**Paintings with Artist Names using**

**Machine Learning**

By

Deepak Rai (2013062)

Supervisor

Dr. Atul Gupta



**Computer Science Engineering Department**

**INDIAN INSTITUTE OF INFORMATIONTECHNOLOGY**

**DESIGN AND MANUFACTURING JABALPUR**

(2016)

**Acknowledgement**

Although I have put my sincere Efforts in this project based internship however, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to **Dr. Atul Gupta** for his constant supervision and providing necessary information regarding the Internship. Also, motivating me every time to get better understanding of research work, has really helped me a lot.

Last, but never the least, I express my gratitude to my family and friends who made sure that I kept receiving the much needed emotional support.

* **Deepak Rai**

**Contents**

* Chapter 1

- *Introduction 4*

* Chapter 2

- *Report on the present investigation 5*

* Chapter 3

- *Results and Discussions 9*

* Chapter 4

- *Conclusions 9*

* Chapter 5

- *Literature Cited 9*

**1 Introduction**

This is the problem based on the similar problem given on the kaggle.com. This project focuses on classification of paintings based on their painter. Paintings are different from normal images because they contain some intrinsic properties. Research on classification of paintings based on their painter has got very little attention compared to other problems related to images. Researchers have proposed several approaches using digital processing, machine learning and computer vision techniques to tackle the problem and also got some interesting results. Still many questions are partially answered till now such as which features are most useful for classification of paintings by painter? Is it possible to give a general method for classification? How are classification based on painters evaluated? To address some of these questions, this project tries to show that the paintings can be tagged with painter names (currently we do this only with a small set) with reasonably high accuracies.

* 1. **Motivation**

Recently virtual art galleries have come up that allow users to view, buy and sell paintings. The aim is to give users an experience that is close to what one can get in a traditional art gallery. This requires the building of automatic tools for tagging paintings with different classes of tags, searching, novelty and similarity detection, visualization etc. Querying the web for paintings of a particular painter can sometimes lead to irrelevant results. Paintings of others painters, which are not interesting to an user, may also be shown in the results. The problem of storing and querying paintings requires effective mechanisms. The retrieval of images for a query can be sped-up by indexing the database using multiple tags that are often used for querying. These are the reasons that motivate us to study the possibility of identifying the painters based on different features

* 1. **Problem Statement**

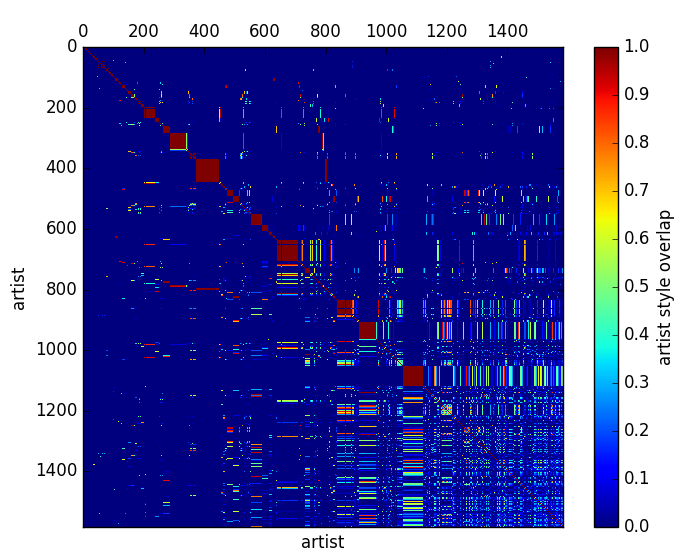
The primary goal is to study and propose an approach for automatic task of painting classification based on painter using different methodologies. The problem of painting classification by painter can be seen as given a set of paintings, whose painters are unknown, predict the painter. It also includes analysis of our own approaches across classifiers and features with existing features and classification methods.

The input to the problem will be a painting painted by painters under consideration. Our task is to give an algorithm which can identify the painter of the given painter. This translates the task in pattern classification problem. This project should answer how features are extracted and which learning methods are used.

**2 Present Investigation/progress**

**2.1 Artist style overlap**

From my given dataset I made an artist-artist graph in which I tried to find the relation between different artistic style among the artist. This style is based on the genre of the paintings they painted for example like imperialism, cubism etc. The plot and relationship between artists look like this.



From the plot we can clearly say that the model is linear.

**2.2 Feature Extraction**

Feature extraction is a special form of dimensionality reduction, without loosing important information of an image, in pattern recognition and in image processing. As input dataset consisting of images themselves is too large to be processed, we need some transformation which captures only interesting points from an image and that important information is called as set of features. Computing set of features from the input data is called as feature vector extraction (also named as feature vector). Feature vectors are designed to perform desired tasks using reduced representation .It is the first operation performed on image which examines every pixel to extract features.

Feature extraction is an important task in image processing because the classifier's performance depends on the feature vector. Several methods are proposed to extract different kinds of features from an image. Commonly used feature vectors for image classification are Color Histogram, SIFT, SURF, HOG, LBP etc.

This time I studied about Texture Feature Extraction LBP.

**2.2.1 Texture Feature Extraction LBP(Local Binary Pattern)**

Local Binary Pattern (LBP) is a texture feature originated from texture analysis. It has been found as a powerful feature for texture classification. It captures intensity statistics in a local neighborhood around a pixel. The basic idea is to capture the local information in an image by comparing center pixel value and pixels surrounding it in a circle. It is computationally efficient and is also invariant to monotonic gray level changes.

Below are the steps to find LBP descriptor for an image

1. Divide the image into segments then divide the examined segment into cells.

2. Every pixel in each cell is then compared to the pixels surrounding it i.e. its 8 neighbors (on its left-bottom, left-middle etc.). We can follow surrounding pixels in any direction i.e. either clockwise or counter-clockwise.

3. If the pixel value of neighbor pixel is greater than center pixel then denote it with 1 and if not then 0. This will give an 8-digit binary number as shown in Figure 1. starting from top-left corner in clockwise direction, which is eventually converted into its decimal representation.

4. For each cell, compute the frequency histogram of each decimal representation occurring in previous step and then normalize it for better comparison.

5. Combine normalized histograms of all the cells that will generate final LBP feature descriptor for an image.

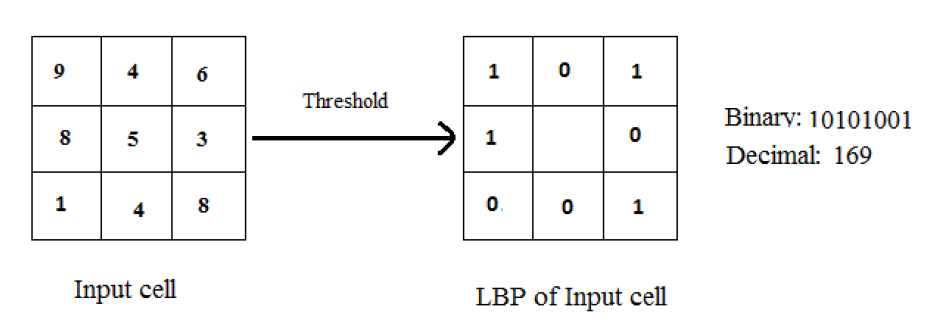
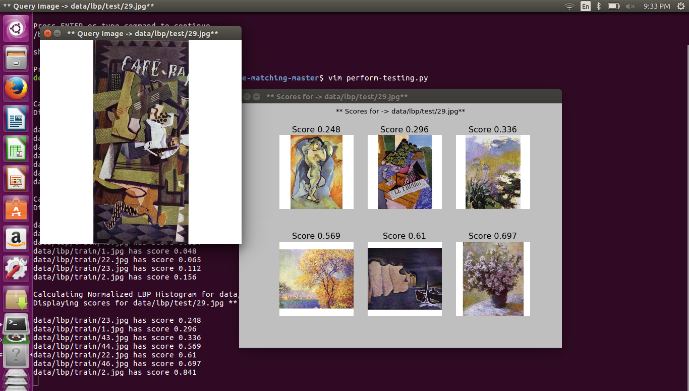
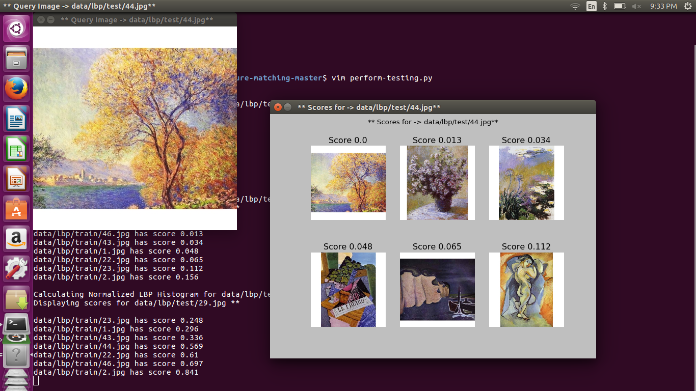


Figure 1.

I have successfully implement this algorithm in on my dataset. Below are the snapshot related to it.



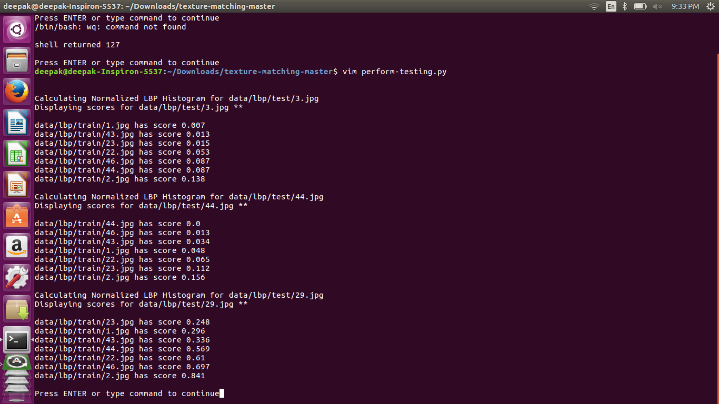
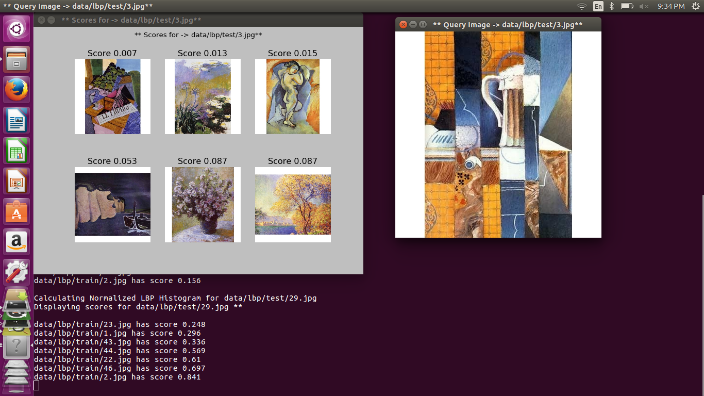
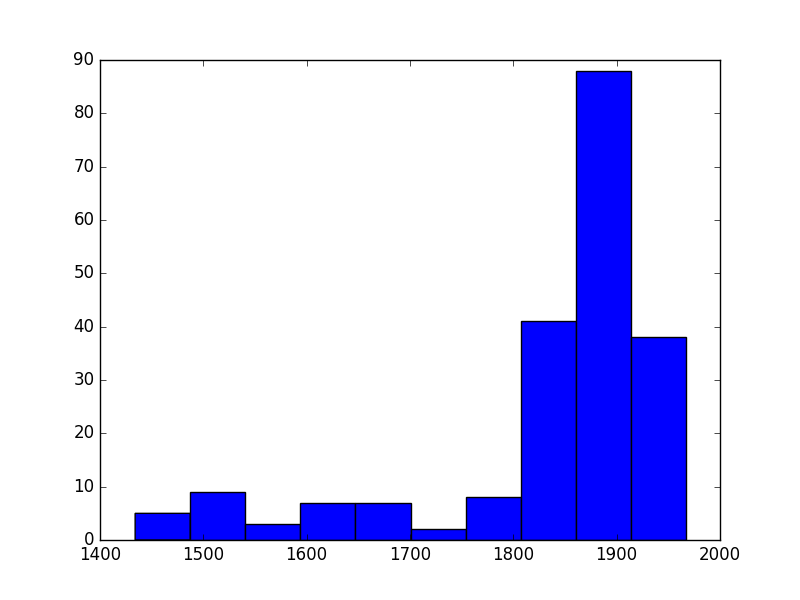


Figure-2

From the figure-2 it is clear that it gives the 6 best possible result that match the given dataset In particular order. It is also seen that from this we can find the painter for the given image given that who painted its respective matching painting.

Also in dataset I got train.info.csv file with additional information on the painting along with the artist name, date, genre ,style ,etc. I tried to find information through it by plotting the graphs one of whose example is given below.



This graph plotted is the relation between year of paining/number of painting in 1000

**3. Results and Discussion**

* The full Data set has been extracted for the project.
* Also while studying the Texture Feature Extraction I implemented on for algorithm that is LBP(local binary pattern)
* I also tried to find the relations between the different attributes of the painting given in the train.info.csv file by plotting different graphs

**4. Conclusion:**

In this work, we demonstrated a simple and efficient approach for texture feature extraction in images using different machine learning concepts and LBP algorithm, for the purpose of identifying the artists of paintings. We will now analyze the performance of our model by modifying the LBP algorithms and compare the result.

**5. Literature Cited**

1.<https://www.kaggle.com/c/painter-by-numbers>

(dataset is also taken from this competion itself)

2.<https://www.kaggle.com/c/painter-by-numbers/data>

3.<http://www.wikiart.org/>

(kaggle has taken the dataset from this website)

4.<http://in.mathworks.com/company/newsletters/articles/creating-computer-vision-and-machine-learning-algorithms-that-can-analyze-works-of-art.html>

5.<http://cs229.stanford.edu/proj2010/BlessingWen-UsingMachineLearningForIdentificationOfArtPaintings.pdf>

6.[https://paws.kettering.edu/~pstanche/paper%20524-ECML%20PKDD%202010-Ivanova%20et%20al.pdf](https://paws.kettering.edu/~pstanche/paper%20524-ECML%20PKDD%202010-Ivanova%20etal.pdf)

7.<http://www.ece.northwestern.edu/~pappas/papers/zujovic_mmsp09b.pdf>

8.<http://gitxiv.com/posts/jG46ukGod8R7Rdtud/a-neural-algorithm-of-artistic-style>

9.<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3170918/>

10.<https://services.math.duke.edu/~ingrid/publications/dating.pdf>

11.<http://www.cs.tau.ac.il/~wolf/papers/genrestyle.pdf>

12.<https://arxiv.org/pdf/1408.3218.pdf>

13.<https://arxiv.org/pdf/1505.00855.pdf>

14.<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.6118&rep=rep1&type=pdf>

15.<https://gautam5.cse.iitk.ac.in/opencs/sites/default/files/12111048_urvesh.pdf>