

# Hierarchical Clustering of Images

Rachel Vancea

Hannah Price

## Introduction/Background:

Image clustering is an important aspect of machine learning in that it allows for pattern recognition in massive amounts of data. Image clustering is essential in the identification of images and patterns such as classifying plants and animals, as well as, processing satellite data. We would like to use image clustering on a large set of images of clothing in order to sort them by color. We will use hierarchical clustering so that the clothes can be sorted on various levels of adherence to the base color.

$$\arg \min_S \sum_{i=1}^k \sum_{X_j \in S_i} \|x_j - u_i\|^2$$

## Methods:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

We have based our approach on the approach of that by Zheng. First, we will run a K-mean clustering algorithm on a single picture. This will allow us to filter out background color information from the photos of clothing. After the background color is filtered out, we will use mean shift clustering on the K-means centers to obtain the correct amount of color clusters in the image. At this step, unlike Zheng, we will use hierarchical cluster to sort the images, so that images are grouped based on color similarity.

## Potential Results:

We hope to sort the clothing images in multiple layers to allow for multiple classifications. For example, we would have a “top” layer of colors within the same family such as: red, orange, yellow in one cluster and a “bottom” layer, going more in depth, where each color is its own cluster.

## Discussion:

Through this project we will gain first hand experience in programming the K-means algorithm on a real data set. Additionally, we will learn and implement the mean shift algorithm. With the combinations of these methods, we are able to effectively cluster clothing based on colors, providing a filter for which consumers can make clothing purchases.

## References:

- » Zheng, X., & Liu, N. (2012). Color recognition of clothes based on k-means and mean shift. *2012 IEEE International Conference on Intelligent Control, Automatic Detection and High-End Equipment*. doi:10.1109/icade.2012.6330097
- » “Mean Shift Clustering.” [Http://Homepages.inf.ed.ac.uk/](http://Homepages.inf.ed.ac.uk/), University of Edinburgh, [homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/TUZEL1/MeanShift.pdf](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/TUZEL1/MeanShift.pdf).
- » Li, J. (2012, March 7). Visualizing High Dimensional Data. Retrieved February 19, 2019, from <http://jamesxli.blogspot.com/2012/03/on-mean-shift-and-k-means-clustering.html>