A Survey on Human Sex Determination Methods

Anand Kunwar	Harkirat Behl	Manikanta Reddy	Parag Bansal
13110	13286	13265	13464

Under Guidance of Prof. Harish Karnick

Abstract

Identifying the sex of an individual is a challenging problem in Computer Vision. In the recent years various methods have surfaced to solve this classification problem and get closer to achieve better human machine interaction. In this survey we study a number of such algorithms which classify human subjects into male and female classes. This task is done by looking at a number of features, like face, body and motion.

Contents

1 Introduction

We provide a survey of human sex recognition using computer vision techniques. We review multiple methods the exploit information from face and whole body. Identifying demographic attributes is a key point of machine-human interaction. While humans can do this task very easily it is a challenging task for computer vision.

In general, we breakdown the problem into multiple phases. Object detection, preprocessing, feature extraction and classification. All these sub problems depend on the choice of feature we wish to encode and use. Once the object of interest is detected we then process it to extract features and their corresponding descriptors. We then use these descriptors to train a classifier.[7]

2 Face Based Methods

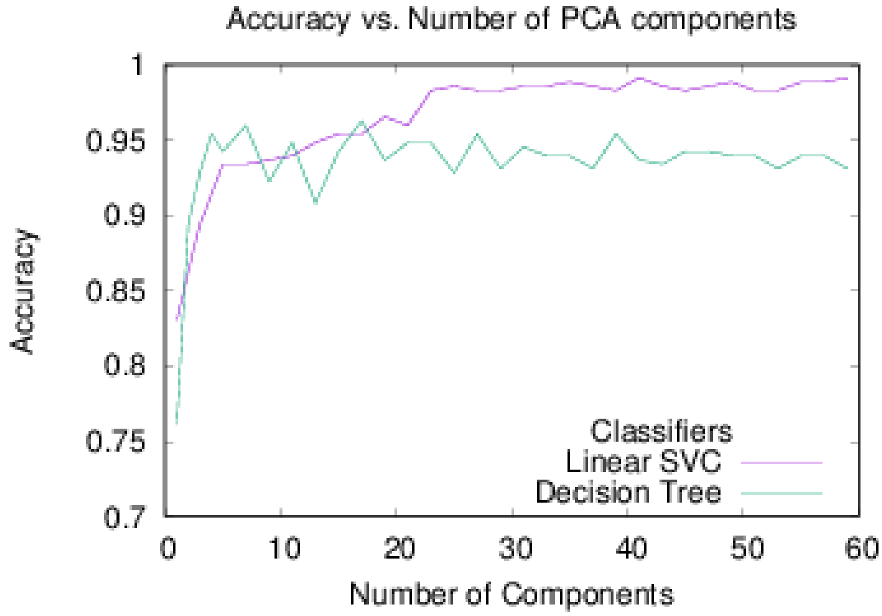
While using faces as our objects, we realize that the face region could also include features from hair and neck. A human also exhibits variation which make it hard for us to encode them using raw pixel values. The features also depend on race, age, expression and accessories worn.[12][13]

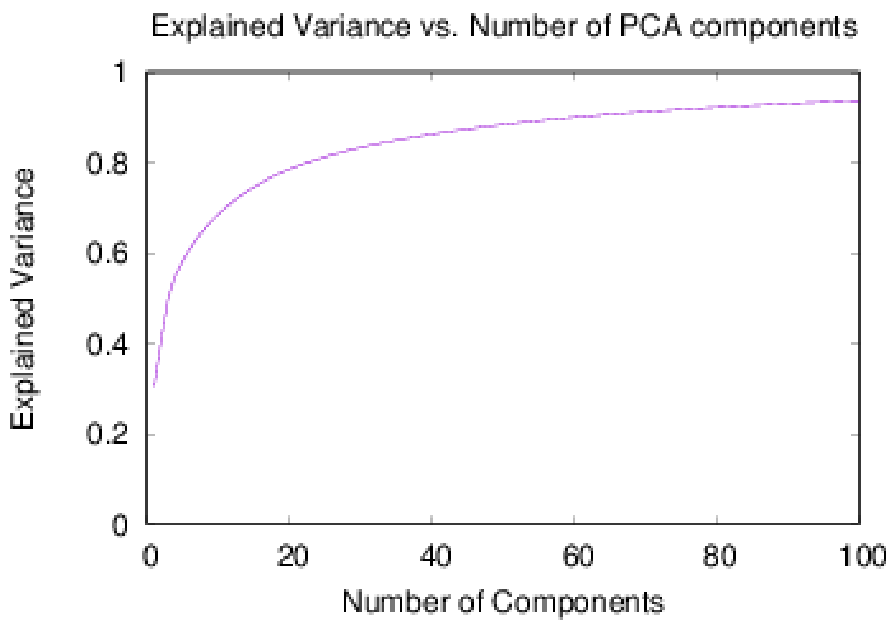
2.1 Dimensionality Reduction using PCA

To extract feature vector from our facial image, we used the technique of dimensionality reduction using PCA[4]. PCA is a transformation scheme which transforms the data to a new coordinate system. The axes in this new coordinate system are found by finding the directions of maximum variance in our data. Direction of maximum variance is found by calculating the Eigenvector of covariance matrix corresponding to large Eigen values.

Around 95% of total variance was explained by 80 dimensions. Following Eigenfaces were obtained after dimensionality reduction of image. We used these Eigenfaces[11] as input to our classifier for classification to male and female.

We tried Linear Support Vector classifier and Decision tree based classification. Following results were obtained.





The classification accuracy seems to be pretty good around 98%. As expected, SVM classifier performs better than decision forests.

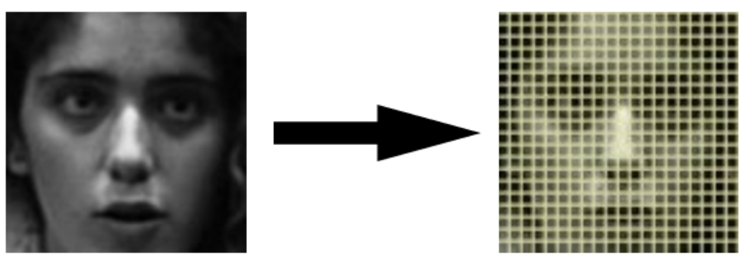
2.2 Data Set Used

Nottingham Dataset[8]

Faces94[10]

2.3 Lowe Key Point Descriptors

We tried to encapsulate the key points of the image by using a Lowe Descriptors in an orderly fashion. We divided the image into grids and generate Lowe descriptors for the intersection points. Lowe descriptor[5] is essentially a histogram of gradient map of the neighborhood of the descriptor.



We generated key point descriptors in a grid wise manner for every person and performed a classification based on them.

We then flattened all the key point descriptors in an order, row by row and used it as a descriptor for the image.

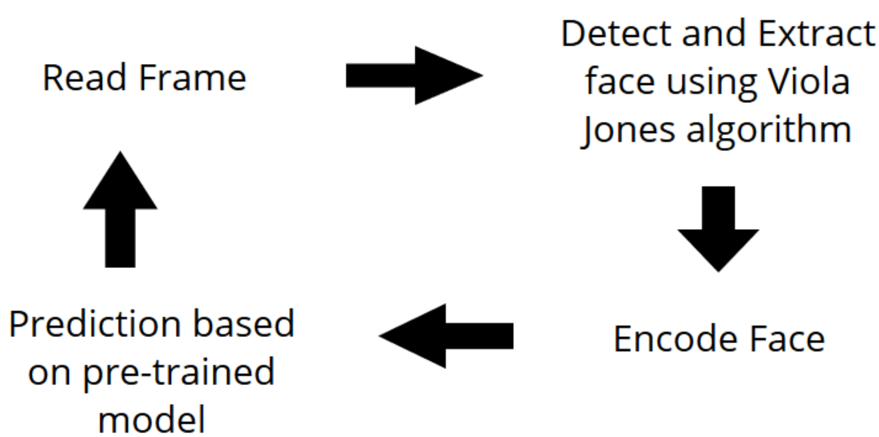
We then used multiple classifying algorithms to identify sex. Here are the results of our experiment.

2.3.1 Accuracies

Classifier	Lowe Descriptor
LinearSVC	0.611 ± 0.002
Random Forest	0.600 ± 0.002
Perceptron	0.628 ± 0.090
Adaboost	0.621 ± 0.005

We realize at this point that using Lowe descriptors wasnt a very good idea as the orderly way we encoded the vectors didnt actually capture the features (Spatial Description in the form of geometric features).
This might be the reason for very low performance of this method.

2.4 Online Port



The blue box around a person's face indicates that his sex is male. As we have haar cascades for frontal face, all the faces in the frame are not detected and hence are not classified.

2.5 Demonstration

We ported the algorithm based on facial features to work on the IIT Kanpur Pedestrian Data set and the results were visually pleasing.

2.6 Caveats

The current implementation is not very robust to uncontrolled environmental conditions, as we rely on Haar cascades for facial detection.

- Only frontal faces are detected
- The resolution of faces in the video is very low

3 Full Body Features

3.1 Pedestrian Detection

There are various difficulties in localizing pedestrians in an image. These can be broadly classified into mainly Illumination, Occlusion, scales, clutter in the background, variable appearances and the different human poses. We have used the famous Histogram of Oriented Gradients [1] for Human detection. The pipeline briefly has the following steps: Scan image at all possible scales and locations, then extract features over these windows and run Linear SVM , then fuse the multiple detections that we have found in 3D space for scale and position and we finally get the true bounding box for the detections. Dalal tried to optimize each of these steps and used every possible regularization at each step. For the learning part, first a fixed resolution and normalized training image data set is created and after encoding images into feature spaces, a binary classifier is learnt.[9]

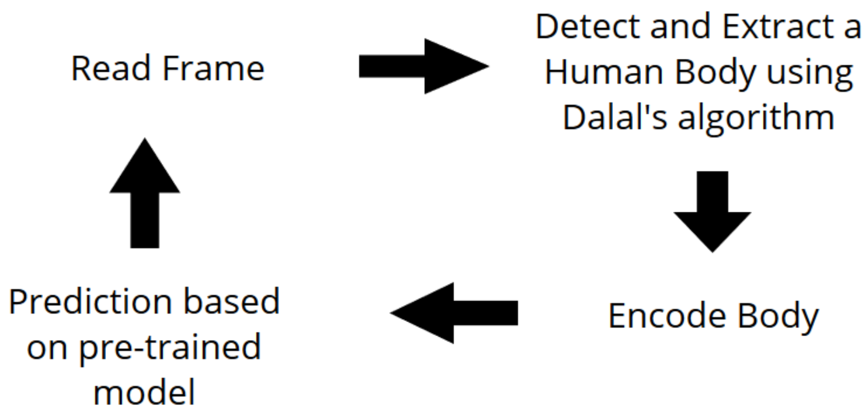
3.2 Data Set Used

MIT Pedestrian Dataset was used[6]

3.3 Results

Classifier	HOG Descriptor	Pixel Values
LinearSVC	0.75 ± 0.04	0.63 ± 0.05
Random Forest	0.67 ± 0.04	0.60 ± 0.05
Perceptron	0.69 ± 0.09	0.63 ± 0.09
Adaboost	0.71 ± 0.05	0.65 ± 0.04
MultinomialNB	0.63 ± 0.04	0.53 ± 0.04

3.4 Online Port



3.5 Demonstration

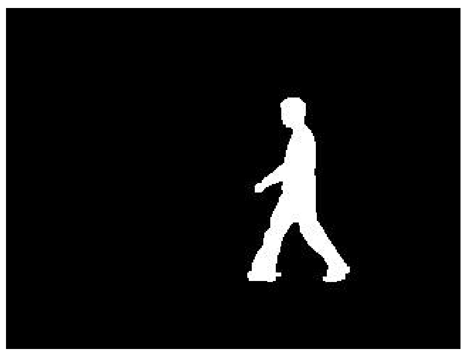
As the human recognition algorithm detects humans in general, it could detect people driving their bycycles too. It can bee seen that this was very much a possible result.

The redbox around the woman's body indicates that here sex is female based on her full body features.

4 Motion Based Features

Gait[2] is defined as the combination of coordinated cyclic movements are that result in human locomotion, which includes walking, running, jogging and climbing stairs. Gait of a person on foot is often used to exploiting process information in some situations, for example when the face is not visible. In a video of a person walking, the gait cycle can be designated as the time interval between two positions. Many factors affect the approach of a person, such as the load, shoes, walking area, injury, mood, age and change over time. Video analysis of this process will also deal with clothes for the camera, walking speed and clutter background.

For our purpose gait images are background subtracted images of persons motion across a scene at different times in an interval.



Data Set used is CASIA Gait Data Set[3]

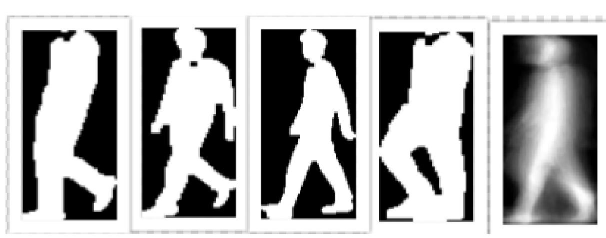
From these gait images we'll extract only the person/silhouette as it is the only part that holds any feature relevant to our problem.

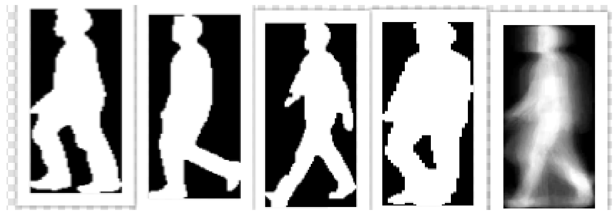
4.1 Gait Energy Image

We define Gait Energy Image as the average of silhouettes in a gait/walk cycle.[14]

$$F(i,j) = \frac{1}{T}\sum_{t=1}^{T} I(i,j,t) \tag{1}$$

where T is the number of frames in the sequence $I(i,j)$, $I(i,j,t)$ is the binary silhouttes image pixel at frame $t(i,j)$ with i,j as coordinates.





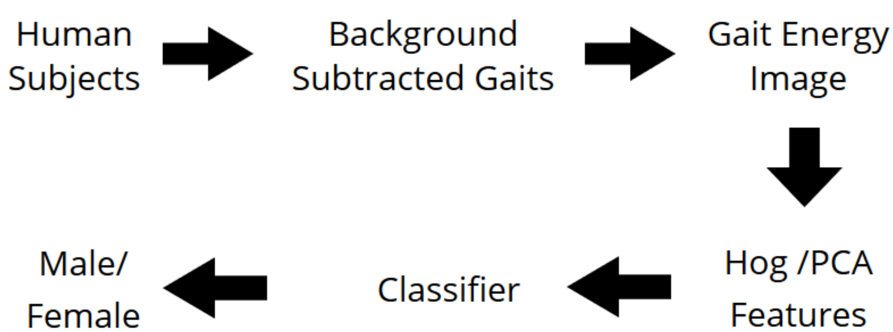
These are the extracted Gait Images. The last image in each row is the Gait Enery Image of the corresponding row.

4.2 Classifiers

We've used multiple classifiers with GEI as the input vectors. SVM as was expected is a very good classifiers for this two class problem.

Classifier	HOG Descriptor	PCA Reduction	Raw GEI Values
LinearSVC	0.83 ± 0.008	0.90 ± 0.006	0.91 ± 0.004
Random Forest	0.60 ± 0.008	0.75 ± 0.020	0.80 ± 0.030
Perceptron	0.74 ± 0.030	0.85 ± 0.010	0.60 ± 0.030
Adaboost	0.66 ± 0.020	0.79 ± 0.009	0.82 ± 0.000

4.3 Online Port



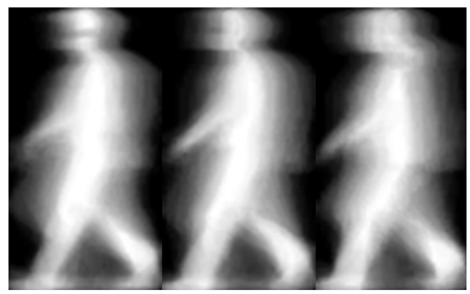
4.4 Problem with an online version

- To generate a Gait Energy Image, we need a set of gaits of the person.
- The problem with generating such gaits, lies on three fronts.
- Detection of Person
- Background Subtraction
- Tracking

4.5 Contribution of Halos and Cores in Gait Energy Images

The images to right are normal gaits extracted. If we use these directly the accuracy is a bit low 70%
This might be because we are unable to quantify what stride to use.

The images to right are Average gaits. As was reported they provide an accuracy of 91%



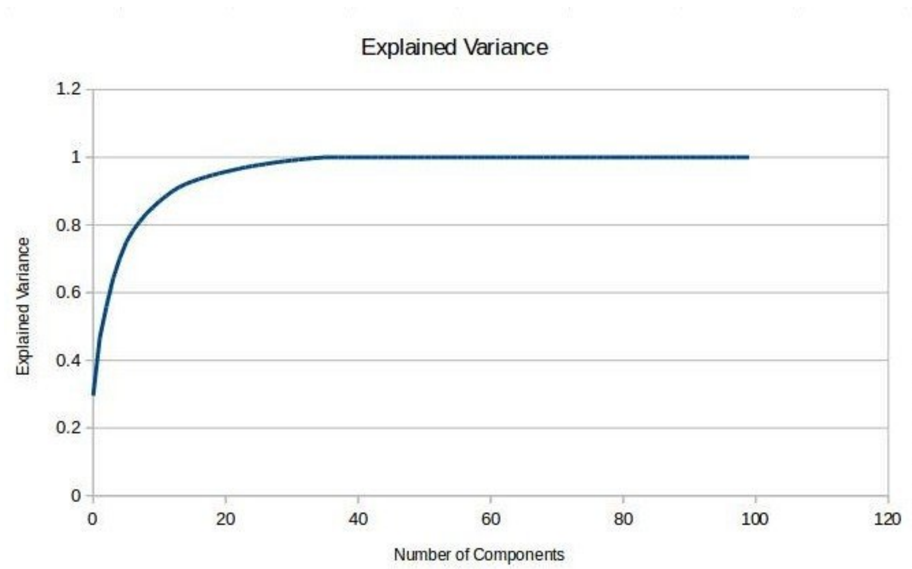
The images to right are Halo subtracted Average gaits (Cores). These seem to capture the body features of an individual in a normalized fashion.
They surprisingly provide us with an accuracy of 87%

he images to right are just the Halos around the cores, they capture the body motions effectively.
Even they provide us an accuracy of 86%

As we can see the peoples strides could be random in single images but on an average, it incorporates all the information about the pattern of movement, thus Gait images provide us with good results.

The results above also suggest that the Halos as well as the cores independently are capable of describing a person, with notable accuracy (¿ 85%). When used together as is the case in Average gait image, the information about an individual seems to have been boosted, hence greater success rates.

4.6 Explained Variance for PCA reduction



Optimal Value of N turns out to be close to 30
This gives us an idea that the entire motion can most probably be described in 30 features.

References

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[2] Jinguang Han and Bir Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):316–322, 2006.

[3] Chinese Academy of Sciences (CASIA) Institute of Automation. CASIA Gait Dataset. `http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp`. Accessed: 2016-04-05.

[4] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[5] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[6] MIT. MIT Pedestrian Data Set. `http://cbcl.mit.edu/software-datasets/PedestrianData.html`. Accessed: 2016-04-01.

[7] Choon Boon Ng, Yong Haur Tay, and Bok Min Goi. Vision-based human gender recognition: A survey. *arXiv preprint arXiv:1204.1611*, 2012.

[8] University of Stirling. Nottingham Faces Dataset. `http://pics.psych.stir.ac.uk/2D_face_sets.htm`. Accessed: 2016-04-05.

[9] C. Papageorgiou, T. Evgeniou, and T. Poggio. A trainable pedestrian detection system. In *Proceeding of Intelligent Vehicles*, pages 241–246, October 1998.

[10] Libor Spacek. Faces94 Faces Dataset. `http://cswww.essex.ac.uk/mv/allfaces/faces94.html`. Accessed: 2016-04-02.

[11] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[12] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[13] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[14] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 441–444. IEEE, 2006.