MACHINE PERCEPTION PROJECT
PROGRESS REPORT

STATE PREDICTION BY ATTIRE RECOGNITION

SUBMITTED BY-

GROUP 11

CHETNA JAIN MT2016044

RACHANA PRAJAPATI MT2016105

SHRADDHA AGRAWAL MT2016129

SHUBHAM SINGH MT2016131

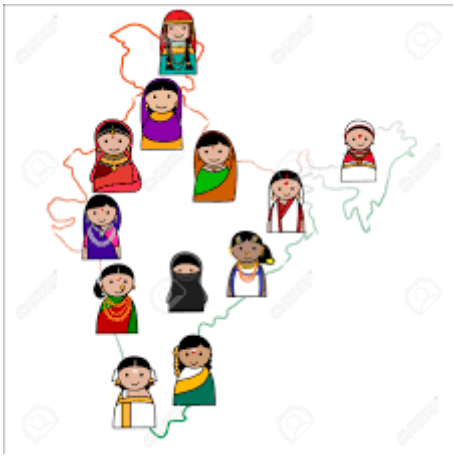
Contents

MOTIVATION

State prediction by attire recognition.

Every country has its own cultural costumes. In Indian ethnic wear, silhouettes are mostly standard and basic; it concentrates more on printing techniques, surface ornamentation, intricate detailings, motifs and dyeing techniques.
In western wear, silhouettes are experimented and it concentrates more on shapes, textures, fabric manipulation, cuts and drapes.

Even in India every state follows different culture. States have there own festivals, language, songs, dance forms, food, believes and attire etc. These characterstics strongly defines a particular state of India.
One of the ways by which we can distinguish states from each other is from its traditional attire. Clothing in India varies depending on the different ethnicity, geography, climate and cultural traditions of the people of each region of India. India also has a great diversity in terms of weaves, fibers, colours and material of clothing. Colour codes are followed in clothing based on the religion and ritual concerned.



From our current work we would like to predict the state/region from the image, depending upon what tradition dress person is wearing.

INTRODUCTION

Initially to train our model we have collected images from the following states as our dataset:

- Gujarat (Ghagara Choli)
- Haryana (Daaman, Chunder and Kurti)
- Jammu and Kashmir (Burkha, Pheran and Scarf on head (called Taranga))
- Uttar Pradesh (Chikkan work)
- Punjab (Punjabi salwar suit)

General Features of Gujarati Costumes

- Brightly colored costumes
- Richly adorned costumes exhibiting intricate stone work, mirror work, bead work
- Beautifully embroidered costumes showing stunning designs done with golden as well as silver zari threads
- Intricate designs, finesse, figurative, floral and geometric patterns exhibited on the costumes
- Ornate jewelries and accessories offer the final touch to the typical Gujarati look
- The traditional costume of Gujarati women consists in the following items - Chaniyo, a richly colored and embellished petticoat
- Choli/Polku, a richly adorned blouse
- Odhani/Churni, a lightweight, transparent and decorated head cloth to add the final glamorous touch

General Features of Haryanvi Costumes

- The dresses of the Haryanvi's are usually very simple i.e. a dhoti, a shirt, a turban and a pair of shoes. Additionally, a Chaddar, a type of blanket, serves as a wrapper.
- A woman's dress is much more vivid and colourful than a male attire. A Jat woman's attire essentially consists of a full dress or Thel, Ghaggri and printed Orhni.
- The colour of dress of different communities is also unique to them. A Gujjar may therefore be differentiated from an Ahir woman, by the colour of the dress.
- Ornaments made of gold and silver are usually worn by men and women. Necklaces, bracelets, gold chains are usually worn on special occasions such as marriages etc.

General Features of Kashmiri Costumes

- The Pherans are the prominent attire for Kashmiri women, as well. Traditionally, there are the Poots and the Pherans, which includes two robes placed on top of the other. The Pherans worn by women usually has Zari embroidery done on the hem line, around the pockets and mostly done around the collar area.
- Females wear a headdress called the Taranga that is stitched to a suspended cap, and it narrows down at the back, towards the heels. Patterns in Brocade style can be found on their long sleeves.
- The Pheran is usually worn alongside a red coloured head cloth known as the Kasaba. It is stitched like a turban and is pinned together by silver brooches and ornaments. A pin-scarf hanging from the Kasaba goes towards the shoulder.

General Features of Lucknowi Costume (chikankari)

- Chikankari is a very delicate and intricate shadow work type of embroidery. Initially, the embroidery was done using white yarn, on colourless muslins known as tanzeb.
- The word Chikankari has been derived from a Persian word Chakin or Chakeen, which means creating delicate patterns on a fabric.
- These are generally dull coloured, with light colour embroidery

General Features of Punjabi Costume

- The traditional dresses effortlessly merge style elements, comfort and colours.
- Salwar and Kameez is traditional costume of Punjab. They are usually of bright, dark and brilliant colours and are matched with a matching or contrasting coloured Dupatta.
- It is mostly shiny in texture and draped over the head as a part of their religious custom. Women also wear shawls which also come with beautiful Phulkari designs.
- Kameez are generally not loose and the fit to the skin closely.
- The length of the kameez is typically short, in comparision to other states.

Brief introduction of Algorithms and classifiers used :

1.K means clustering

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure is as follow:
Algorithmic steps for k-means clustering
Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.
1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster centre whose distance from the cluster center is minimum of all the cluster centers..
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'ci' represents the number of data points in ith cluster.
5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step (3)

2.KNN

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.

Algorithm

A case is classified by a majority vote of its neighbours, with the case being assigned to the class most common amongst its K nearest neighbours measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbour.

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

3.Bag of Visual Words (BOW)

Bag-of-words is a technique used in the field of computer vision for image classification. It draws a parallel between images and text documents. Image features are compared to words. In text classification, a vector is created for the counts of words. Similarly, in image classification, a vector called the bag-of-keypoints contains the dictionary of occurrence of local image features. Csurka et. al. presented a novel method for image classification. This process has been designed to maximize classification accuracy and minimize computational cost. The steps involved in this process of categorization are as follows:

1. Detection And Description of Image Patches: SIFT features have been selected for detection and description of image patches as they are robust to image rotation, scaling and illumination. They are vectors of 128 dimension.
2. Assignment of Patch Descriptors: To reduce the large number of descriptors involved in computation, K-means clustering is performed. Dictionary is created where the nearby descriptors are attached to the cluster centers. Whenever a new query arrives, each of its descriptors are assigned to the existing cluster centers. The dictionary of keypoints should be large enough to distinguish relevant changes in the image parts but not so large that noise starts affecting the categorization.
3. Construction of Bags of Keypoints: The next step is to create a histogram for each image. This is created on the basis of number of occurrences of various keypoints(words) in the entire image.
4. Application of Multi-Class Classifier: In the last step, a multi-class classifier is applied. The bag of keypoints are treated as features and

categories are assigned to the images. Here, Support Vector Machines(SVM) has been used for classification using a linear kernel. SVM uses the concept of finding the hyperplane which maximizes the margin between the support vectors.

4.LinearSVC

Linear Support Vector Classification.
Similar to SVC with parameter kernel='linear', but implemented in terms of liblinear rather than libsvm, so it has more flexibility in the choice of penalties and loss functions and should scale better to large numbers of samples.
This class supports both dense and sparse input and the multiclass support is handled according to a one-vs-the-rest scheme.

5.HumanDetection

OpenCV ships with a pre-trained HOG + Linear SVM model that can be used to perform human detection in both images and video streams.First, we make a call to hog = cv2.HOGDescriptor() which initializes the Histogram of Oriented Gradients descriptor. Then, we call the setSVMDetector to set the Support Vector Machine to be pre-trained human detector, loaded via the cv2.HOGDescriptor_getDefaultPeopleDetector() function.we then handle loading our image off disk and resizing it to have a maximum width of 400 pixels. The reason we attempt to reduce our image dimensions is two-fold: Reducing image size ensures that less sliding windows in the image pyramid need to be evaluated (i.e., have HOG features extracted from and then passed on to the Linear SVM), thus reducing detection time (and increasing overall detection throughput).Resizing image also improves the overall accuracy of our human detection (i.e., less false-positives).

Actually detecting huamns in images is handled by making a call to the detectMultiScale method of the hog descriptor. The detectMultiScale method constructs an image pyramid with scale=1.05 and a sliding window step size of (4, 4) pixels in both the x and y direction, respectively.
The size of the sliding window is fixed at 64 x 128 pixels.The size of the sliding window is fixed at 64 x 128 pixels, as suggested by the seminal Dalal and Triggs paper, Histograms of Oriented Gradients for Human Detection. The detectMultiScale function returns a 2-tuple of rects , or the bounding box (x, y)-coordinates of each person in the image, and weights , the confidence value returned by the SVM for each detection.

A larger scale size will evaluate less layers in the image pyramid which can make the algorithm faster to run. However, having too large of a scale (i.e., less layers in the image pyramid) can lead to humans not being detected. Similarly, having too small of a scale size dramatically increases the number of image pyramid layers that need to be evaluated. Not only can this be computationally wasteful, it can also dramatically increase the number of false-positives detected by the human detector. That said,the scale is one of the most important parameters to tune when performing human detection.
we make initial bounding boxes and draw them on our image after this.

However, for some images we notice that there are multiple, overlapping bounding boxes detected for each person.
In this case, we have two options. We can detect if one bounding box is fully contained within another (as one of the OpenCV examples implements). Or we can apply non-maxima suppression and suppress bounding boxes that overlap with a significant threshold — and that's exactly what we do.
After applying non-maxima suppression, we draw the finalized bounding boxes on image and use this as input image to to our futher classifiers.

6.FaceDetection

Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

OpenCV comes with a trainer as well as detector.Here we will deal with detection. OpenCV already contains many pre-trained classifiers for face, eyes, smile etc. XML file 'haarcascade_frontalface_default.xml' contains them.

1. First we need to load the required XML classifiers. Then load our input image (or video) in grayscale mode.

```
face_cascade =
cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
```

1. Now we find the faces in the image. If faces are found, it returns the positions of detected faces as Rect(x,y,w,h).

```
faces = face_cascade.detectMultiScale(gray, 1.3, 5)
for (x,y,w,h) in faces:
```

```
cv2.rectangle(img,(x,y),(x+w,y+h),(255,0,0),2)
roi_gray = gray[y:y+h, x:x+w]
roi_color = img[y:y+h, x:x+w]
```

7.CNN

A convolutional neural network is like a usual neural network. It also comprises of neurons, biases and weights. They take in an input, compute its dot product and optionally treat it with a non-linearity. Traditionally, a Convolutional Neural Network has convolutional layers followed by fully-connected layers. Its architecture is designed to take benefit of the two-dimensional nature of an image. It has a loss function on the last layer i.e. the fully connected layer. It is a feedforward neural network which is inspired by the animal visual cortex. It was created because the traditional networks were not invariant to linear translation. The types of layers present in a Convolutional Neural Network are as follows:

- INPUT
  Input Layer contains the raw values of image pixels and three layers for Red, Green and Blue channels.
- CONV
  Convolution layer computes the output of neurons that are locally connected to the input. They calculate the dot product between the pixel values of the connected input and their weights.
- RELU
  This stands for Rectified Linear Unit. This layer introduces non-linearity in the architecture.
- POOL
  The pooling layer is responsible for down sampling across height and width of the input matrix

- FC
  The fully connected layer appears at the end of the network. It is used for computing scores for all the classes. The number of outputs at this layer correspond to the number of categories for classification. As the name suggests, every neuron of this layer is connected to every neuron of the previous layer.

Along with this we have developed modules namely:

- **Collar Detection**
- **Forehead silver jewellery Detection**
- **Dress Length Ratio Detection**
- **Upper Body Dress Color Detection**

CURRENT WORK

Approach 1: Color histogram as feature and LinearSVC Classifier.
In this approach we form color histogram from the image pixels and give it to our classifier(LinearSVC,KNN) as a feature that defines the image.
Accuracy obtained via color histogram quantization is 43% and 45% respectively.

Approach 2: Bag of Visual Words as feature vector and LinearSVC Classifier, KNN Classifier.
We first obtain (100) cluster Centers using K-Means Algorithm. We then form Histograms of features of image using these centers and thus pass this bag of visual words obtained to train our classifier.
Accuracy obtained via BagOfVisualWordsApproach is 46%.

Approach 3: Human Detection + Bag of Visual Words
We thought that above approach would be extracting unnecessary features of background so we first applied human detection on our dataset,extracted this as new image input for finding descriptors and keypoints but it did not make a significant difference and accuracy obtained was 48%.

```
chetna@chetna-HP-Pavilion-dv6:~/pythonWorkspace$ python human_detection_bow1.py
/home/chetna/anaconda3/lib/python3.6/site-packages/sklearn/cross_validation.py:4
4: DeprecationWarning: This module was deprecated in version 0.18 in favor of th
e model_selection module into which all the refactored classes and functions are
 moved. Also note that the interface of the new CV iterators are different from
that of this module. This module will be removed in 0.20.
  "This module will be removed in 0.20.", DeprecationWarning)
['punjab', 'kashmir', 'lucknow', 'gujarat', 'haryana']
check point 1
check point 2
check point 3
check point 4
check point 5
For Method 1 [ 0.56603774  0.52830189  0.48076923  0.5         0.50980392]
Accuracy: 0.52 (+/- 0.06)
For KNN Classifier  [ 0.39622642  0.49056604  0.53846154  0.55769231  0.43137255
]
Accuracy: 0.48 (+/- 0.12)
check point 6
```

All above codes use **cross validation** method to fetch accuracy.

Approach 4: State Classification by individual unique feature detection as a module

Since above approaches did not give us satisfactory results we decided to work on individual feature which could uniquely identify a state and thus help in predicting the state that is represented by image.

Features and States

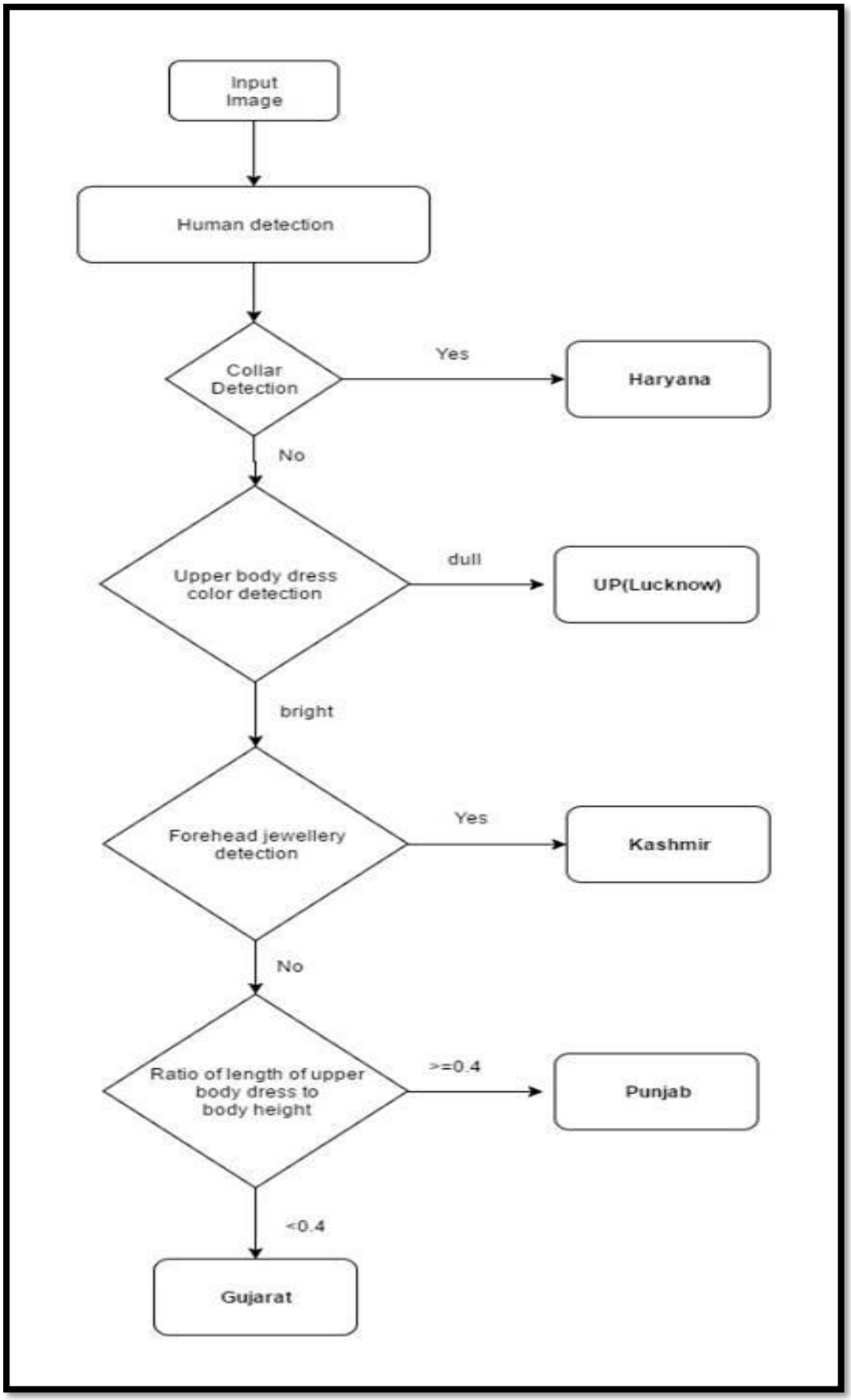
Module 1 : collar detection : Haryana
Module 2 : upper body dress color detection: Lucknow (light, dull color, remaining states have bright color dresses)
Module 3 : silver jewellery covering major forehead : Kashmir
Module 4 : ratio of length of upper body dress to body height : punjab (patiala with short kurta) else: colorful Ghaghra choli : Gujarat

Explanation:

Digging deeper into dataset for states ,we found that each state had one unique feature that could help detect that state from others (*One Vs rest*) and thus in order to learn various techniques and add on our own logics we decided to work on these features. For predicting state of given input we decided to follow the steps shown in below flow chart.

We first extract human from the given image using humanDetectionAlgorithm which uses harcascade and pass this as our image to work on to different modules. we did this so that we can find our required region of interest(ROI) for individual modules easily.We then apply our collar detection algorithm which is built over face detection algorithm.This will help us check whether the dress in image is a traditional haryana costume. If not then we pass our image through upper body dress color detector which tells whether the dress is of dull color or bright color. This helps predict lucknow , as remaining states namely gujarat,punjab and kashmir have bright colored traditional costumes. If image is not classified still , we pass it through our next detector which is silver jwellery on forehead detector. This helps in uniquely identifying kashmiris as only their dress code involves forehead jewellery that extends on major part of forehead.Haryanvi costume also includes forehead jewellery but since we already checked for haryana using collar detector , we need not worry about it.The states which now remain are Punjab and Gujarat,Since both wear bright colorful dress therefore we classify punjabi traditional costume, which involves patiala and short kurta using dress length ratio to extract this feature. If this doesn't predict the state, we classify it as gujarat.

Project Approach :



Approach 5: Convolutional neural network using ALEXNET

Accuracy obtained by using CNN is : 95%

```
chetna@chetna-HP-Pavilion-dv6:~/pythonWorkspace$ python cnn_3.py
Using Theano backend.
['punjab', 'kashmir', 'lucknow', 'gujarat', 'haryana']
['punjab', 'kashmir', 'lucknow', 'gujarat', 'haryana']
------------check point 1------------
------------check point 1.1------------
Timestamp 1: 2017-04-18 23:05:36
Timestamp 1.1: 2017-04-18 23:13:55
------------check point 2------------
------------check point 2.2------------
Timestamp2: 2017-04-18 23:13:55
Timestamp 2.2: 2017-04-18 23:14:36
------------check point 3------------
<class 'numpy.ndarray'>
(246, 4096)
------------check point 4------------
<function accuracy_score at 0x7f9fee5e4ea0>
------------check point 5------------
<class 'numpy.ndarray'>
(24, 4096)
[0 0 0 0 0 1 1 1 0 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4]
[[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4]
 [0 0 0 0 0 1 1 1 0 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4]]
acurracy is:  0.9583333333333334
------------check point 6------------
chetna@chetna-HP-Pavilion-dv6:~/pythonWorkspace$
```

PROGRESS IN CURRENT WORK & CHALLENGES FACED

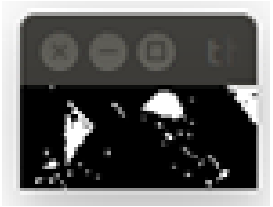
Module 1: Collar Detection

Training the model:

1. Collected sufficient images with collar and without collar.
2. For each training image, we have done:

I. Face detection:

II. Extracting Region of Interest and apply thresholding:



III. Extract SIFT descriptors from ROI

3. Stacked all descriptors vertically
4. Applied K-means clustering.
5. We have created histogram for each image and assigned them to the cluster they are nearest to.
6. Trained KNeighborsClassifier model.

Prediction
1. Extract ROI from test image
2. Also extract SIFT descriptors from ROI of the test image.
3. Trained model will classify the test image (collar or non-collar).

Module 2 : Upper Body dress color detection:

1.Detect and extract upper body image.
2.Convert it to HSV and HSL individually .
3.Accordingly extract ROI (Region of interest) : a rectangular patch having same center as that of upper body detected but relatively smaller length and width.
4.From this ROI we calculate mean of the pixel values using numpy's mean() method and compare it with our requirements which is :

FOR HSV:
Saturation >= 45
Value >=80
Hue = no bounds

For HSL:
Saturation <= 0.4
lightness >= 70%
Hue = no bounds

If it satisfies any of the above criteria , we predict it as Uttar Pradesh (Lucknow) Since from our research we found that above combination gives us light and dull color shades which best represents Lucknowi kurti's.

Challenges faced :
We first used the approach of detecting face and then detecting ROI which failed for cases where face detection algorithm did not predict face and thus we migrated to upper body detector which sometimes fails for cases where we don't have complete human in picture and have just the upper body dress.Thus, we are working on this part,rest code that follows up is completed.

Module 3: Detecting Kashmiri Forehead Jewellery

Women in Jammu and Kashmir wear huge circular ear rings locally called kundalas, large anklets called nupura while in the Rajouri region a silver cap known as chaunk phool is worn on the head under the veil.
Here we focus on predicting whether the lady is wearing "chaunk phool" or not (mainly silver color).
In the dataset collected, kashmiri ladies are wearing silver chaunk phool

Steps for detecting the jewellery:

1. Detect face in the images using “haarcascade_frontalface_default”
2. This will give a square containing face of the lady; the ornament is just above the face.
3. Take a small rectangle above the face; depending upon the size of square, the size of rectangle is:

   Length of rectangle=1/4th of square
   Width of rectangle=width of square
   Convert the image to HSV
4. Take a range of values in HSV which represents silver colour:

   H in range: 0 to 180
   S in range: 0 to 21
   V in range: 125 to 255
5. Convert all the silver pixels to white and rest of the pixel values to black
6. In the rectangle if more than 30 percent of the pixels are white then we can say that the lady is wearing chaunka phool, else she is not.
7. Using this we can detect wearing head ornament is present in the image of or not.

Challenges faced:

In some of the images the ornament is not purely silver, and it has golden touch, at that time the ornament does not comes under the range specified above.

Face detection algorithm detects only frontal face and not side faces.

FUTURE WORK:

- We plan to complete all modules with good efficiency and accuracy and then include more states like Kerala , Mizoram etc
- We are also aiming to overcome challenges that we are currently facing in each module.
- After accomplishing above mentioned things we would like to develop two more modules if time permits which are :
- Saree Detection along with border color Detection.
- Dress Fitting.

FUTURE WORK CHALLENGES

- Overcoming the challenges we are facing in current modules require us to try several methods which is Brute Force approach with a lot of algorithms which we might have not looked into but do exists. Its a trial and hit method which does not guarantee anything.
- Reason why we are so selective with states is because while studying states we realised that most south indian states have similar traditional dresses which are difficult to classify among themselves as they have off white saree with some border. We also found that west bengal dress code significantly differs from kerala dress only in border color (Red Vs Golden respectively).Thus, if time permits , we would certainly like to build saree border detector and expand states to be classified.
- Also, we saw that kashmiris wear loose dresses whereas Punjabi's prefer tight body fit dresses.Thus, we would like to work in this field also and predict whether dress wore is loosely fit or tight fit, which in itself is a complex task and we have little clue on how to proceed.