PROJECT SYNOPSIS

on

Image Similarity Detection For Context Based Search

towards partial fulfillment of the requirement

for the award of degree of

Bachelor of Computer Applications

From

Babu Banarasi Das University

Lucknow

Academic Session 2019 - 2020

School of Computer Applications

Submitted by:

Mohit Kumar Dixit

Sushant Singh

Navneet Kumar

Under Guidance of:

Mr. Sarfaraj Alam

Table of Contents

Image similarity detection for context based search

1. Introduction

Image similarity is used in image registration, content based image retrieval, medical application etc .According to the adopted feature, image similarity detection methods can be divided into two categories, namely, global-feature-based detection methods and local-feature-based detection methods. The global feature of the image refers to the use of one or a few feature vectors to represent whole image content. Common global features include color histograms, texture features, and block features. Because the number of feature points is small, the calculation speed of image content similarity detection based on global feature is usually very fast. However, due to the singularity of its feature selection and the roughness of the description image, the global feature is very susceptible to edits and local transformations. For example, image similarity detection with color histogram as global feature is very sensitive to the illumination of the image.

Image similarity detection helps to better understand images and improve the accuracy and density of recommendations and search results across our production surfaces

2. Objective and Scope:

The future of search will be about pictures rather than keywords.Image similarity detection is to judge the similarity of visual content by matching the images.We partition our image universe into classes of images that are nearly identical as perceived by human observers. While this concept is somewhat subjective, the set of image pairs below will give you an idea of what falls within the NearDup threshold. Notice the image may not necessarily come from the same source photo or have the same background. It can have noticeable geometric distortions or may be a rotational, crop or flip variant.The future of search will be about pictures rather than keywords

1. The ability to learn from images to understand our content is important.
2. To Detect similar type of images
3. Reduced computational complexity by skipping unlikely-to-be-similar pairs of images.
4. Context similarity based seaching and ranking
5. low-level optimization to minimize memory
6. Understanding of context with the images

3. Resources (Hardware & Software)

1. Hardware Requirements

Client Side

Processor	Dual Core or above
RAM	1 GB
Disk space	500 GB
Monitor	15"
Others	Keyboard, mouse, Internet Connection

Server Side

Processor	Dual Core or above
RAM	1 GB
Disk space	500 GB
Monitor	15"
Others	Keyboard, mouse, Internet Connection

2. Software Requirements

Client Side

- Web Browser (Google Chrome, Firefox, IE9 or above)

Server Side

- Web Browser (Google Chrome, Firefox, IE9 or above)
- Windows 7 or above / Linux
- VScode
- SQliteManager
- Python 3.7
- Flask

4. Project Schedule Plan:

The objective of Software Planning is to provide a framework that enables the manager to make reasonable estimates of resources, cost, and schedule. These estimates are made within a limited time frame at the beginning of a software project and should be updated regularly as the project progresses. In addition, estimates should attempt to define best case and worst case scenario so that project outcomes can be bounded.

A Gantt chart is a popular type of chart that illustrates a project schedule. Gantt Chart illustrates the start and finish dates of the terminal elements and summary elements of a project. Terminal element and summary comprise the work breakdown structure of the project.

Task	4Jan-30Jan	31Jan-9Feb	10Feb-12 Mar	13Mar-16 Apr	17Apr-22 Apr	23Apr-28 Apr
Develop project proposal	 27 days					
Analysis		 10 days				
Designing			 30 days			
Coding				 34days		
Unit Testing					 5 days	
Implementation						 5 days

Gantt Chart

5. Project Team:

Instructor	**:**	**Mr. Zaid Kamil**
Mentor	**:**	**Mr. Sarfaraj Alam**
Project Team Members	**:**	**1. Mohit Kumar dixit**
		2. Sushant Singh
		3. Navneet Kumar

6. Process Description

1. **Image collection:** In this module ,we will collect a lot of images from the different sources including the websites and social media. All images will be stored in database of the system.

2. **Image convolution and filtering:** This modules deals with filtering of images using convolution algorithm that modifies the value of each pixel in an image by using information from neighboring pixels.

3. **Pixel maxpooling system:** In this module, we will convert continuous image context attribute into a finite set of time with no loss of image context by using maxpooling.

4. **Scikit mean Shift Algorithm:** This module deals with the clustering algorithm that assigns the datapoints to the clusters iteratively by shifting points towards the mode**.**

5. **Data base management:** Here we will use one of database management sytem from MySQL, Postregsql, Oracle, Microsoft SQL Server for the data management.

6. **Sparse dictionary learning :** Here we will find a sparse representation of the input data as sparse coding in the form of a linear combination of basic elements as well as those basic elements themselves. .

7. **Real time image matching algorithm:** In this module, we will find matching and tracking over image sequences is presented. The goal is that a quadrotor helicopter autonomously approaches a selected target.

8. **User dashboard system:** Here we will create a dashboard for user to check the similarity of the images and search the similar images.

9. **Model creation and compiling**: Here user can choose the model on which he want to work and compile the tasks of the model before the add the module to the module.

10. **Model training:** In this module,we trained the system elements according to a model of the image similarity detection for context based searching.

11. **Evalution and visualization:** Here we will do evalution module by module (task by task).and visualize the whole modules as a system after adding all module.

12. **Image recognition and prediction:** This module helps to detect instances of objects in image and predict the elements and values of the image for the detection of the similarity.

13. **Image uploading system :**This module contain uploading option for images for similarity detection and search option for searching similar type of images.

14. **Dynamic display manager:** This model manages the display dynamically as the user interface which intract the user better.There will be more option for display management.

15. **Preference manager:** This module deals about the preference of the image processing for the similarity detection and search similar type images.

7. Contribution of the student in the project:

1. **Mohit Kumar Dixit**: Image collection,Image convolution and filtering, Pixel Maxpooling,Shift algorithm, Database management.
2. **Sushant Singh:** Sparse dictionary learning, Real time Image algorithm, User dashboard, Model creation , Model training.
3. **Navneet Kumar**: Evaluation & visualization, Image recognition & predi -ction ,Dynamic display manager, preference manager.

8. Conclusion: At the end user do not need to waste time on browsing for trending and best suited tags.

9. References:

https://towardsdatascience.com/image-similarity-detection-in-action-with-tensorflow-2-0-b8d9a78

https://www.hindawi.com/journals/mpe/2018/1917421/

https://medium.com/pinterest-engineering/detecting-image-similarity-using-spark-lsh-and-tensorflc fc939

https://medium.com/ai-techsystems/image-detection-recognition-and-image-classification-with-ma ng-92226ea5f595

https://www.researchgate.net/post/What_is_the_best_image_recognition_algorithm