Entity resolution in sports videos using image to video matching

M.M.D. Maduranga

139171X

Faculty of Information Technology

University of Moratuwa

May 2017

Entity resolution in sports videos using image to video matching

M.M.D. Maduranga
139171X

Thesis submitted to the Faculty of Information Technology,
University of Moratuwa, Sri Lanka for the partial fulfillment of the
requirements of the Degree of Master of Science in
Information Technology.

May 2017

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student | Signature of student

Date:

Supervised by

Name of Supervisor | Signature of Supervisor

Date:

Acknowledgments

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mr. SC Premaratne my supervisor for his guidance and constant supervision as well as for providing necessary information regarding the project & also for his support in completing the project.

I would like to express my gratitude towards my parents & member of University of Moratuwa faculty of Information Technology for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

ABSTRACT

Using both text and image content features, a hybrid image retrieval system for Word Wide Web is developed in this paper. We first use a text-based image meta-search engine to retrieve images from the Web based on the text information on the image host pages to provide an initial image set. Because of the high-speed and low cost nature of the text-based approach, we can easily retrieve a broad coverage of images with a high recall rate and a relatively low precision. An image content based ordering is then performed on the initial image set. All the images are clustered into different folders based on the image content features. Experimental results confirm the efficiency of the system.

Current image search engines on the web rely purely on the keywords around the images and the filenames, which produces a lot of garbage in the search results. Alternatively, there exist methods for content based image retrieval that require a user to submit a query image, and return images that are similar in content. We propose a novel approach named Entity resolution in sports videos using image to video, that is a hybrid of the two methods. Our algorithm first retrieves the results of a keyword query from an existing image search engine, clusters the results based on extracted image features, and returns the cluster that is inferred to be the most relevant to the search query. Furthermore, it ranks the remaining results in order of relevance

Table of Contents

Table of Figure

Chapter 1
Introduction

1.1 Background

"A picture worth thousand words". What does this idiom tell us? Obviously, image play a much more important role than text. The information contained in an image even can't be described by words. However, the search and organization of image databases is often of less importance / attention than text databases. Plainly, there must be difficulties lie ahead for us to solve.

The World Wide Web contains a great quantity of information. The World Wide Web has grown very considerably in size, and is increasing enormously. Due to this enormous information retrieving the information of interest becomes very difficult. A lots of search engines are available for retrieving this information. This information can be text, images or visual information.

If focused on visual information, the World Wide Web contains several kinds of images and other visual information, such as videos, movies, and comic strips, in various formats such as JPG and GIF for still images, and MPG, AVI, and RAM for moving images. Photographs of people, museum collections and paintings, medical libraries, erotic, maps and charts, star photos and movie shots, advertising tapes, greeting cards, logos, sports images, humorous images, and comic strips are good examples of collections on the web.

The most of the search engines available today are textual based i.e., they are all keyword based. One or more keyword is to be provided only then it can find the related web document. This web documents also contains images and most of the image search engine present today are primitive. There are currently two dominant interface types for searching and browsing large image collections: keyword based search, and searching by overall similarity to sample images

The image based search engines available on the web are basically categorized into two categories, one in which images are searched based on certain keywords. These keywords are used to describe which images or which type of images user is looking for. And based on those keywords similar images can be searched. But many a times, user finds their self in a situation when they don't find any keywords to describe it.

The other type of image search engines however searches the images based visual feature. They are based on the concept that the pictures speak a lot than thousands of words. For this, these engines are to be provided with a query image rather than keywords. These image search engines retrieve the images based on the query image. For searching the images based on query image many algorithms have been proposed.

Also the volume of video data is rapidly increasing, more than 4 billion hours of video are being watched each month on YouTube and more than 72 hours of video are uploaded to YouTube every minute, and counters are still running fast. This is attributed to recent advances in technology, cheap digital cameras and the madness of web streaming either for personal or advertising purpose. The majority of available video data exists in compressed format MPEG-1, MPEG-2 and MPEG-4. Extracting low level features from compressed videos, without decompression, is still an open issue and has not been efficiently resolved. Extraction of low level features, directly from compressed domain, is the first step towards efficient video content retrieval. Such approach avoids expensive computations and memory requirement involved in decoding compressed videos.

The proposed solution on this paper is video search engine base on image. This research is combined both textual and visual feature to retrieve result.

1.2 Problem in Brief

Content-Based Image Retrieval (CBIR) was proposed for nearly ten years, yet, there are still many open problems left unsolved. According to some researchers the learning of image similarity, the interaction with users, the need for databases, the problem of evaluation, the semantic gap with image features, and the understanding of images are keys to improve CBIR. In this research, I target understanding of images and the interaction with users, areas.

Motivated by the success of web search engines, many researchers have tried to map image retrieval problems to text retrieval problems, hoping that the proven effective indexing and ranking schemes can be used to handle the scale. The basic idea is to map image features to words. Typically, images are first represented by local features, and then by clustering, each local feature is mapped to a discrete keyword. Such an image representation is called "bag of features," similar to "bag-of-words" for document representation. With this representation, comparing two images becomes matching words in them, and there-fore, text-based search engine technologies can be utilized to reduce the computational and memory cost.

To develop this type of image search engines for a web-scale database, there are still many technical challenges and problems that we need to address

Vocabulary: What kinds of image features should be used? How to map them to words? The most generally utilized method is clustering. Some researchers also adopted a hierarchical clustering method to generate a vocabulary tree. But it is clear that we need to develop some kinds of visual language models to solve the problem.

Long query: The reason why the text search engine is effective is because text queries usually contain only a few words. So, the query-document matching can be con-ducted efficiently by an inverted index. Although images can be represented by "bag-of-features," the retrieval problem is still very different from text retrieval because query-by-example is actually equivalent to using a whole document as a query. So, the search is more like document-to-document matching. How can we deal with this kind of "long query' effectively?

Content quality: Web search engine is effective because it can use link analysis to obtain quality and importance measurement (e.g., PageRank) for web pages. Based on PageRank, an efficient cache can be designed to select a small portion of high-quality web pages and keep them into memory. In most cases, top-k documents can be found in this cache and therefore there is no need to go through the disk index. This strategy significantly improves the index scalability. For images, it is hard to obtain a similar kind of measurement because the links are typically not directly associated with images. Without PageRank for images, it will lead to the lack of efficient cache of index, because we do not know how to select high-quality images and keep them into memory to speed up the search process, and therefore we will not be able to take advantage of many top-k search techniques typically used in web search.

Relevance ranking: The similarity measure between two images is quite different from text. How are image words weighted in computing the relevance? And how to deal with "word proximity" in images?

Distributed computing for web-scale multimedia analysis: Because of the large volume of image data we need to process and index, the system has to be a distributed system, consisting of hundreds of powerful servers. It is inevitably to confront with the challenges as in text-based search engines, such as fault tolerance, data redundant backup, auto configuration, and so on.

1.3 Proposed Solution

Using both text and image content features, a hybrid video retrieval system for Word Wide Web is developed in this paper. We first use a text-based image meta-search engine to retrieve videos from the Web. By identifying the given image, we generate best match text to search videos. based on the text information on the image host pages to provide an initial video link. Because of the high-speed and low cost nature of the text-based approach, we can easily retrieve a broad coverage of Videos.

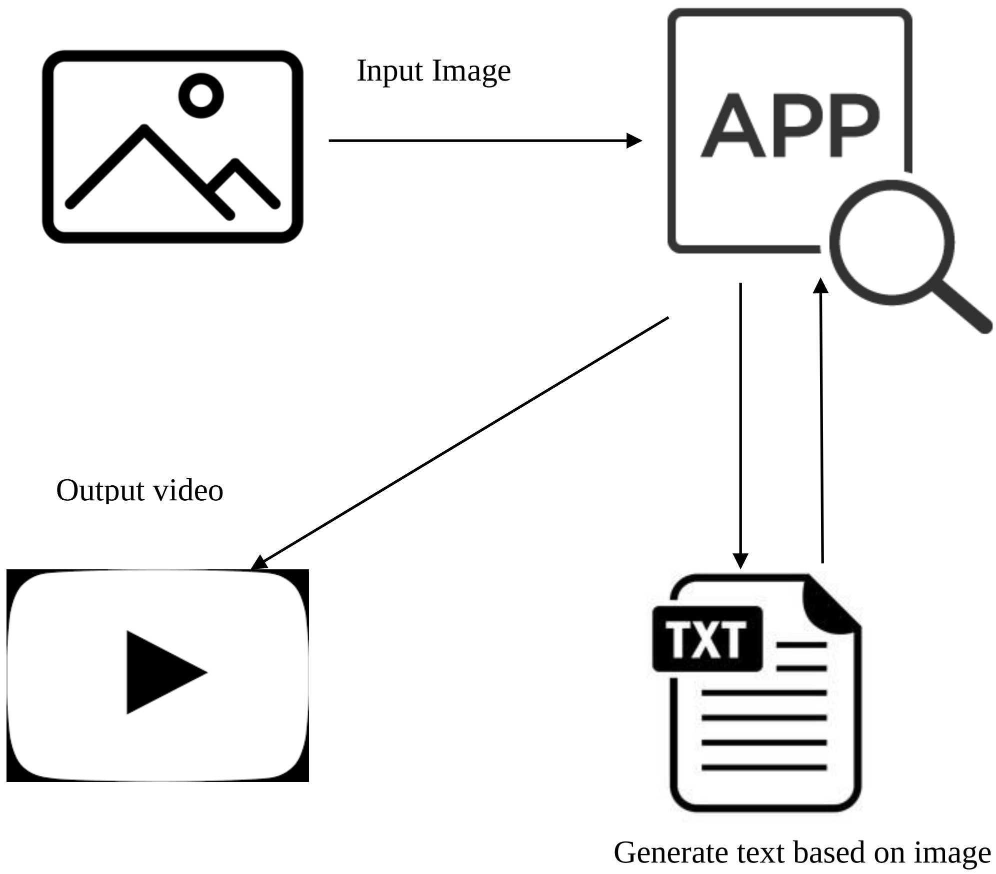


Figure 1:Basic architecture of the system

1.4 Goals

Provide a high-speed and low cost image base video search engine.

1.5 Achievement in brief

Developed system is a web-based application, which is identifying and describe an image for search related video from You Tube.

First, I generate text using input image and then create a URL to retrieve videos from YouTube. The proposed system has major 5 parts including client, image server, Cloudinary storage, Microsoft vision and YouTube API

Chapter 2

Review of others' work

2.1 Introduction

This paper lists all the available techniques for searching images on the web. And will try to find which technique is more feasible for searching the images based on query image and finally based on the retrieved results performance evaluation is done.

The bulk of image retrieval research falls under the rubric of "content-based" image retrieval; this term refers to systems that perform image analysis in order to extract low-level visual properties, such as color and texture or object segmentation. Some systems also incorporate information extracted from associated text. A good summary of content-based image retrieval can be found in [1].

There has been a great deal of research on these systems, but only a small subset of the past work has included usability studies. Rodden et al. [2] performed a series of experiments whose goal was to determine if and how organization by visual similarity is useful, using as features global image properties (colors and textures) and the spatial layout of image regions. Their results suggested that images organized by category labels were more understandable than those grouped by visual features.

Ethnographic studies of image search needs have indicated that there is a great need for more conceptually rich image search. In a study of art directors, art buyers, and stock photo researchers [3], Garber & Grunes found that the search for appropriate images is an iterative process: after specifying and weighting criteria, searchers view retrieved images, then add criteria, add restrictions, change criteria, or redefine the search. The concept often starts out loosely defined and becomes more refined as the process continues.

Markkula and Sormunen [4] reported on a field study of journalists and newspaper editors choosing photos from a digital archive in order to illustrate newspaper articles. Journalists stressed the need for browsing, and considered searching for photos of specific objects to be a "trivial task". Selection of search keys for general topics was considered difficult; journalists emphasized the need for photos dealing with places, types of objects, and themes. The journalists had access to an “advanced search” interface that allowed them to search on many different features at once, but its format, which consisted of about 40 entry forms and dropdown boxes, was seen as too complex, and was rarely used. Thus, although they had the desire to do searches on multiple categories, the interface discouraged them from doing so.

A query study also supports the notion that users want to search for images according to combinations of topical categories. Armitage and Enser [5] analyzed a set of 1,749 queries submitted to 7 image and film libraries. They classified the queries into a 3-by-4 facet matrix; for example, Rio Carnivals fell under Geographic Location and Kind of Event. They did not summarize how many queries contain multiple facets, but showed a set of 45 selected queries, to which they assigned an average of 1.9 facets per query.

The system proposed by Garber & Grunes [3] is the interface most similar to our approach. The interface operated in two modes: (i) showing metadata associated with a target image, and presenting images in an order reflecting the number of categories they had in common with the target image; and (ii) allowing the user to select a set of category labels, and showing sample images for similar categories (e.g., showing images labeled New England, Africa, and Egypt when the category label Florida is selected). Hierarchy information was not shown, and no information was provided about how many images are available in each category. Focus groups observing the demonstration were very enthusiastic about it, but no follow up work appears to have been done.

2.2 CCV (Color Coherence vector)

Color coherence vector is an image processing technique when applied to an image it exacts contents from that image. And stores content information in a coherence vector which can then be used for matching two images. CCV defines color coherence as degree to which the pixel of an image are members of large similarity colored region. If the query image is compared with the images stored in the database and the difference between them is found to be greater than 1000 then the difference is considered to be significant and the change is detected. [6]

However, many a times this technique alone fails to find similarity between two images. Hence is not so effective.

2.3 Sobel Edge Detection and CCV

The technique presented in [6] [7] presented an idea that when a combination of sobel edge detection and CCV is used the percentage of similarity for matching query image with images in the web increases. Firstly, it performs Sobel edge detection, contents from the uploaded image and image on the web are exacted Then CCV matrix technique is used for the comparison which checks the content similarity between the images being compared for similarity.

2.4 New Wavelet Feature

Wavelet based features discussed in [8] were first introduced in Jacobs et al. it selects sixty-four largest Haar wavelet coefficients in each of the 3 color band and stores them in feature vector as +1 or -1 along with their position in the transformation matrix. Low frequency coefficients tend to be more dominant than those of the high frequency coefficients and this makes this algorithm ineffective for images with sharp color changes. In addition to that, Haar wavelet basis is not suitable for natural images. [9].

2.5 Discrete wavelet Transform

Wavelet transform has emerged as an effective tool for analyzing texture features as it decomposes the image into various sub-bands which is multi-scale oriented. For computational convenience, the given input color image with size $N \div N$ is converted to a sequence $X(n)$ where $n = N*N$. The Discrete Haar wavelet transform described in [10] is adopted to decompose the input image by computing the approximations and details, which result in the low-pass and high-pass filters respectively. The decomposition process in [11] is performed till obtaining the optimum level i.e. fine level of the image.

2.6 RISE DCT Transform

RISE DCT transform is a technique in [12] was designed for the robust image search engine. The main idea of this technique is to develop an index for images database using the subset of JPEG coefficients in a compressed image. It creates an average of color components of each 8 x 8 pixel blocks in a JPEG image, which is the same as the DC-coefficient of each block after applying discrete cosine transform (DCT). The DC coefficient for each color component is used to develop an index. The main advantage of using average is that it can be applied to any image format. It is also more practical since it does not use the AC components of the DCT-transformed image [12].

2.7 Image Feature Inverted Indexing(FII)

Despite a lot of research on extraction of image feature, the design of feature index for high efficiency is yet to be worked out. The present content-based image retrieval engine takes the image feature into database so that more image features lead to more frequent modification of table field. Therefore, a new type of data indexing suitable for retrieval engine is needed. Combining image shape and texture feature with vector machining-supporting type, a new technique was designed named feature inverted indexing

This FII technique in [13] was used for image search engine Eva, it extracts image shape feature by using Zernike moments. Eva extracts image texture feature by using Gray level co-occurrence matrix algorithm depicting characteristic of texture. According this algorithm, Eva retrieves seven texture characteristics: uniform, contrast, correlation, variance, inverted-deviation matrix, standard-average, standard deviation. Based on this features the query image is then compared with the images in the web

2.8 FCTH and CEDD

FCTH and CEDD techniques explained in [9] deals with the extraction of a new low level feature that combines, in one histogram, color and texture information. These features are named FCTH - Fuzzy Color and Texture Histogram and CEDD - Color and Edge Directivity Descriptor. FCTH results from the combination of 3 fuzzy systems and size is up to 72 bytes per image. CEDD size is up to 54 bytes per image and so these descriptors are suitable for large image databases. CEDD need less computational power for extraction than most of the MPEG-7 descriptors. And thus is efficient in retrieving distorted and transformed images. Compact Composite Descriptors (CCD) are able to capture both, color and texture characteristics, and so very useful in a very compact representation [14] [15].

2.9 Content Oriented Image Retrieval (COIR)

COIR technique presented in [16] partitions the original image into several regions based on visual characteristics like color, edge, position, and texture using image processing techniques like edge detection [11] [17], color analysis and region division. Each such region is considered as objects. The number of regions depends on the image contents.

The indexing engine calculates the visual attribute values for each region. The attributes assigned to each region are color, shape, texture, size, object location and object composition. The extracted attribute values are stored as metadata [18] [19] and used during the query processing phase to determine similarity. Image index are stored as matrix and by comparing the matrix of the original image with that of in the web images are retrieved.

Chapter 3

Methodology

3.1Technology adopted

3.1.1 Web Development Technology

HTML

CSS

Java Script

AJAX

JSON

QUERY

3.1.2 You Tube API for search Videos

The YouTube Application Programming Interface, or the YouTube API, allows developers to access video statistics and YouTube channels' data via two types of calls, REST and XML-RPC.

3.1.3 Microsoft Computer Vision API

The cloud-based Computer Vision API provides developers with access to advanced algorithms for processing images and returning information. By uploading an image or specifying an image URL, Microsoft Computer Vision algorithms can analyze visual content in different ways based on inputs and user choices. With the Computer Vision API users can analyze images

3.1.4 Cloudinary

Cloudinary is a SaaS technology company headquartered in Sunnyvale, California, with an office in Israel. The company provides a cloud-based image and video management solution

3.1.5 Node.js

Node.js is a platform built on Chrome's JavaScript runtime for easily building fast and scalable network applications. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient, perfect for data-intensive real-time applications that run across distributed devices.

3.2 Approach

When using a standard web-based image search engine, one is likely to find garbage output even in the first couple of pages of results. Nevertheless, for simple objects, it can be conjectured that most of the top results will contain the object of interest. Our system exploits this consistency, and attempts to find the object that appears in most of these images.

in this paper we first use a text-based image meta-search engine to retrieve videos from the Web. By identifying the given image, we generate best match text to search videos. based on the text information on the image host pages to provide an initial video link. Because of the high-speed and low cost nature of the text-based approach, we can easily retrieve a broad coverage of Videos.

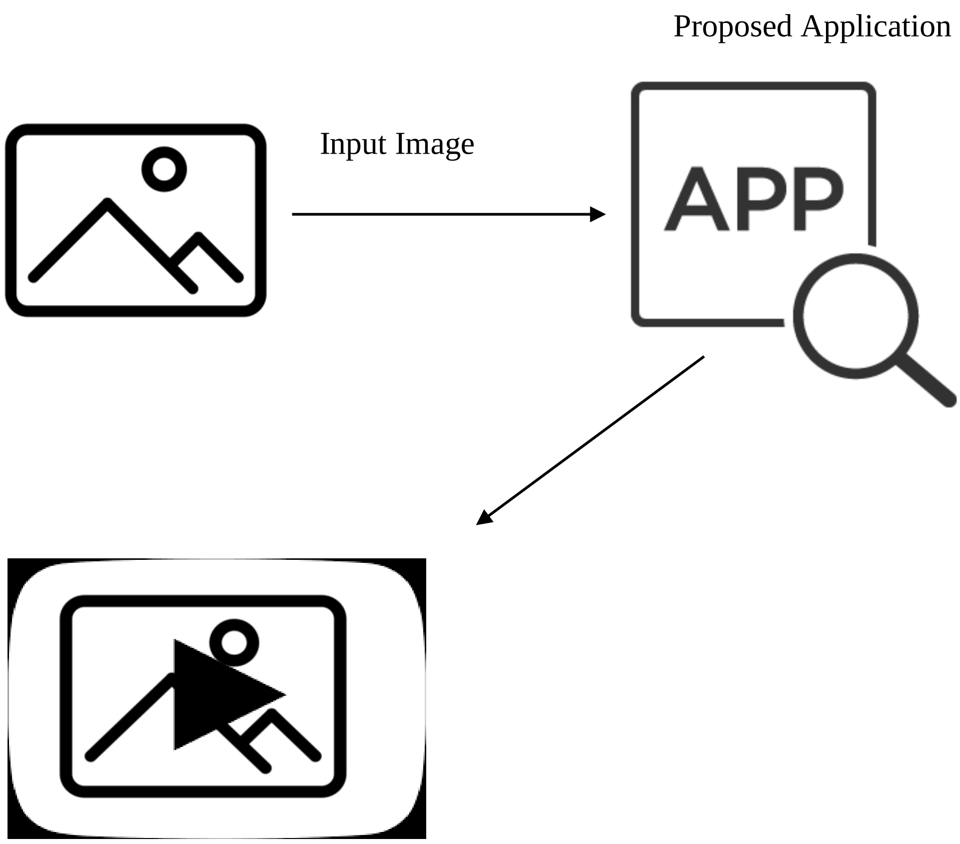


Figure 2:Process of the system

3.3 Development stages

Preliminary Study

Studied on existing image identification project like Google image search, Microsoft vision with you tube API.

Submission of Proposal and approval

Develop the project proposal on the problem identified by proposing the IT driven solution.

Inertial prototype development

First develop a prototype to identify given image. And studied You tube API to match video with image.

System design

After having good result with prototype I start the video searching part.

Constructing the final system

Since the area is too large after having discussion with my supervisor we narrow down the video searching part to sports video search.

Evaluation and testing

The system is tested on more than 100 ICC players profile.

Chapter 4

Design

Using both text and image content features, a hybrid video retrieval system for Word Wide Web is developed. We first use a text-based image meta-search engine to retrieve videos from the Web. By identifying the given image, we generate best match text to search videos. Based on the text information on the image host pages to provide an initial video link. Because of the high-speed and low cost nature of the text-based approach, we can easily retrieve a broad coverage of Videos

Image vs Text search

Text search is high speed and low cost in internet. So major disadvantage in image search engine is when input image large system is getting slow. Due to this draw back lots of available image searching engine limited their input image size.

To avoid above disadvantage first, I generate text using input image and then create a URL to retrieve videos from YouTube.

The proposed system has major 5 parts including client, image server, Cloudinary storage, Microsoft vision and YouTube API.

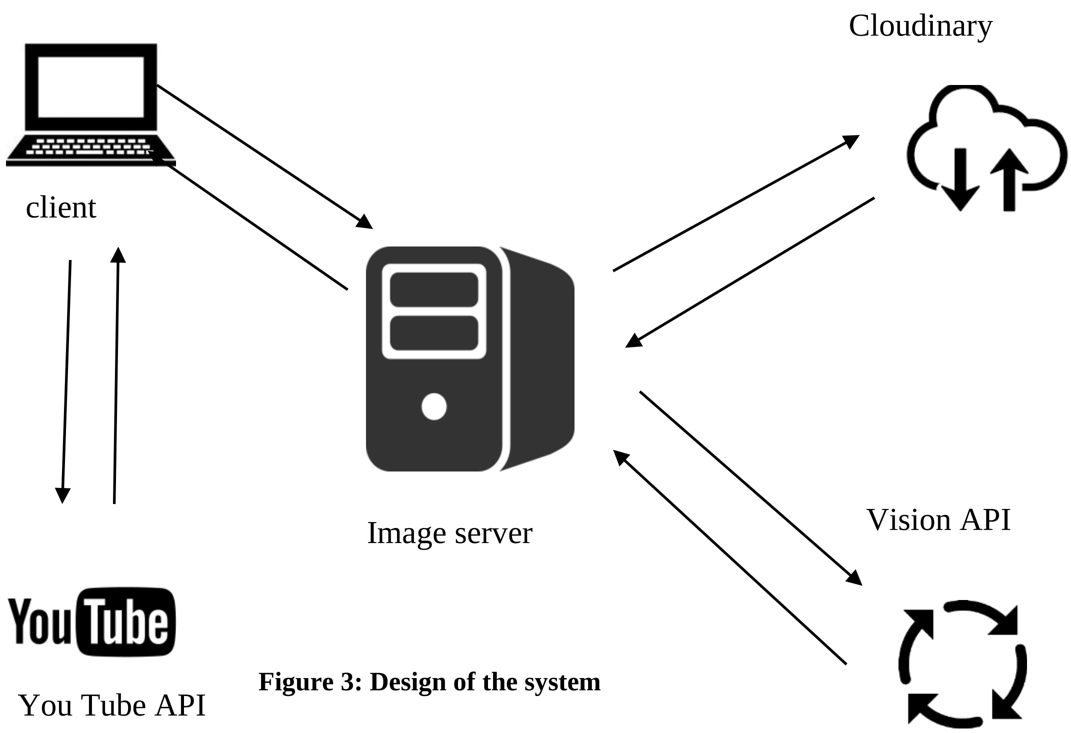


Figure 3: Design of the system

4.1 Purpose of the Client in proposed system

Client is use for upload image to the server. After analyzed the upload image server will return URL to access YouTube API. Client has to send the request to the You tube and You tube API will send the result to the client. This is the final output of this process.

The developed image server identified and analyzed image therefor this system is much faster and low cost when comparing other system.

4.2 Image server

This is the core of this system and server communicate with all other to process the image. Server has upload, identify, response major role in proposed system.

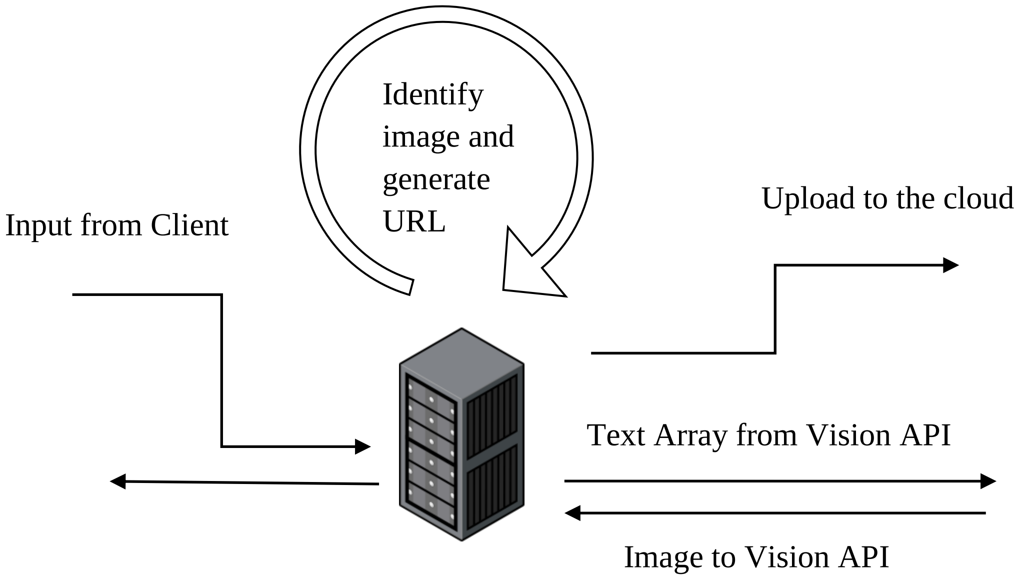


Figure 4: Design of the server

4.2.1 Upload

As a storage device in this system I used online Cloudinary space to stored input image. When client upload the image to the server, server will communicate with Cloudinary and request the URL from Cloudinary to provide the Microsoft Vision API.

4.2.2 Identify

Computer Vision API returns tags based on more than 2000 recognizable objects, living beings, scenery, and actions. Base on vision response server will filter and manipulate the result for generate URL to search related videos.

4.2.3 Response

The next role main role of the server is response for other section of the system such as client, cloud space, vision API.

4.3 Cloudinary

Cloudinary is a Software-as-a-Service (SaaS) solution for managing all your web or mobile application's media assets in the cloud. Cloudinary offers an end-to-end solution for all your image and video needs, including upload, storage, administration, manipulation and delivery. Media upload, processing, and delivery are done on Cloudinary's servers and automatically scale for handling high load and bursts of traffic.

Resources are uploaded, managed, and transformed using Cloudinary's Management Console and easy to use APIs, and delivered from high-performance servers through Content Delivery Networks (CDNs). While you can use Cloudinary's APIs directly within your own custom code, it is simpler (and recommended) to use one of Cloudinary's client libraries (SDKs), which wrap the Cloudinary APIs and simplify integration with web sites and mobile applications.

Cloudinary is a cloud-based service that provides an end-to-end image management solution, including upload, storage, administration, manipulation, optimization and delivery.

With Cloudinary we can easily upload images to the cloud and automatically perform smart image manipulations without installing any complex software. Cloudinary provides a secure and comprehensive API for easily uploading images from server-side code, directly from the browser or from a mobile application. You can either use Cloudinary's API directly or through one of Cloudinary's client libraries (SDKs), which wrap the upload API and simplify integration with web sites and mobile applications.

4.4 Microsoft vision API

Computer Vision API returns tags based on more than 2000 recognizable objects, living beings, scenery, and actions. When tags are ambiguous or not common knowledge, the API response provides 'hints' to clarify the meaning of the tag in context of a known setting. Tags are not organized as a taxonomy and no inheritance hierarchies exist. A collection of content tags forms the foundation for an image 'description' displayed as human readable language formatted in complete sentences.

After uploading an image or specifying an image URL, Computer Vision API's algorithms output tags based on the objects, living beings, and actions identified in the image. Tagging is not limited to the main subject, such as a person in the foreground, but also includes the setting (indoor or outdoor), furniture, tools, plants, animals, accessories, gadgets etc.

4.5 YouTube

At YouTube, we understand the power of video to tell stories, move people, and leave a lasting impression. One part of storytelling that many people take for granted is sound, yet sound adds color to the world around us. Just imagine not being able to hear music, the joy of a baby laughing, or the roar of a crowd. But this is often a reality for the 360 million people around the world who are deaf and hard of hearing. Over the last decade, we have been working to change that.

YouTube allow to perform many of the operations available on the website. It provides the capability to search for videos, retrieve standard feeds, and see related content. A program can also authenticate as a user to upload videos, modify user playlists, and more. This integration can be used for a variety of uses such as developing a web application allowing users to upload video to YouTube, or a device or desktop application that brings the YouTube experience to a new platform. The Data API gives users programmatic access to the video and user information stored on YouTube. This can be used to personalize a web site or application with the user's existing information as well as perform actions like commenting on and rating videos.

4.5.1 YouTube API v3

YouTube API v3 is support to search and watch videos from YouTube. YouTube offers various powerful ways to play around with videos. We can create awesome applications using the API and also can embed videos from YouTube. It has client libraries in various programming languages to do that.

The YouTube Data API (v3) lets us incorporate YouTube functionality into our own application. We can use the API to fetch search results and to retrieve, insert, update, and delete resources like videos or playlists.

4.6 Functional overview

This system is a completely web based solution. It could be use in smart phone or smart Tab. Only required input is an image. System will provide you a video link with analyzed more details.

Chapter 5

Implementation

5.1 Introduction

This section describes implementation of the system and the layouts of the proposed video retrieval system. it includes some of the graphical interfaces of the system. Due to the time limitation I selected only one area as a prototype. I selected sports videos, and select cricket as a sports.

5.2 System functionalities

Following are functionalities of developed video retrieval system

5.2.1. Upload image

Use has to upload an image which has only present one player

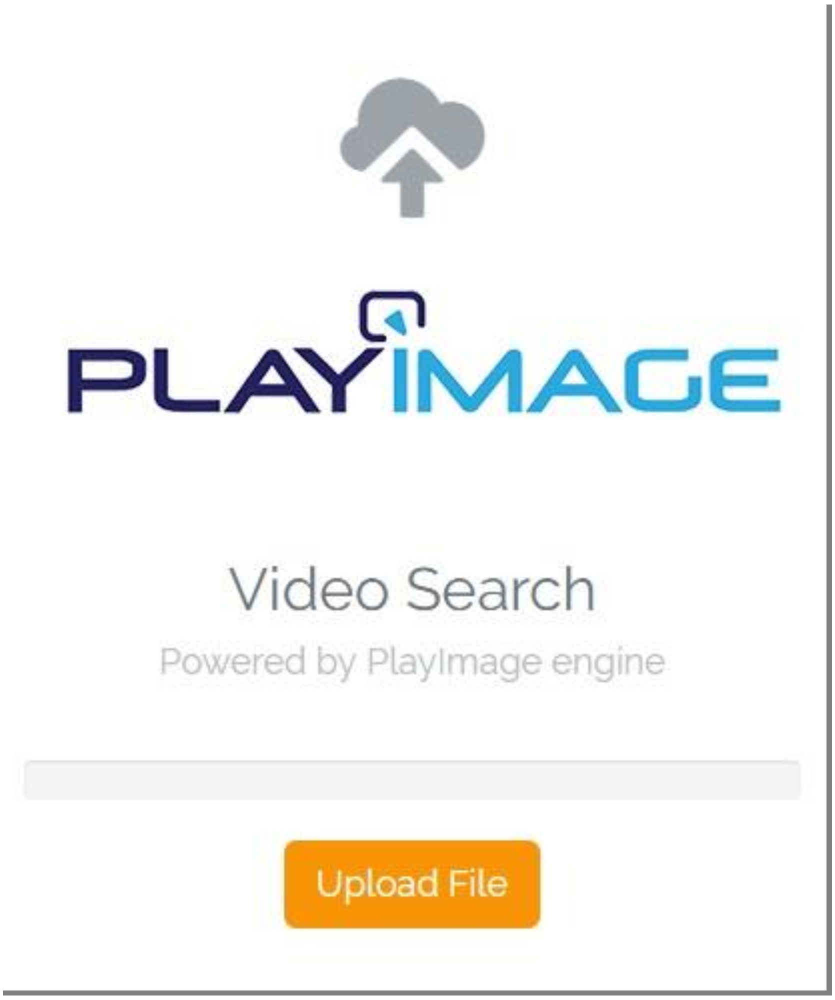


Figure 5: Upload screen

5.2.2. Play image server

Image server is main role of this application it works as intermediate communicate process with other resource and APIs. Following are the main rolls of this server

i. Upload image to the cloud and generate URL.
ii. Send the URL + Tags to Computer Vision API for identify image.
iii. Read the Computer Vision API output (Text) and description
iv. Send the description to client
v. Using generated text crate URL for YouTube API.
vi. Get the result from YouTube
vii. Add filter to with Videos Title and thumbnails
viii. Display result

5.2.3 Cloudinary

Cloudinary is a cloud-based service that provides an end-to-end image and video management solution including uploads, storage, manipulations, optimizations and delivery.

Input

```
cloudinary.v2.uploader.upload("sample.jpg",
    { public_id: "sample_id" },
    function(error, result) {console.log(result); });
```

output

http://res.cloudinary.com/demo/image/upload/v1371750447/sample.jpg

Upload response

An upload API call returns a response that includes the HTTP and HTTPS URLs for accessing the uploaded image, as well as additional information regarding the uploaded image: The Public ID and current version of the image (used in the Media Library, Admin API, and for building manipulation and delivery URLs), the image's dimensions, the image format and a signature for verifying the response. The following is an example of the JSON response returned:


```
{
 public_id: 'sample',
 version: '1312461204',
 width: 864,
 height: 564,
 format: 'jpg',
 created_at: '2015-10-10T09:55:32Z',
 resource_type: 'image',
 tags: [],
 bytes: 9597,
 type: 'upload',
 etag: 'd1ac0ee70a9a36b14887aca7f7211737',
 url:
'http://res.cloudinary.com/demo/image/upload/v1312461204/sample.j
pg',
 secure_url:
'https://res.cloudinary.com/demo/image/upload/v1312461204/sample.
jpg',
 signature: 'abcdefgc024acceb1c1baa8dca46717137fa5ae0c3',
 original_filename: 'sample'
}
```


5.2.4 Identify image -Computer Vision API.

The cloud-based Computer Vision API provides developers with access to advanced algorithms for processing images and returning information. By uploading an image or specifying an image URL, Microsoft Computer Vision algorithms can analyze visual content in different ways based on inputs and user choices. With the Computer Vision API users can analyze images to:

- Tag images based on content.
- Categorize images.
- Identify the type and quality of images.
- Detect human faces and return their coordinates.
- Recognize domain-specific content.
- Generate descriptions of the content.
- Use optical character recognition to identify text found in images.
- Distinguish color schemes.
- Flag adult content.
- Crop photos to be used as thumbnails.

After uploading an image or specifying an image URL, Computer Vision API's algorithms output tags based on the objects, living beings, and actions identified in the image. Tagging is not limited to the main subject, such as a person in the foreground, but also includes the setting (indoor or outdoor), furniture, tools, plants, animals, accessories, gadgets etc.

Input

Output tag array will be


```
Returned Json
{
'tags': [
          {
            "name": "grass",
              "confidence": 0.999999761581421
          },
          {
            "name": "outdoor",
              "confidence": 0.999970674514771
          },
```

```
            {
                "name": "sky",
                  "confidence": 999289751052856
            },
            {
                "name": "building",
                  "confidence": 0.996463239192963
            },
            {
              "name": "house",
                "confidence": 0.992798030376434
            },
            {
              "name": "lawn",
                "confidence": 0.822680294513702
            },
            {
              "name": "green",
                "confidence": 0.641222536563873

},
            {
              "name": "residential",
                "confidence": 0.314032256603241
            },
          ],
}
```

Chapter 6

Discussion

6.1 Discussion

New algorithm was proposed to find visually similar video clips with different temporal durations and spatial variations. Contributions are as follows. First, perform a text-based meta-search to obtain an initial video set with relatively high recall rate and low precision rate. Then the image content based processing is employed to produce a much more relevant output.

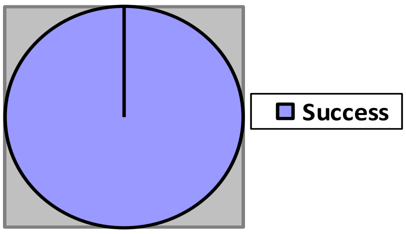
In this paper, various techniques available for video retrieval have been discussed. These traditional techniques face many difficulties when exact videos are to be retrieved, hence a concept of reverse image search has been proposed and designed. And thus the results generated from this system have been discussed. This system mainly aims at retrieving the exact as well similar videos from web based on a query image. According to the concept of this system till now it is able to retrieve the exact videos from a given and image. Will try to implement this system on large videos databases which can then be used for detecting unauthorized use of brands and copyright videos.

6.2 Product Evaluation

Used more than 500 images to test this application with deferent test scenario

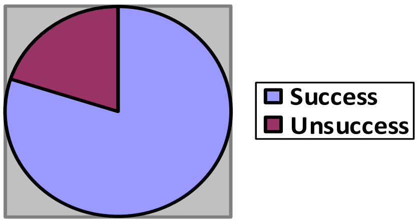
1. using ICC profile Pictures

Confidence level 100 %



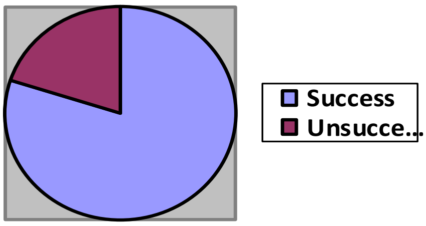
2. Using non ICC profile picture

Confidence level -80 %



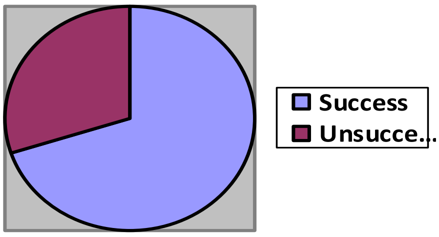
3. Select 10 deferent angle with selected player

Sample test scenario with Mahela Jayawardana



Confidence level 80 %

Sample test scenario Kumar Sangakkara



Confidence level 70 %

Conclusion

The main purpose of this paper is to provide an efficient and truly realizable approach for WWW based videos retrieval. We first perform a text-based meta-search to obtain an initial video set with relatively high recall rate and low precision rate. Then the image content based processing is employed to produce a much more relevant output. There are three ways to combine the text-based method and the visual content-based method: use the text-based method first; use the visual content-based method first; use the two methods at the same time. The key to the success of this system is using the first approach. By using the high-recall, low-precision, and low-cost text-based method first, proposed application can easily collect as many relevant videos as possible over the Internet at a very low cost. Then the high-precision and high-cost visual based method is used to improve the relevance precision on a significantly smaller video set.

In this paper, various techniques available for image retrieval have been discussed. These traditional techniques face many difficulties when exact images are to be retrieved, hence a concept of reverse video search has been proposed and designed. And thus the results generated from this system have been discussed. This system mainly aims at retrieving the exact as well similar videos from web based on a query image. According to the concept of this system till now it is able to retrieve the exact video from a given image.

References

1. R. C. Veltkamp and M. Tanase. Content-Based Image Retrieval Systems: A Survey. Technical Report UU-CS-200034, Dept. of Computing Science, Utrecht University, 2000.

2. K. Rodden, W. Basalaj, D. Sinclair, and K. R. Wood. Does organisation by similarity assist image browsing? In Proceedings of ACM SIGCHI 2001, pages 190–197, 2001.

3. S. R. Garber and M. B. Grunes. The art of search: A study of art directors. In Proc. of CHI-92, Monterey, CA, 1992

4. M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. Information Retrieval, 1:259–285, 2000

5. L. H. Armitage and P. G. B. Enser. Analysis of user need in image archives. Journal of Information Science, 23(4):287– 299, 1997.

6. Divya venkata, divakar yadav.” Image query based search engine using content based image retrieval”. 2012 14th International Conference on Modelling and Simulation.

7. Divya Ragadha, Deepika Kulshreshtha and Divakar Yadav ”Techniques for refreshing Images on Web Document ”, 2011 International Conference on Control, robotics and Cybernetics

8. A. Lakshmi Subrata Rakshit “New Wavelet features for image indexing and retrieval”. 2010 IEEE 2nd International Advance Computing Conference.

9. P.Praveen Kumar, Aparna, Dr K Venkata Rao PhD.” Compact Descriptors for Accurate Image Indexing And Retrieval: Fcth And Cedd”. International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October – 2012 ISSN: 2278-0181

10. Ting Yao, Chong-Wah Ngo, Tao Mei." Circular Reranking for Visual Search". IEEE Transaction on image processing, vol. 22, no. 4, April 2013.

11. Dr. K. Seetharaman, M. Kamarasan "A Smart Color Image Retrieval Method Based on Multiresolution Features", Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on. A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

12. Goswami, D, Bhatia, S.K. "RISE: A Robust Image Search Engine" , 2006 IEEE International Conference on Image Processing.

13. Tan Hao, Chen Yu, Qiu Hang ,"Implementation Of FII-based image retrieval engine", Apperceiving Computing and Intelligence Analysis, 2009. ICACIA 2009. International Conference on

14. Konstantinos Zagoris, Savvas A. Chatzichristofis, Nikos Papamarkos and Yiannis S. Boutalis, "img (Anaktisi): A Web Content Based Image Retrieval System" , 2009 Second International Workshop on Similarity Search and Applications.

15. Harsh Kumar Sarohi, Farhat Ullah Khan ,"Image Retrieval using Perceptual Hashing", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 9, Issue 1 (Jan. - Feb. 2013), PP 38-40.

16. Sougata Mukherjea, Kyoji Hirata and Yoshinori Hara, "AMORE: A World Wide Web image retrieval engine", ACM, Journal of World Wide Web , Volume 2, Issue 3, 1999

17. Piyush Kansal and Vinay Krishnamurthy,"Reverse Image Search".

18. Xinmei Tian, Yijuan Lu, Linjun Yang." Query Difficulty Prediction for Web Image Search". IEEE Transaction on multimedia, vol. 14, no. 4, August 2012.