Footprint Classification

Swetha Mandava
Electrical and Computer Engineering
Pittsburgh, PA 15213
mmandava@andrew.cmu.edu

1 Left/Right Footprint Detection

Suppose the police has collected a bunch of footprints. The first task is to help with detecting whether a new one belongs to left or right foot. The result is a list of the prediction on the validation data with your trained classifier, where each row is a integer "1" if the the prediction is "right foot" or "2" otherwise.

1.1 Data Processing

Gaussian Blur: gaussian blur is typically used to reduce image noise in images. Since the footprint images had noise included, I used a gaussian blur to smoothen the image.

Discrete Cosine Transform: expresses data as a sum of cosine functions. In particular it is useful while separating different frequency components in an image. Since noise usually has a high frequency, this can be used to separate and use the stable low frequency components. DCT also has a strong energy compaction property. 2 dimensional DCT returns an array the same size as the image with the upper left corner having low frequency components: increasing towards the right and down. I truncated 40*40 block from upper right block of DCT from every image for our purpose.

Normalizing: Normalizing data before computing principal components is a an important exercise as PCA tries to maximize variance. Without normalizing, we would wrongly weight the data with high standard deviation highly and hence get undesired behaviour.

Principal Component Analysis: is an orthogonal transformation from a set of correlated variables into principal components of uncorrelated variables.By picking m < n principal components that maximize the variance of the data in low dimensional space, we can reduce dimensionality by losing minimum possible information. I used PCA for dimensionality reduction of the flattened 40*40 upper left block obtained from DCT into a 30 dimensional array for faster classification efficiency.

1.2 Machine Learning Techniques

Support Vector Machines with RBF kernel: Support vector machines are supervised classifiers ideal for high dimensional space classification. RBF kernels were used even though linear classifiers are faster as it generally gives much better performance. RBF kernels represent feature vectors in some input space defined as below:

$$K(X, X') = exp(-\gamma||X - X'||^2) \tag{1}$$

1.3 Parameter Setting

Grid Search Cross Validation: Hyper parameters are those parameters whose values are not learned with the estimator but are input constraints on the estimator. Grid search cross validation is used to

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Table 1: Results for Left/Right Footprint Detection

Data Set	Accuracy(%)
Validation	90.2
Testing	89.95

search a hyper- parameter space to obtain the best hyper parameters for a good cross validation score. We obtained C=10 and alpha = 0.005 for SVM rbf kernel for optimal classification.

1.4 Results

The results on test and validation dataset have been summarized in Table 1.

1.5 Other Approaches

Convolutional Neural Networks

Implemented a CNN consisting of one convolution layer with 20 filters of 5*5, a reLu Layer, a max pooling layer with stride 2 of size 2 followed by fully connected layer of size 2 and a softmax layer on MATLAB. I used a learning rate of 0.001 with a batch size of 32. However, my accuracy plateaued at 80.7 after which it starts to over fit to the data. Adding layers and augmenting data had not much effect on the accuracy. One possible solution could be to use pre-trained networks to obtain extracted features which could then be fed into a support vector machine or to change the convolution layer input sizes. Another possible approach could be to use drop outs to reduce over fitting.

2 Footprint Matching

Now our police collects quite a few footprints from the recent crime scenes. Each of them is believed to belong to one of the 1,000 recidivists and fortunately their footprints are well stored in the database. Now, your task is to tell the owner of those footprints. Note that different footprints can be of the same person. The result is a list of the prediction on the test data with your approach, where each row is an integer indicating the owner of the footprint (ranging from 1 to 1,000).

2.1 Data Processing

Augmenting of Data: Since the training image is clean unlike the testing data and gaussian noise of 0 mean and 0.2 variance has been added to the images. Since images in training set are all right foot images in the upright direction, they have been rotated in intervals of 45 degrees and also flipped to resemble the left foot.

Speeded up Robust Features: is a local feature descriptor that finds corner(interest) points using determinant of Hessian blob detector. SURF features are scale invariant as it finds interest points on difference of Gaussians. In order to also be rotation invariant, the haar wavelet responses in both x and y directions around the interest point are computed.For our purpose, surf descriptors of 64 dimensions each are computed for the strongest 60 points.

2.2 Machine Learning Techniques

Bag of Words: is a simplified model mainly used in language processing or computer vision. In this model, the major concern is of frequency and not the spatial location of the descriptor. A dictionary of all the descriptors found in the training data set is formed along with their labels. During testing, we find the closest match to each descriptor in an image and assign label to the most visible class.

In our model, let N be the number of images with each having 60 SURF descriptors. In the bag of words model, the training data now has 60N * 64 dimensional array where each row has a label that describes which image that particular descriptor belongs to among our 1000 samples.

Table 2: Accuracy vs K

K	Accuracy(%)
1	19.8
5	13.3
50	11.1

Table 3: Results for Footprint Matching

Data Set	Accuracy(%)
Validation	19.7
Testing	19.8

K Nearest Neighbor classifier: is a supervised model that simply returns a class membership from the majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. If k = 1, then the object is simply assigned to the class of the training descriptor closest to itself. The distance metric used to compute the closeness is the euclidean distance.

During testing, we compute the labels of each descriptor in an image using KNN. The class membership of the image would therefore be the mode of the labels of all descriptors in an image.

2.3 Parameter Setting

K for Nearest Neighbour Classification:

I have tried different Ks for nearest neighbor classification and obtained accuracies on the validation data set are summarized in Table 2. We can see that the best accuracy was obtained for K=1. Hence, I used 1NN for footprint matching with our dataset.

2.4 Results

The results on test and validation data set have been summarized in Table 3. The results obtained are 19700% above chance and can be possibly increased by using boosting algorithms like Adaboost on our 1NN weak learner.

2.5 Other Approaches

Histogram Vectorization followed by SVM

After computing 60N * 64 bag of words model, KMeans clustering is done on the training data to cluster it into 200 clusters. Upon clustering, 60 SURF descriptors from each image are clustered into the above obtained 200 clusters to compute a histogram frequency. After normalizing these histograms from all images, they are passed into a support vector machine classifier for multi class classification.

DCT followed by SVM

Low frequency DCT components have been obtained from images (similar to Problem 1) and the features have been sent to multi class SVM classification upon dimensionality reduction using PCA.

3 Pattern Discovery

During the past few years, the police has collected a bunch of footprints without annotations. Due to various reasons, those data are cropped, twisted or blurred. However, they are still very important hints and believed to be related to some undecided cases. Could you please find some patterns in those data, so as to assist the police with further investigation? You can apply whatever you think appropriate to attack this open-ended question. One possibility is to use cluster techniques, to see if there are meaningful grouping structure.

Table 4: Results from failed Approaches on Validation Set for Footprint Matching

Approach	Accuracy(%)
Histogram Vectorization followed by SVM	0.8
DCT followed by SVM	0.1

3.1 Analysis

Clustering of images can be done by finding patterns in the shoe prints. For clear understanding, let us consider a data set of shoe prints that contain only circles or lines. We can therefore perform clustering using two features 1. Number of lines in an image and 2. Number of circles in an image. Drawing from this understanding, we can cluster the local descriptors in an image into bins. We can then obtain histograms of descriptors in each image to cluster the images.

3.2 Machine Learning Techniques

K Means: partitions all data into k clusters in such a way that all observations that belong to the same cluster has minimum variance. This can be used to cluster similar descriptors together and later to cluster similar shoe prints together.

3.3 Algorithm

`Step 1`

Compute SURF features of all the images as they are scale invariant.

`Step 2`

Perform Kmeans clustering on all the descriptors to obtain bins with similar descriptors grouped together.

`Step 3`

Obtain histograms of each image depending on how many descriptors belong to each bin obtained in Step 2.

`Step 4`

Perform K means clustering on normalized histograms of images obtained in step3 to find meaningful patterns in the data.

3.4 Results

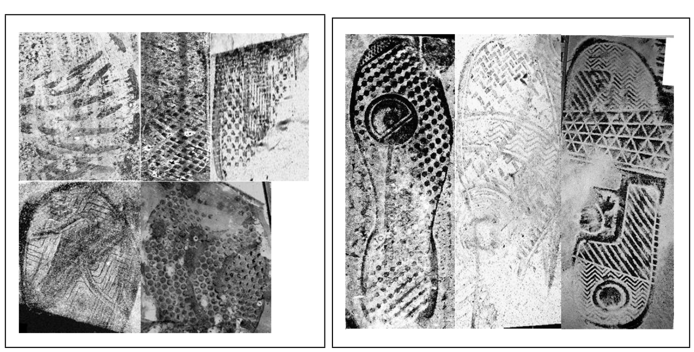

Figure 1: Result images in clusters

References

[1] H. Su, D. Crookes, A. Bouridane, and M. Gueham, *Local image Features for Shoeprint Image Retrieval* British Machine Vision Conference 2007

[2] Alex Krizhevsky Ilya Sutskever Geoffrey E. Hinton*ImageNet Classification with Deep Convolutional Neural Networks* Advances in Neural Information Processing Systems 25 (NIPS 2012)

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

[3] P. D. Chazal, J. Flynn, and R.B. Reilly,*Automated processing of shoeprint images based on the Fourier transform for use in forensic science* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp. 341-350, Mar. 2005.

[4] Adam Kortylewski, Thomas Albrecht, Thomas Vetter *Unsupervised Footwear Impression Analysis and Retrieval from Crime Scene Data* ACCV 2014, Workshop on Robust Local Descriptors. PDF