

# Tagging and Clustering of Objects in Videos

Vikrant Gangan

C.N.S.Sisma

Monisha Menon

Aditya Gupta

Department of Information  
Technology

Department of Information  
Technology

Department of Information  
Technology

Department of Information  
Technology

Fr. C.R.I.T, VashiNavi  
Mumbai.

Fr. C.R.I.T, VashiNavi  
Mumbai.

Fr. C.R.I.T, VashiNavi  
Mumbai.

Fr. C.R.I.T, VashiNavi  
Mumbai.

## ABSTRACT

"Tagging" in videos is an extension of the image tagging application. Just as people can be tagged in images, a system can be developed to carry out tagging in video clips. Using this, people can automatically be identified ("tagged") wherever they are present in a video clip. The paper discusses the design of the system which identifies the real world objects in a video clip. The object to be searched can be provided by the user in the form of an image. The object is specified by a user outlining it in an image, and the object is then delineated in the retrieved shots. The aim is to detect and track presence of required objects in a video and for thesearching to be view-point invariant. Finally, the searched object can be tracked throughout the video by clustering all of its occurrences into a single video. This can be used for analyzing any possible changes the object might have undergone during the length of the video.

## General Terms

Object detection, object matching, tagging, video analysis, image processing.

## Keywords

tagging, clustering, searching, object, video, frames.

## 1. INTRODUCTION

With modern-day social networking, the world has become a smaller place to live in. Sites like Facebook, Twitter, Orkut, and Hi5 allow us to upload a lot of pictures which we can 'tag' ourselves and our friends in.

According to the modern dictionary, tagging is signing your name or other representation of yourself on anywhere public. "Tagging" here, however, refers to labeling. It is a feature that allows you to select a portion of a picture and have that image mapped to a user. For example, you can select a friend's face in a picture and tag it with their name. Then a list can be generated automatically of all the people present in the picture, each user can have a list of all the pictures they appear in. Users can also be notified of their presence in pictures with the help of such tags. The need for video tagging has been felt in order to recognize people in videos as well. The use of tagging in videos can have various uses ranging from social networking to security systems.

Depending on who you ask, the concept of tagging is either incredibly convenient or a potential privacy nightmare. Hence, video tagging can help a person know whether he/she is present in a particular video. This will enable the person to detect videos, which have been uploaded without his/her consent. A similar thing can be done with objects in videos wherein the presence of a particular object can be affirmed or dismissed. This could be added to the arsenal of security

services, the ones provided by the government and also otherwise.

This method can also be used to confirm the presence of a person in a video footage which has been taken by a CCTV camera or other such devices and can have its applications in banks, colleges etc.

## 2. EXISTING SYSTEMS

Currently, video-tagging is prevalent on the Internet, but in different interpretations of the term. Youtube and other video-sharing websites allow users to associate words or phrases with entire videos. However, the simple fact that video, unlike still images, tends to be a linear experience that unfolds over duration of time limits the usefulness of a tag applied to the entire clip, especially with longer clips. This does ensure, though, that the tagged videos show up in relevant searches. Different types of video tagging are discussed below:

### 2.1 Facebook

The well-known social networking site Facebook allows users to tag videos by associating users' names with the video(s) they appear in. As of now, this functionality is open to users of Facebook only; non-members cannot be tagged in Facebook videos..

### 2.2 BlipSnips®

BlipSnips®, a Colorado-based startup, does just this while providing free consumer products designed to make video sharing simple and fun across a user's social networks such as Facebook and Twitter. The products offers moment by moment video captioning allowing the user to inject commentary at specific moments of videos. Viewers can then skip straight to the moment they wish to see. Each listed video mentions, along with its name, the number of "moments" in it, marked by users themselves. BlipSnips® also allows users to tag friends by way of their social network (e.g. Facebook) and notify them of their "appearances" at specific points in a video. BlipSnips® offers these capabilities through a variety of means: a destination website at [www.blipsnips.com](http://www.blipsnips.com); an iPhone app designed to shoot, tag and upload video; a Facebook app to perform these functions against Facebook-hosted and YouTube videos.

### 2.3 Deep Tagging

Deep tagging is the advanced form of tagging which allows users to make a direct link to a small part of a larger piece of media for example, an image or video. This type of deep tagging is also used by other websites like Veotag to enhance efficiency of video searches through search engines. The Veotag web video publishing platform proclaims itself to be a better way to watch online video and audio. Veotag allows end users to add visual search to audio and video. It also makes videos searchable from outside search engines. This has become one of the important applications of video-

tagging. With a customizable player and the ability to deliver targeted and time-sensitive content like advertising, Veotag is revolutionizing the way people think about how they can use audio and video as a way to communicate. With Veotag, content producers can make active tables of contents, and descriptions that enrich their audience's experiences. The audience sees the text menus whenever they play a Veotagged file on the web and that text is linked to the file on a time basis, so whenever the audience clicks on a Veotag, they jump right to that part of the file. Once a media file is Veotagged, the content in the file is instantaneously exposed to all major industry search engines. This allows the world to search within the audio or video file, not just for it. Thus, users will no longer have to go through entire media files to obtain information they require; they can, instead, directly look up and view only those parts of the files which have been tagged with the words they are looking for.

These are new developments in social media. However, these established systems are lacking in certain respects.

## 2.4 Issues in Existing Systems

- 1) Existing systems do not yet have a provision to tag the people appearing in videos along with a frame or box identifying the person within the video.
- 2) Also, there exists no system that would allow one to search for the presence of a person within a video using an image of the person.

## 3. PROPOSED SYSTEM

Current commercial systems that implement video tagging do not provide tagging of objects that interest the user within a video. They only allow one to mark out "moments" in a video along the timeline, which users can view separately as per their choice. This is useful, but has limited functionality. Tagging, in this case is not an extension of the concept popularly used in images. Therefore, the user will not be able to see object(s) of interest highlighted in the video using a box or frame. The system under discussion aims at making up for the lack of these functions that are missing in current systems.

The system is one that will basically highlight the occurrences of a specific object in a stored video. Technically, this object can be anything, ranging from simple two-dimensional logos of organizations, products etc. to human faces and figures. Our system aims to identify inanimate objects, starting with simple two-dimensional objects such as corporate logos and insignia. The specialty of this system is that it allows the user to give an image as input which will form the basis of the search. The user can then view search results and play those lengths of video clips which the selected object appears in. The user can also see a clustered video which contains the frames where the object is present in case the user wants to analyze object's behavior in the video.

The user cannot provide more than one image as input at one time. Also, the image provided should have the object of interest as its focus. The image must not have other objects that are as prominent as the one to be searched to prevent ambiguity.

## 4. DESIGN

Block diagram of the video tagging system is shown in Figure 1.

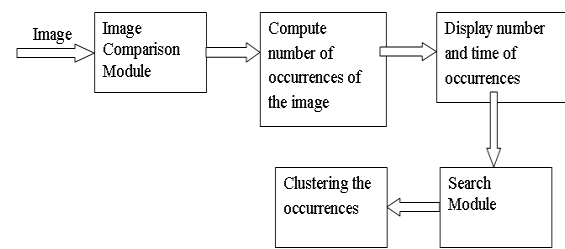


Figure 1: Block diagram of video tagging system.

Blocks of the diagram are explained as follows:

### 4.1.1 Image:

The input is taken from the user in the form of an image. This image will be the basis of the search that will be carried out in the video.

### 4.1.2 Image comparison module:

In this module, the image that has been accepted from the user is compared with each frame in the video. This is done in order to check for the presence of the object of interest in the video. If one or more matches are found, the object is present in the video.

### 4.1.3 Compute number of occurrences of the image:

In this module, the frames showing success are grouped together and considered as a single entity, if they are sequential. Each such sequence of frames will correspond to one occurrence of the object. The total number of such occurrences is calculated. Also, the duration of each occurrence is noted.

### 4.1.4 Display number and time of occurrences:

The user is notified of the success or failure of the search process. If the search has been successful, a message is displayed to the user telling him the number of matches found which refers to the number of occurrences of the object in the video. Along with the number of matches found, the user will be shown the duration of each occurrence, i.e., starting time and ending time of each occurrence.

### 4.1.5 Search module:

This module allows the user to go directly to any one of the occurrences and play just that sequences. This is made possible by providing the user links to each occurrence. This is an optional module; the user need not view any of the links.

### 4.1.6 Clustering the occurrences:

The clustering module finally clusters all occurrences of the object into one video. This is done to enable the user to easily view and track the object's progress in the video. Irrespective of whether or not any of the occurrences were viewed individually by the user, all occurrences of the objects will necessarily be clustered into a single video.

## 5. CONCLUSION

The video tagging system allows the user to search for the presence of any object in a specified video. An image of this particular object is used to do so. Once the object is searched, it can be tracked through the entire video. This can be done

easily onceall occurrences of the object are clustered into a single video.

## 6. REFERENCES

- [1] Jia Wang and Wei Tsang Ooi, "Detecting Static Objects in a Busy Scene," Department of Computer Science, Cornell University, Ithaca, Technical Report TR99-1730, Feb. 8, 1999.
- [2] "Automatic image annotation," [http://en.wikipedia.org/wiki/Automatic\\_image\\_annotation](http://en.wikipedia.org/wiki/Automatic_image_annotation), June 20, 2011.
- [3] "Content-based image retrieval," [http://en.wikipedia.org/wiki/Content-based\\_image\\_retrieval.html](http://en.wikipedia.org/wiki/Content-based_image_retrieval.html), September 16, 2011.
- [4] "Facebook Wins Patents For Tagging in Photos, Other Digital Media," <http://www.insidefacebook.com/2011/05/17/facebook-patent-photo-tagging.html>, May 17, 2011.