Investigating the Use of Synthetic Data and Feature Extraction Methods for Training Support Vector Machines for Classification Tasks on a Rock, Paper, Scissors Dataset

Sonali Mahendran, Johan Ospina

Princeton University, Department of Computer Science, Princeton, NJ

ABSTRACT

Currently, the quality of machine learning vision algorithms are directly tied to the quality and quantity of images collected for training. It can be difficult and inefficient to collect and label these Natural Images (NI). Thus, Computer Generated Imagery (CGI) can pose as a scalable solution to this ubiquitous machine learning need for big data. Recent advances in CGI have allowed the increase in throughput of photorealistic renders.

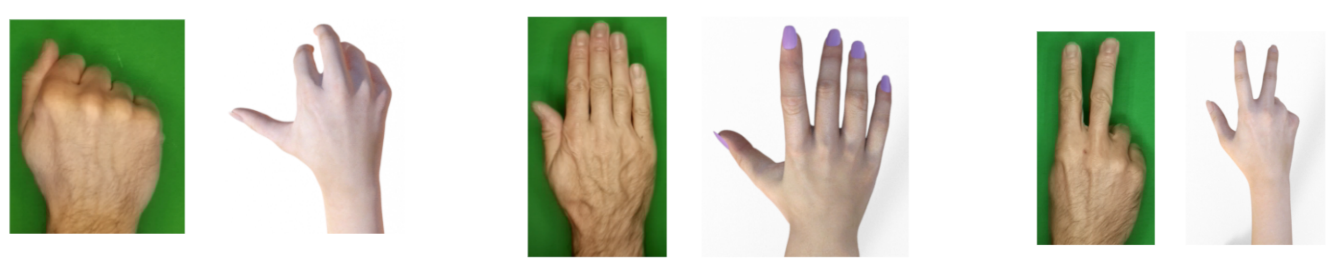
In this project, we investigate whether salient features can be identified in CGI data and used in classification tasks for NI. We focus this investigation on the simple 3-way classification of poses in the classic Rock-Paper-Scissors (RPS) game. This is a simple task for humans, which means there exist some features that uniquely describe each of the poses, regardless of dataset membership.

We present an analysis of the impact of three feature extraction methods on Support Vector Machine (SVM) classifier performance for this RPS classification task. The SVM trains on CGI data, and its performance is evaluated on a test set of NI.

BACKGROUND AND APPROACH

Description of the Data

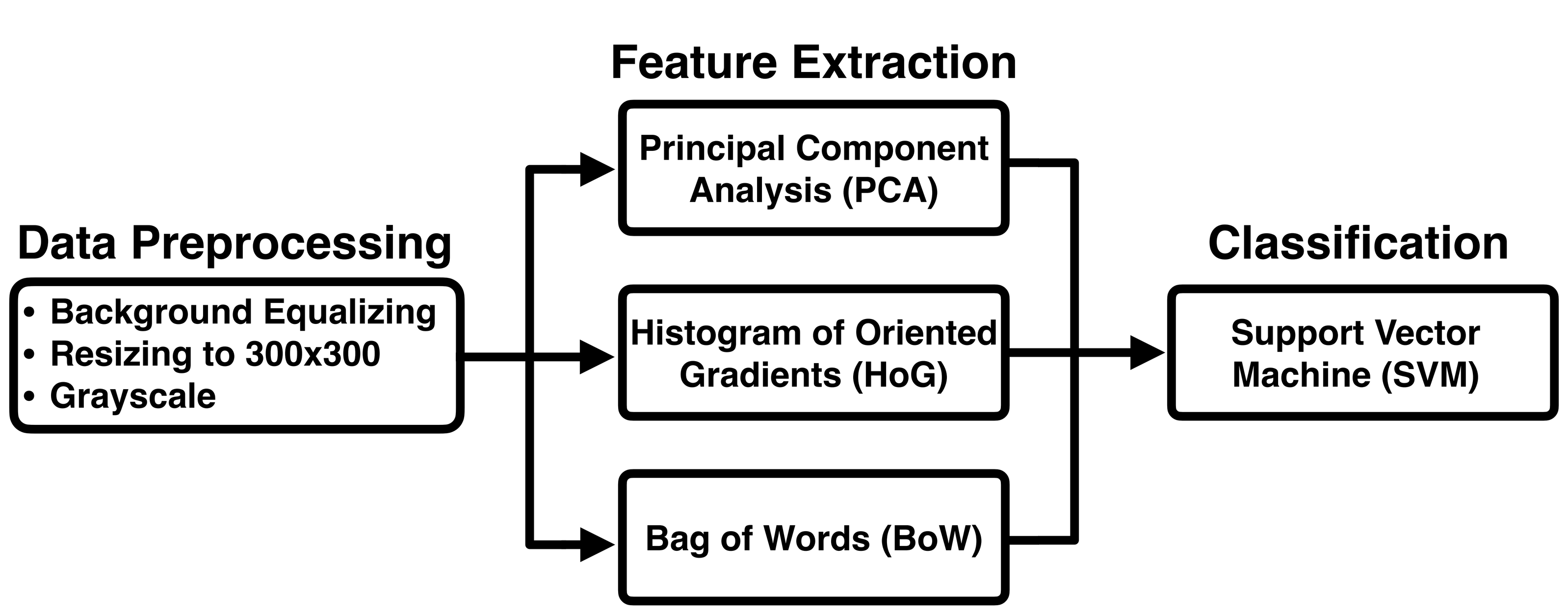
The natural images for this analysis come from Julien de la Bruere-Terreault's Raspberry Pi camera dataset which consists of a total of 2188 300x200 pixel images[1]. The CGI come from Laurence Moroney's CGI-based dataset which consists of a total of 2892 300x300 pixel images[2].



Rock, Paper, and Scissors images from both datasets: NI (Green Background), CGI (White Background)

Methods

We are interested in analyzing how SVM classifier performance varies with respect to different feature extraction methods. Our goal is to analyze how the classifier performs on the task of classifying NI after training on CGI.



ACKNOWLEDGEMENTS

We would like to thank Jimmy Wu for advising us over the course of this project.

REFERENCES

1. Julien de la Bruère-Terreault. Rock-Paper-Scissors Images. https://www.kaggle.com/drgfreemrockpaperscissors#README_rpc-cv-images.txt. Accessed: 2019-11-05
2. Laurence Moroney. Rock Paper Scissors Dataset. http://www.laurencemoroney.com/rock-paper-scissors-dataset/. Accessed: 2019-11-05.
3. Julien de la Bruère-Terreault. 2019. rps-cv.https://github.com/DrGFreeman/rps-cv.
4. Olga Russakovsky. 2019. Lecture 9 object classification. https://www.cs.princeton.edu/courses/archive/fall19/cos429/slides/cos429_fall2019_lecture9_ObjectClassification.pdf. Accessed: 2020-1-12.

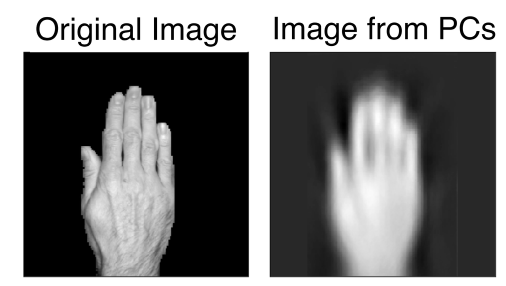
PCA FEATURES

Results

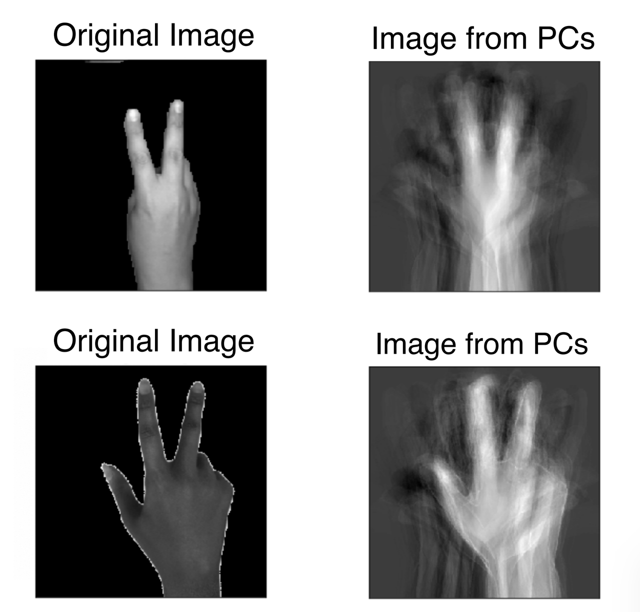
Principal Component Analysis (PCA) is a dimensionality reduction technique used to summarize data as vectors that capture the highest variance[3].

Our results show that PCA is not suitable for this task since there is high variance between these two datasets.

Below, we use the principal components (PCs) generated on the NI to reconstruct a lower dimensional representation of the original image. The reconstruction images shown on the right show that PCs extracted from the CGI dataset cannot be used to effectively reconstruct an NI.



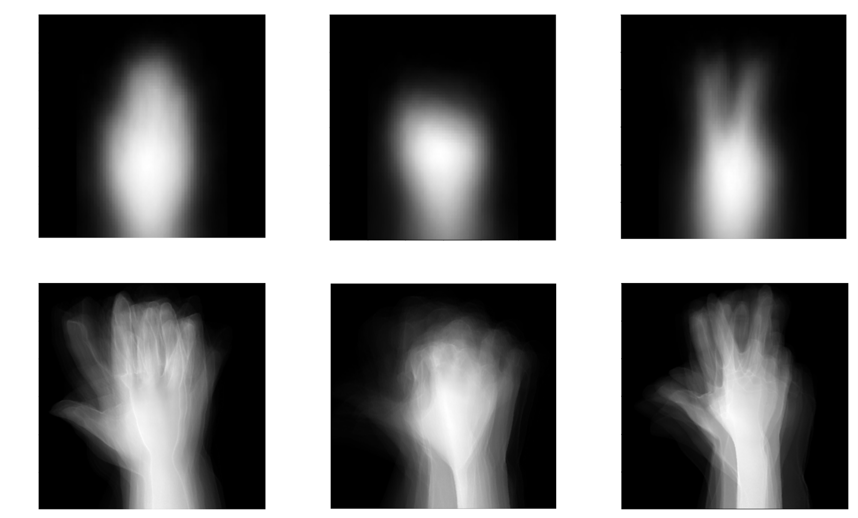
Reconstruction



Reconstruction of an NI and a CGI image from PCA components of the CGI data set

SVM Performance Results

Train Set	Test Set	F1 Score
Natural	Natural	0.981
CGI	CGI	1.0
CGI	Natural	0.425



Average Images for each class (NI top, CGI bottom)

HoG FEATURES

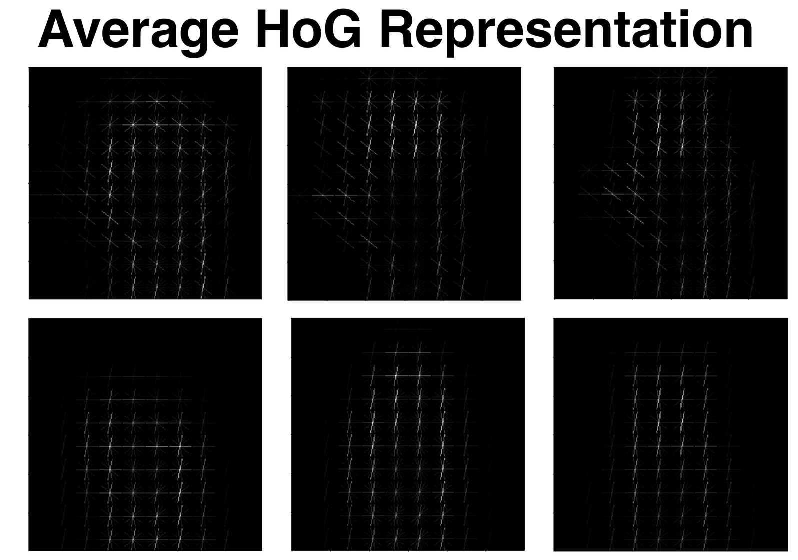
Results

Histogram of Oriented Gradients (HoG) is another dimensionality reduction technique used to summarize data as histograms of gradient directions[4].

The results show that HoG works sightly better than PCA since the features are now capturing structural cues from their respective poses.

The HoG representations on the right show the average HoG features extracted for each class.

Average HoG Representation



CGI (Top Row) and Real (Bottom Row) for Rock, Paper, Scissors (Left to Right)

SVM Performance Results

Train Set	Test Set	F1 Score
Natural	Natural	0.984
CGI	CGI	0.994
CGI	Natural	0.530

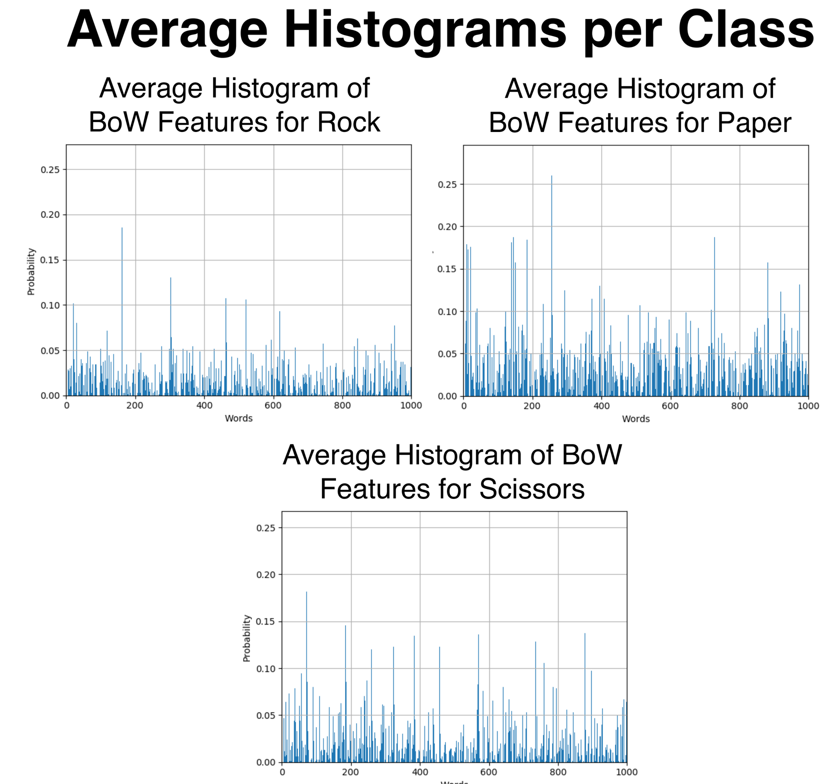
BoW FEATURES

Results

Bag of Words (BoW) is a dimensionality reduction technique used to summarize data as histograms based on extracted features. This is done by analyzing images with the SIFT feature detection algorithm and then clustering the feature descriptors using the K-Means algorithm. This method relies on finding spatially independent features[4].

Our results show that this method achieved the best performance on both datasets due to the fact that it focused more on the features present in the images instead of using the general shape of each image.

Average Histograms per Class



SVM Performance Results

Train Set	Test Set	Train Set Size	F1 Score
Natural	Natural	1785	0.997
CGI	Natural	1200	0.511
CGI	Natural	1500	0.600
CGI	Natural	2100	0.606

All experiments used a vocabulary size of 1000.

Conclusion

Based on these results, we found that the CGI and NI datasets are inherently different. Differences, such as pose shape and lighting conditions, introduced significant variance between corresponding classes across the two datasets. For example, there is significant variance among the images within each class in the CGI dataset. Further, the poses for each class of the CGI dataset are different from that of the NI dataset.

These conditions led the PCA-SVM setup to achieve low precision and recall. The HoG-SVM setup achieved slightly better performance since HoG features captured the structural composition of the images. The BoW-SVM setup achieved the highest performance of all three setups. We believe this is due to the fact BoW relies on features that are not specifically tied to the spatial orientation of pixels. Finally, these results demonstrate that though these two datasets are different, they share key features that can be used to improve a classifier.

As an extension of this work, one could try to design systems that allow him/her to specify the generation of CGI to more closely align with the NI data that will be encountered in downstream tasks.