

Creative Design Visualizer

Aahan Gupta, Yash Kokane, Shubham Khedekar, Sulochana Madachane
Student, Student, Student, Professor(Guide)

Computer Engineering,

K.C College of Engineering & Management Studies & Research (Mumbai University), Thane, India

Abstract: The CDV (Creative Design Visualizer) software is aimed to generate N-number of iterations of designs, patterns, images, concepts from various images/designs/patterns of any category that we need by using combination of designs over it which gives us an image file of any format we need. The idea here is basically to create beautiful creative designs for different purposes. For example, designs for printing on fabric, showcase portraits, website designing, posters, tattoos, decals, catalogue covers, interior designs for walls

IndexTerms - Web Crawling, Image Mining, Image retrieval, Image Clustering.

I. INTRODUCTION

Web scraping, often called web crawling or web spidering, or “programmatically going over a collection of web pages and extracting data,” is a powerful tool for working with data on the web. You can mine data about a set of products, a large corpus of text or quantitative data using a web scraper, get data from a site without an official API, or just satisfy your curiosity.

Image mining refers to set of tools and techniques to explore images in an automated approach to extract semantically meaningful information. The image mining techniques used by miners are image acquisition, object recognition, image indexing and retrieval, image classification and clustering, association rule mining and neural networks. Without the retrieval process image mining is incomplete. Images can be retrieved based on colour, texture, shape, size, the spatial location of image elements. Image indexing, retrieval, image segmentation, feature extraction are the principal tasks in image mining. The categorization is done as content-based techniques

II. LITERATURE SURVEY

During the development of this software, we read through papers. A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques, from Department of Computer Science, University of Wales Swansea SA2 8PP, United Kingdom, 18th April 2008 I-22, 2008. In this paper, Filter bank based methods have been very popular in textural defect detection. The filters can be manipulated and designed in all sorts of directions and scales to decompose textures in order to highlight defects. There are significant and increasing amount of work on color texture analysis, limited work has so far been reported in visual inspection using color texture analysis. The majority of the existing methods decompose the color image into separate channels and process them independently or with limited interactions.

In Design and Development of an algorithm for Image Clustering in textile image retrieval using color descriptors, International Journal of Computer Science, Engineering and Applications (IJCSA) Vol. 2, No. 3, June 2012. This research is concerned with the study and analysis of image mining, image retrieval, image clustering of textile images and the retrieval method design is based on relevance feedback, color layout, scalable color and edge histogram. The algorithm for image clustering is designed based on k-means algorithm and the developed software prototype which allows one to retrieve the images of the textile based on categories such as shirts, t-shirts and color descriptors.

Image mining:

Image mining is an extension of data mining to image domain and it is an interdisciplinary endeavour that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database and artificial intelligence [1]. It reveals valuable information to the users. It deals with the extraction of hidden knowledge, image data relationship, and other patterns not openly stored in the images. The most important areas belonging to image mining are content based image retrieval, model learning as well as object recognition.

Image Clustering and retrieval:

Clustering is a method of combining data objects into different groups [1], such that alike data objects belong to the same group and different data objects to different clusters. Current research increasing interest in digital image searching, classification, identification, management and storage. Some common but significant applications of are person identification, biometric recognition system, natural scene classification for robot vision, segmentation of important topics in lectures.

Retrieval Methods and components

The storing of image data is complicated than textual data stored in traditional databases. Image retrieval techniques provides support for user queries in an effective and efficient manner and just as conventional information retrieval does for textual retrieval. The image retrieval categorization is explained below.

Exact Matching – This type is valid to static environment where the features of image don't change over time. The databases containing industrial drawings are example of such environments.

Low-Level Similarity-Based Searching – In many cases, it is difficult to determine which images please the query and different users might have different needs of images. The user might have diverse predilections under different situations.

High-Level Semantic-Based Searching – In this searching, the notion of similarity is not centred on feature matching and usually results from extended user interaction with the system. At the grainiest level, it identifies two main components of the retrieval problem would be representation and learning.

Similarity Function – The feature representation for each database image, retrieval consists of extracting a set of feature vectors from a query image and relying on a similarity function to evaluate which feature representation best explains those features.

OpenCV:

OpenCV is a free library available for commercial and academic use. It uses Python, Java interfaces. It supports various types of OS. It was designed for computational efficiency with focus on real-time applications. OpenCV is written in optimized C/C++, due to which it can take advantage of multi-processing elements. OpenCL enables to take advantage of hardware acceleration of heterogeneous platforms.

OpenCV usage ranges from like mine inspection, stitching maps on web.

A. Smoothing images: Smoothing is also known as blurring. It's frequently used in image processing operations. Smoothing is used for various reasons but its many done to reduce noise in the image.

To perform a smoothing operation we will apply a filter to our image. The most common type of filters are linear, in which an output pixel's value (i.e. $g(i,j)$) is determined as a weighted sum of input pixel values (i.e. $f(i+k,j+l)$):

$$G(i,j)=\sum_k,l f(i+k,j+l)h(k,l)$$

$h(k,l)$ is called the *kernel*, which is nothing more than the coefficients of the filter.

B. Normalized Box Filter:

This filter is the simplest of all! Each output pixel is the mean of its kernel neighbors (all of them contribute with equal weights). The kernel is below

$$K = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

This is how a 3x3 normalization box filter would look like.

C. Median Filter:

The median filter run through each element of the signal (in this case the image) and replace each pixel with the median of its neighboring pixels (located in a square neighborhood around the evaluated pixel).

D. Changing colorspaces:

There are more than 150 color-space conversion methods available in OpenCV. But we will look into only two which are most widely used ones, BGR \leftrightarrow Gray and BGR \leftrightarrow HSV.

Feature Detection & Description:

Well, it is difficult to say how humans find these features. It is already programmed in our brain. But if we look deep into some pictures and search for different patterns, we will find something interesting. Feature detection is looking for regions of the image which have maximum variation when moved in regions around it.

Beautiful soup 4:

Beautiful soup is a Python library for pulling data out of HTML and XML files. It works with your favourite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

These instructions illustrate all major features of Beautiful Soup 4, with examples. I show you what the library is good for, how it works, how to use it, how to make it do what you want, and what to do when it violates your expectations.

You might be looking for the documentation for Beautiful Soup 3. If so, you should know that Beautiful Soup 3 is no longer being developed, and that Beautiful Soup 4 is recommended for all new projects.

Scrappy Spiders:

Spiders are classes which define how a certain site (or a group of sites) will be scraped, including how to perform the crawl (i.e. follow links) and how to extract structured data from their pages (i.e. scraping items). In other words, Spiders are the place where you define the custom behaviour for crawling and parsing pages for a particular site.

Machine Learning Library:

The Machine Learning Library (MLL) is a set of classes and functions for statistical classification, regression, and clustering of data.

Most of the classification and regression algorithms are implemented as C++ classes. As the algorithms have different sets of features (like an ability to handle missing measurements or categorical input variables), there is a little common ground between the classes. This common ground is defined by the class `cv::ml::StatModel` that all the other ML classes are derived from.

DjangoRESTframework:

Django REST framework is a powerful and flexible toolkit for building Web APIs. REST framework can be used for reasons like Authentication policies including packages for OAuth1a and OAuth2, Serialization that supports both ORM and non-ORM data sources.

Django framework:

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. It's free and open source. Django framework is very fast, it is loaded with features that help take care of various tasks, it is exceedingly scalable that helps to scale to meet the heaviest traffic demands.

PostgresQL:

Postgres is an object-relational database system. PostgreSQL is an open-source descendant of this original Berkeley code. It supports a large part of the SQL standard and offers many modern features: complex queries, foreign keys, triggers, updatable views, transactional integrity and multi version concurrency control.

Keras:

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible

III. METHODOLOGY

In the software we allow the user to provide input via two methods, image input in which the user provides images that user prefers or wants the system to work on or word input in which the software will search for the parameter given over the internet using a spider and fetch images relevant to the parameter for the user to select from. The downloaded images will be stored into PostgreSQL database in the format the images are available.

Feature extraction is detection and isolation of various key points of the images of any image. For extracting, the software uses various OpenCV libraries. For eg, we will be using harrish corner detection algorithm. He took this simple idea to a mathematical form. It basically finds the difference in intensity for a displacement of (u, v) in all directions.

The software will then apply segmentation libraries on the selected image and generate various iteration of images where each image has a set of features applied like smoothening, filtering, color adjust, etc.

Keras is mainly used to cluster images that selected, for the first time use the software will be trained by a ML model that will help the user to identify the good and bad designs. After some use, the software can decide on the currently processed images and help user to identify good designs. The selected good designs will later be displayed as recommended designs to the user.

Once these iterations are developed they will be displayed to the user to pick whichever image fits best to the user's needs and that image can then be exported by the user. The user can then download the image he finds suitable for his needs. The user can also download more than one image by using Zipfile library.

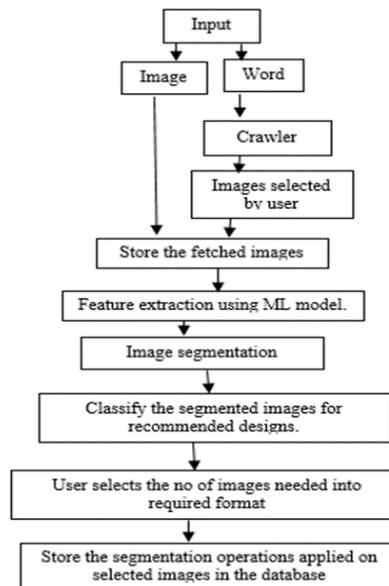


Fig. 1. Operational flow of the software

IV. CONCLUSION

This research is about with study and analysis of images mining, clustering, retrieval. The retrieval method designed based on relevance feedback, color layout, scalable color and edge histogram. The algorithm for image clustering is designed based on k-means algorithm and the developed software prototype allows us to retrieve the images of the textile based on categories such as shirts, t-shirts, pants. The prototype is tested with test data and found successful.

V. ACKNOWLEDGEMENT

We would like to thank our guide Prof. (Mrs.) Sulochana Madachane, Associate Professor, K. C College of Engineering & Management studies & Research, Thane, for initiating me into this field of research and for providing me with the necessary guidance, great encouragement throughout the preparation of this paper. I take this opportunity to express my gratitude to the Staff members, Non-Teaching Staff members and Research Scholars of the Department of Computer Engineering, K. C College of Engineering & Management studies & Research, Thane, for their timely help and encouragement. I record my deep indebtedness to them for their support.

REFERENCES

- [1] S.Balan1 and T.Devi2, "Design and Development of an algorithm for Image Clustering in textile image retrieval using colour descriptors," International Journal of Computer Science, Engineering and Applications (IJCSA) Vol. 2, No. 3, June 2012.)
- [2] <https://www.djangoproject.com/>
- [3] <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- [4] <https://docs.scrapy.org/en/latest/topics/spiders.html>