**Movie Genre Classification using posters**

The aim of this project was to classify the movies according to their genres based on their posters.

**Motivation behind the Idea**

Whenever I visit a multiplex, I am always overwhelmed with so many movie choices that it becomes very difficult to decide which movie to watch. Hence a thought came into my mind that what if I could know the genre of the movies in the multiplex – it would help me decide which movie to watch.

So I decided to develop a classifier which would predict the genre of the movies from their posters.

I started searching for some material in this field and came across this paper *– “Movie Genre Classification via Scene Categorization” by Howard Zhou, Tucker Hermans, Asmita V. Karandikar, and James M. Rehg.* In this paper, they have discussed how they have used movie trailers to classify the movie genres.

**Goal of the project**

So I started with the goal to classify the movie genres from their posters into 3 main categories-

* Action
* Horror
* Comedy

I have developed the project in MATLAB, where it trains on a given input of movie posters and their genres and then it is tested on a set of new movie posters to predict their genres and then the final accuracy is reported.

**Implementation**

**Dataset**

I downloaded 105 posters of each genre from the internet and then labelled them to generate the dataset.

**Feature Extraction**

For the feature extraction, I decided to use the following 3 descriptors –

* GIST
* SIFT (vl\_phow with color information)
* Bag of Visual Words

Here is brief description of these descriptors and how I used it.

**GIST**

It is a holistic representation of a scene. It is used to represent the spatial structures.

I referred to the paper *“Modeling the shape of the scene: A holistic representation of the spatial envelope” by A. Oliva and A. Torralba*. In this paper, they have discussed how GIST descriptors have achieved high accuracy in recognizing the natural scene categories e.g. mountain and coast. However, when the categories of indoor environments are added, its performance drops automatically.

I used the code provided by “*A. Oliva and A. Torralba”* and modified it to fit my project needs.

**SIFT**

I used PHOW features from the image along with color variant information. It was an important descriptor as the posters of the movies have a certain color composition which varies in different genres. In fact, this is the first feature which humans notice to classify the posters. Hence, we could use that feature by extracting the color composition of the image through VLFeat.

Since the number of feature descriptors returned by vl\_phow() is very large, so I am using a uniform subset of these features to prevent the problems from high dimensionality of data.

**Bag of Visual Words**

In this model, the images are represented as vocabulary of words. In this model, first of all the SIFT descriptors are extracted for all the images which are then clustered to generate the vocabulary of the dataset. Then all the images are mapped to that vocabulary to extract a set of visual words which are used to classify them.

It was also an important feature in this project as there are certain objects which are uniform across a particular genre. E.g. in the action movie posters, all kinds of weapons appear in some form in the movie poster. Hence the bag of visual words is very important to classify such kind of features.

I used the code provided by “*Andrea Vedaldi and Andrew Zisserman*” and modified it to fit the project.

**Classification**

After feature extraction and normalization, I used SVM (Support Vector Machines) to classify the data. I used the rbf kernel with the default configurations. I tried with other types of kernels and some configuration parameters of the rbf kernel but the best performance was reached with the default configuration.

In addition to this, I tried to use PCA (Principal Component Analysis) in order to reduce the dimensions of the data. But the accuracy suffered when I used PCA.

**Results**

I used the dataset which consisted of 105 images each for all the 3 genres – comedy, action, horror.

Then I divided it so that 30 images formed the test set and 75 images formed the training set.

With this configuration I was able to achieve an accuracy of around 68%.

Observations

* This accuracy is highly dependent on the dataset. As during cross-validation, the accuracy % varies from 55% to 80% depending on the types of posters.
* I tried to analyze this behavior but was not able to find any concrete reason behind this.

**Improvements/Future Work**

As I already mentioned that the accuracy was highly dependent on the type of posters and was highly unpredictable, hence here is a list of areas I think should help fix this –

* **CENTRIST Descriptors**

I tried to use the CENTRIST descriptors in my code, but they were not giving the desired results. As mentioned in the paper – “CENTRIST: A Visual Descriptor for Scene Categorization” by Jianxin Wu, and James M. Rehg, CENTRIST fulfills the shortcomings of GIST descriptors. They provide good results even for the indoor environments.

* **Color Decriptors**

# Although I used the PHOW features with color information, but there is still some more color information which can be extracted from the images. For this purpose, the color descriptors can be used. I found their reference in “*Color CENTRIST: a color descriptor for scene categorization*” by *Wei-Ta Chu and Chih-Hao Chen.* But I was not able to implement the color descriptors the way they referred in this paper.

# Larger Dataset

# Because of the manual work involved in downloading each poster of the movie and then labelling it and also the limited time, I was able to download only 105 posters for each genre. Had it been a bigger dataset, the accuracy could have improved further.

# Smartphone Application

# I also want to develop a smartphone application for this classifier, where you can take a picture of the movie poster and it would tell you the genre of the movie. As already mentioned before, this would be very useful when you are in a movie theatre and you are faced with so many movie options.