eyebrows, whereas men have a rougher skin particularly within the presence of beard. the foremost widespread texture descriptors for face pictures square measure the histograms of native binary patterns (LBP) [6, 7, 8]. One may additionally observe a variation within the form of the face. The face of a girl is usually additional rounded, whereas the face of a person is additional elliptical. In [9], the authors exploited this facet and planned the utilization of bar chart of gradients (HOG) descriptor [10] for the popularity of gender. In different works, shape-based options are combined with different forms of options so as to own a additional sturdy classifier [11, 12, 13, 14]. Finally, there also are several delicate variations within the pure mathematics of the faces. the common face of a person has nearer eyes, a agent nose and a narrower mouth. These observations triggered the investigation of what square measure referred to as facial fiducial distances, that square measure primarily the distances between bound facial landmarks (e.g. nose, eyes contour, eyebrows) [15]. The fiducial points is also detected fix active form model [16] or deep learning techniques [17, 18].

Gesture-driven human laptop interaction (HCI) has been an energetic topic in recent years. The key element of gesture-driven HCI could be a gesture recognition system, that identifies the gestures to produce input to the choice system of HCI. several ways are developed for each hand gesture and full body gesture recognition [7]. These recognition tasks square measure difficult as a result of each human hand and body square measure extremely articulated. so as to develop associate degree economical gesture recognition system, a vital issue is to extract options to explain the extremely articulated body. in concert resolution, marker positions extracted fix infraredreflective marker systems square measure used in several interactive environments, love in [10]. Marker-based systems will dependably capture the 3D coordinates of the markers placed on bony landmarks of the body. However, carrying markers is cumbersome to the topic fix the system. in addition, business marker-based motion capture systems square measure terribly overpriced. Therefore, a video-based gesture recognition system is most well-liked for non-intrusive and low price sensing. In existing video-based gesture recognition systems, trailing bound “landmarks” within the motion pictures could be a wide applied strategy to extract motion options. In [28], human hands and head square measure caterpillar-tracked fix skin detector and face detector. In [26] and [24], some “visual attention-grabbing points” or
“visual cues” square measure accustomed describe the motion in every image frame. Landmark-based feature extraction is vulnerable to trailing failure in several cases, particularly once the angle of view of the camera changes, or equivalently the body orientation (facing direction) of the topic changes. In such cases, several landmarks may be occluded. The system delineate in [6] is landmark-free, however it's a read dependent system. There square measure some systems that perform view-invariant gesture recognition, love in [22] and [35], however these systems principally believe getting 3D data of the topic fix additional sophisticated camera systems.

In our technique, we propose a facial geometry based gender recognition system, and combine it with window & wavelet based gesture recognition system in order to combine both the fields and create a sufficiently accurate system, which can work for both standard and real time video datasets. The next section describes the various techniques used in gender and recognition, followed by our proposed approach for performing the task, and it's result evaluation. We conclude the paper with some good observations and a way forward for researchers in order to take this work further.

2. Literature review

In this section, we are going to be reviewing the various techniques for gender and gesture recognition. Golomb et al. [12] utilized a neural network, denoted as GenderNet, to classify gender from ninety face pictures sampled at thirty × thirty pixels. Gutta et al. [13] utilized radial basis perform and call trees for gender classification, wherever normalized pictures of sixty four × seventy two pixels, manually metameric, from FERET dataset [14] were utilized in their experiments. Moghaddam et al. [15] planned a gender recognition technique from twenty one × twenty one pixels from FERET dataset victimization SVM classifier. Nakano et al. [16] explored edge info to classify gender victimization neural networks. Yang et al. [17] planned a texture normalisation technique for up gender classification. Baluja ANd Rowley [18] developed an Adaboost classifier for distinctive gender on pictures from FERET dataset with resolution twenty × twenty pixels. Li et al. [19], [20] planned a gender recognition technique supported motion extracted from human silhouettes metameric from variety of action elements. Alexandre [1] planned a fusion theme supported form and texture options from multiple resolutions of normalized pictures (resized to 20×20, 36×36 and 128×128 pixels). Bekios-Calfa et al. [1] mentioned the utilization of linear techniques, cherish Linear Discriminant Analysis [2], for gender recognition. Perez et al. [3] explored intensity, texture and form options combined with mutual info. Results were evaluated on the FERET dataset. Dago-Casas et al. [4] conducted experiments with cross-database benchmarks for gender classification beneath at liberty conditions. Shan [5] developed a gender recognition approach based mostly a mixture of texture options victimization native binary patterns (LBP) with AN AdaBoost classifier. Experiments were conducted on the tagged Faces within the Wild (LFW) benchmark [6]. Toews and Arbel [7] delineate a framework for police work, nativeizing and classifying visual traits of objects from a viewpoint-invariant model derived from local scale-invariant options. A Bayesian classifier was utilized to spot the visual traits. FERET dataset was accustomed judge the gender classification technique. Mansanet et al. [8] planned a neighborhood deep neural network, denoted as local-DNN, that integrates native options and deep architectures. Experiments were conducted on the tagged Faces within the Wild (LFW) [6] and also the Gallagher’s dataset [9]. Fellous [10] investigated the utilization of twenty four normalized horizontal and vertical distances calculated from fiducial points extracted from a collection of 109 pictures. The model was trained on FERET dataset and different pictures nonheritable in his laboratory so as to predict the gender of varied facial expressions. Gupta [11] given AN approach to police work gender of individuals through frontal facial pictures supported data processing and Delaunay
triangulation techniques. Numerous classifiers, cherish useful trees, random forests, naïve Thomas Bayes, AdaBoost and J48, were accustomed acknowledge a gender as male or feminine on the FEI Face dataset [8]. Patel et al. [12] developed a facial gender recognition technique supported a compass native binary pattern descriptor, that was evaluated on CUFs and CUFSF datasets. Saint Matthew ANd Hassner [33] developed an automatic gender classification employing a convolutional neural network (CNN) to find out representations from the info and improve the performance of the classification task. Experimental results were evaluated on the urge benchmark.

The recognition of gesture involves many ideas cherish pattern recognition [21], motion detection and analysis [21], and machine learning [21]. Completely different tools and techniques square measure used in gesture recognition systems, cherish laptop vision [22], image process [22], pattern recognition [23], applied math modeling [24]. The utilization of neural networks for gesture recognition has been examined by several researchers. Most of the researches use ANN as a classifier in gesture recognition method, whereas some others use it to extract the form of the hand, as in [25]. Tin H. [26] presents a system for hand chase and gesture recognition victimization NNs to acknowledge Union of Burma Alphabet Language (MAL). Adobe Photoshop filter is applied to seek out the sides of the input image and bar chart of native orientation utilized to extract image feature vector which might be the input to the supervised neural networks system. Manar M. [27] used 2 repeated neural network architectures to acknowledge Arabic signing (ArSL). Elman (partially) repeated neural networks and absolutely repeated neural networks are used individually. A coloured glove used for input image knowledge, and for segmentation method, HSI color model is applied. The segmentation divides the image into six color layers, one for the articulatio radiocarpea and 5 for fingertips. Thirty options square measure extracted and sorted to represent one image, fifteen components accustomed represent the angles between the fingertips and between them and also the articulation of data [27], and fifteen components to represent distances between fingertips; and between fingertips and also the articulation of data [28]. This input feature vector is that the input to each neural networks systems. 900 coloured pictures were used as coaching set, and three hundred coloured pictures for system testing. Results had shown that absolutely repeated neural network system (with recognition rate ninety five.11%) higher than the Elman neural network (with eighty nine.67% recognition rate). Kouichi M. in [29] given Japanese signing recognition victimization 2 completely different neural network systems. Firstly, back propagation algorithmic program was used for learning postures of Japanese alphabet. For input postures, knowledge glove is employed, and normalisation operation was applied as a preprocessing tread on the input image. The feature extracted from input pictures was thirteen knowledge things, 10 for bending, and 3 for angles within the coordinates. The output of the network was forty two characters. The network consists of 3 layers, the input layer with thirteen nodes, the hidden layer with a hundred nodes, and also the output layer with forty two nodes that corresponds forty two recognized characters. The popularity rate for learning forty two educated patterns was seventy one.4%, and for unregistered folks forty seven.8%, whereas the speed improved once extra patterns accessorial to the system, it became ninety eight.0% for registered, and 77.0% for unregistered folks. Elman repeated Neural Network was the second system applied for recognition gestures. The system might acknowledge ten words. The information item are taken from data glove and also the same preprocessing applied for input image. options extracted square measure sixteen knowledge things, ten for bending, three for angles within the coordinates, and three for angles within the coordinates. The network consists of 3 layers, the input layer with sixteen nodes, the hidden layer with one hundred fifty nodes, and also the output layer with ten nodes that corresponds ten
recognized words. Some improvement within the point knowledge and filtering knowledge house square measure accessorial to the system [28]. Integration of those 2 neural networks, in a way, that once receiving knowledge from knowledge glove, a determination of the beginning sampling time and if the info item thought-about a gesture or a posture is shipped to successive network, for checking the sampling knowledge and also the system hold a history, that decide the top of signing. the ultimate recognition rate with the secret writing ways was ninety six. Stergiopoulou E. [29] recognized static hand gestures victimization Self-Growing and Self-Organized Neural Gas (SGONG) network. A camera used for effort the input image, and YCbCr color house is applied to discover hand region, some thresholding technique accustomed discover colouring. SGONG network use competitive Hebbian learning algorithmic program for learning method, the training begin with solely 2 neurons and continuous growing until a grid of neurons square measure made and canopy the hand object which is able to capture the form of the hand. From the resultant hand form 3 geometric options was extracted, 2 angles supported hand slope and also the distance from the palm center determined, wherever these options accustomed verify the amount of the raised fingers. For recognizing tip, statistical distribution model employed by categoryifying the fingers into 5 categories and calculate the options for every class. The system recognized thirty one predefined gestures with recognition rate ninety.45%, in time interval one.5 second. Shweta K. in [29] introduced gesture recognition system victimization Neural Networks. Web-cam used for capturing input image at slow rate samples between 15-25 frames per second. Some preprocessing are created on the input image that convert the input image into sequence of (x, y) coordinates victimization MATLAB, then passed into neural classifier, within which it'll classify the gesture into one in all many classed predefined categories which might be known by the system. poet and Geoffiey [30],[31] used neural networks to map hand gestures to speech synthesizer victimization Glove-Talk system that translated gestures to speech through adjustable interface that is a very important category of neural networks applications [31].

Many researches are applied based mostly the bar chart, wherever the orientation bar chart is employed as a feature vector [32]. the primary implementation of the orientation bar chart in gesture recognition system and real time was done by William F. and Michal R. [32]; they given a way for recognizing gestures supported pattern recognition victimization orientation bar chart. For digitized input image, black and white input video was used, some transformations were created on the image to calculate the bar chart of native orientation of every image, then a filter applied to blur the bar chart, and plot it in polar coordinates. The system consists of 2 sections; coaching phase, and running section. within the coaching section, for various input gestures the coaching set is keep with their histograms. In running section an input image is given to the pc and also the feature vector for the new image is created. Then comparison performed between the feature vector of the input image with the feature vector (oriented histogram) of all pictures of the coaching section, victimization euclidian distance metric and also the less error between the 2 compared histograms are elect. the full method time was a hundred millisecond per frame. Hanning Z., et al. [33] given hand gesture recognition system supported native orientation bar chart feature distribution model. colouring based mostly segmentation algorithmic program were accustomed notice a mask for the hand region, wherever the input RGB image born-again into HSI color house, so map the HSI image H to a chance quantitative relation image L the hand region is metameric by thresholding price, 128 components within the native orientation bar chart feature were used. The increased of the native orientation bar chart feature vector enforced by adding the image coordinates of the sub-window. To compact options illustration, k-means clump has been applied over the increased native orientation bar chart vectors. In Recognition
stage, euclidian distance accustomed calculates the precise matching score between the input image and keep posture. Then neighbourhood Sensitive Hashing (LSH) accustomed notice the approximate nearest neighbors, and cut back process value for image retrieval. Wyososki et al. [34] given a rotation invariant static-gesture recognition approach victimization boundary histograms. Colouring detection filter was used, followed by activity erosion, dilation as preprocessing operation, and clump method to seek out the teams within the image. for every cluster the boundary was extracted victimization a normal contour-tracking algorithmic program. The image Divided into grids, and normalized the boundary in size, that offer the system invariability distance between the camera and hand. homogeneous background was applied, and also the boundary is drawn as chord’s size chain. The image was divided into variety of regions N. and also the regions were divided in a very radial type [34], in step with a selected angle as shown within the Figure. The bar chart of boundary chord’s size was calculated. that the whole feature vector consists of a ordered chain of histograms. Multilayer perceptor (MLP) Neural Networks and Dynamic Programming (DP) matching were used as classifiers. twenty six static postures from yank signing, for each posture, forty photos were taken, twenty photos for coaching and twenty for check. completely different variety of histograms were used varies from eight to thirty six increasing by 2, with completely different bar chart resolutions.

Clustering algorithms could be a general term contains all ways that partitioning the given set of sample knowledge into subsets or clusters [35] supported some measures between sorted components [36], in step with this live the pattern that share identical characteristics square measure sorted along to create a cluster [32]. Clump Algorithms are wide unfolded attributable to their ability of grouping sophisticated knowledge collections into frequently clusters [35]. In fuzzy clump, the partitioning of sample knowledge into teams in a very fuzzy approach square measure the most distinction between fuzzy clump and different clump algorithmic program [32], wherever the one knowledge pattern would possibly belong to completely different knowledge teams [32]. Xingyan L. In [32] given fuzzy c-means clump algorithmic program to acknowledge hand gestures in a very mobile remote. A camera was used for acquire input raw pictures, the input RGB pictures square measure born-again into HSV color model, and also the hand extracted once some preprocessing operations to get rid of noise and unwanted objects, and thresholding accustomed section the hand form. thirteen components were used as feature vector, initial one for ratio of the hand’s bounding box, and also the rest twelve parameters represent grid cell of the image, and every cell represents the mean grey level within the three by four blocks partition of the image, wherever the mean of every cell represents the common brightness of these pixels within the image. Then FCM algorithmic program used for classification gestures. numerous environments square measure utilized in the system cherish complicated background and invariant lighting conditions. half-dozen hand gestures used with twenty samples for every gesture within the vocabulary to form the coaching set, with recognition accuracy eighty five.83%.

Many researches were applied within the field of gesture recognition victimization HMM. HMM could be a model [36], with a finite variety of states of Markov chain, and variety of random performs in order that every state encompasses a random function [36]. HMM system topology is drawn by one state for the initial state, a collection of output symbols [36][32], and a collection of transitions state [32][28]. HMM contained loads of mathematical structures and has verified its potency for modeling spatio–temporal information knowledge [36]. signing recognition, square measure one in all the foremost applications of HMM [28], and speech recognition [30]. In [29] Keskiin C., et. al. given HCI interface supported real time hand chase and 3D gesture recognition victimization hidden mathematician models (HMM). 2 coloured cameras for 3D
construction square measure used. To beat the matter of victimization colouring for hand detection attributable to hand overlapping with different body elements, markers square measure accustomed cut back the quality in hand detection method [29]. Markers accustomed section the hand from complicated backgrounds beneath invariant lighting conditions. The markers square measure distinguished victimization marker detection utility, and connected elements algorithmic program was applied to seek out marker regions victimization double thresholding. For tip detection, straightforward descriptors were used, wherever the bounding box and 4 outermost points of the hand that shaping the box is decided [29]. The bounding confine some cases has to be elongate to see the mode of the hand, and also the points accustomed predict the tip location in several modes of the hand. Kalman filter was used for filtering flight of the hand motion. For 3D reconstruction of finger coordinates, standardization utility was enforced for specific standardization object [29]. Least sq. approach accustomed generate tip coordinates, and kalman filter applied for smoothing the flight of 3D reconstructed coordinates. To eliminate reference frame dependency, the 3D coordinates square measure born-again into sequences of quantal rate vectors. HMM interprets these sequences [29], that square measure directional characterizing of the motion [26]. The system designed for game and painting programs application. Hand chase is used to imitate the movements of the mouse for drawing, and also the gesture recognition system used for choosing commands. Eight gestures are used for system coaching, and one hundred sixty for testing, with 98.75% recognition performance.

3. Proposed hybrid Gesture and Gender recognition system

The proposed technique can be represented using the following block diagram shown in figure 1,

![Figure 1. Overall system block diagram](image)

From the block diagram, we can observe that the input video set is first collected for both gender and gestures. The gender video dataset has videos which have frontal facing images of persons, this video set has a collection of more than 120 people and is taken from the UCI learning repository. The gesture video dataset is taken from various sources, including but not limited to berkley video dataset, open surveillance video dataset and stanford video dataset. These datasets contain videos of users performing more than 50 different gestures over more than 100 persons. Such a wide variety of dataset is selected for proper evaluation of the system.

The frames from each of these videos is given input to the system. In the gender detection block, the viola jones algorithm is used to detect the face from the input video. The viola jones algorithm gives the location of various geometric points on the face, like eye locations, nose location, and mouth location. We then evaluated the following features from the facial points,

- Eye to eye distance
- Left & right eye to nose distance
- Left & right eye to mouth distance
- Left and right eye width and height
- Nose to mouth distance
- Nose width and height
- Mouth width and height

When combined together, these features form a Feature Vector (FV) of about 12 different facial geometry features. These features are saved into the database of Matlab along with the training
classes of Male & Female for each of the videos. For every new video frame, the database is evaluated and the class is found for each of the video frame. For N frames, we get N class values, the highest probability of occurring value is selected as the gender for the given video.

The gesture recognition block is the lower part of the block diagram. In this part, we first select the human silhouette for which the gestures are to be evaluated. The selected silhouette is given to a daubechies level 8 filter bank (DB8), which produces approximate, horizontal, vertical and diagonal components of the selected object. These features are stored into a feature vector (FV). For each of the consecutive frames, these FVs are compared within the MxM window using a standard Spiking Neural Network (SNN) with 10 input neurons, 50 hidden neurons and 1 output neuron, and the most matching FV frame from the SNN is selected as the location of the selected silhouette, and it's location is marked. At the end of the video, these marked locations are tracked in order to evaluate the speed and direction of the silhouette. The speed and direction is then given to a decision block in order to evaluate the gesture of the silhouette.

The evaluated gesture and gender is given to a combination block, which shows the combined output of the gender and gesture for the given video. The next section describes the results and analysis of our proposed approach and compares it with some existing approaches in order to provide a complete scenario of where our algorithm stands in terms of accuracy of detection.

### 4. Results and Analysis

We compared our algorithm with kNN and HMM. kNN is a linear classifier, while HMM is a non-linear classifier. The results of both the classifiers when compared with our proposed approach is as shown in the following tables.

<table>
<thead>
<tr>
<th>Train Size</th>
<th>Test Size</th>
<th>Acc. (%) kNN</th>
<th>Acc. (%) HMM</th>
<th>Acc. (%) Prop.</th>
<th>Improv. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
<td>80.00</td>
<td>85.00</td>
<td>85.00</td>
<td>5.88</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>81.50</td>
<td>86.80</td>
<td>87.20</td>
<td>6.54</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
<td>82.60</td>
<td>88.20</td>
<td>88.50</td>
<td>6.67</td>
</tr>
<tr>
<td>75</td>
<td>90</td>
<td>82.80</td>
<td>89.30</td>
<td>90.20</td>
<td>8.20</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>83.10</td>
<td>89.90</td>
<td>90.60</td>
<td>8.28</td>
</tr>
<tr>
<td>200</td>
<td>300</td>
<td>83.40</td>
<td>90.20</td>
<td>91.80</td>
<td>9.15</td>
</tr>
<tr>
<td>500</td>
<td>700</td>
<td>83.50</td>
<td>90.50</td>
<td>93.50</td>
<td>10.70</td>
</tr>
</tbody>
</table>

Table 1. Comparison of accuracy of gender recognition

![Graph showing gender detection accuracy vs number of tested frames](image)

Figure 2. Visual analysis for gender detection accuracy v/s number of tested frames

From our analysis we can observe that the proposed system outperforms both kNN and HMM based systems, due the fact that SVM classifier is a strong 2 class classifier, and can strongly distinguish between frames with different geometrical features. Thus, SVM is the selected choice for any gender recognition system.

The results for gesture recognition which can be observed from the following table, also conclude that the proposed approach is about 15% improved as compared to kNN and HMM, which is due to the fact that the proposed system uses a complex set of DB8 features and uses a bigger search window of 16x16, which ensures that the object under tracking is never out of the frame, and thus the gesture tracking is done effectively.
Table 2. Comparison of accuracy of gesture detection

<table>
<thead>
<tr>
<th>Train Size</th>
<th>Test Size</th>
<th>Acc. (%) kNN</th>
<th>Acc. (%) HMM</th>
<th>Acc. (%) Prop.</th>
<th>Improv. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
<td>60.00</td>
<td>75.00</td>
<td>80.00</td>
<td>25.00</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>72.50</td>
<td>78.50</td>
<td>81.50</td>
<td>11.04</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
<td>78.60</td>
<td>79.80</td>
<td>82.60</td>
<td>4.84</td>
</tr>
<tr>
<td>75</td>
<td>90</td>
<td>81.50</td>
<td>82.60</td>
<td>85.80</td>
<td>5.01</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>83.60</td>
<td>85.80</td>
<td>89.50</td>
<td>6.59</td>
</tr>
<tr>
<td>200</td>
<td>300</td>
<td>83.80</td>
<td>89.20</td>
<td>91.60</td>
<td>8.52</td>
</tr>
<tr>
<td>500</td>
<td>700</td>
<td>84.10</td>
<td>90.50</td>
<td>93.80</td>
<td>10.34</td>
</tr>
</tbody>
</table>

Figure 3. Visual analysis for gesture recognition

We implemented the system in Matlab, the following screenshots demonstrate the output figures from the simulations.

Figure 4. Male detection in video

Figure 5. Female detection in video

Figure 6. Occlusion based gesture tracking

Figure 7. Varying object size for gesture tracking

Figure 8. Summary of the tracked walking object
Figure 9. Summary of moving closer to camera

These figures are some certain examples of the given videos, we tried our system on various other video sequences and found the results to be following the same results.

5. Conclusion

From the obtained results, we can conclude that the proposed system is more than 10% accurate than the kNN and HMM based gender detection systems, and it is more than 15% accurate in terms of gesture recognition accuracy when compared with similar techniques. Evaluation of the algorithms using other datasets also produce similar results, and thus can be used in real time systems.

6. Future work

The proposed protocol demonstrates good accuracy for both stored and real time videos. But it is observed that the system has moderate to high delay as the training set increases, thus it needs to be reduced by using machine learning optimizations, which can be carried out as a future work for this research.

References


