CHAPTER 1

INTRODUCTION

This chapter consists of several subtopics: (1) rationale which identifies the background of study, describes the problem situation, justifies the existence of problem situation using reference, and makes a clinching statement that will relate the background to the research problem; (2) theoretical framework which describes the theories or concept which are useful in the research; (3) statement of the problem that describes the research problem to be solved; (4) objective which explain the purpose of this study; (5) hypotheses that provide the method used to solve the problem based on the theory or empirical evidence, and they should be measurable; (6) conceptual framework which identifies and discuss about variable/schematic diagram related to the problem; (6) scope and delimitation that indicate the area covered in this research; (7) importance of the study which describes the contribution of the study as new knowledge.

1.1 Rationale

In recent years, the development of production, distribution, and processing of multimedia documents has grown very fast. The extreme growth of the internet has made the process of transmitting data, distribution and access to digital media content easier. So the media industry like music, movies, and other artwork will more often intersect and dealing with illegal activities such as copying, modifying and distributing their products without permission, in other words piracy works. Based on a study conducted by David Price [32] in November of 2011, there were 297 million internet users accessing illegal content, and in January 2013, this expanded to 327 million Internet users have made access to illegal content at least once a month. In 2010, a total of 3,690 petabytes of illegal content have been discovered, and in 2012 has grown rapidly to 9,570 petabytes. Here we can see how many violations are done in multimedia content that contains copyright.

In search of copy of a document/media, all begins with the emergence of many search engines such as w3catalog (1993), Altavista (1994), Google (1997) etc. A search engine is a program that scans for and recognizes things in a database that compare to keywords or characters determined by the user. The search results are presented in a list and ordinarily known as hits. Initially search engines start from text based search, where currently is being used to search similarity in text to look for plagiarism. Then the tools extended into image based search. Examples of the applications that are widely used are Googles Goggle, tineye.com etc. These image search tools only tell us about the information of the object that we uploaded. For the audio based search there are widely used applications such as SoundHound, Shazam, Audiosear.ch etc. In video-based search, many methods are used, among others, by using video metadata, text recognition, audio fingerprint, watermark search, and use of video fingerprint. For example in YouTube, they already used the audio fingerprint in their server to recognize the uploaded videos contain copyrighted audio or not.

In the last twenty years there were efforts to develop copyright marking such as a watermark [4, 29, 37]. But the watermark could be removed by using several parameters. This watermark disadvantages [27, 31] has provided an opportunity for the development of new methods, in this case, video fingerprinting. Video fingerprinting has characteristics resistant against modifications on changing the resolution and file format change. They can be used to identify the full video, part of the video and, even more importantly, in a very short piece of video.

1.2 Theoretical Framework

There are several methods for detecting video copies. One method is called watermarking, that is a method to insert unique and invisible code of the user into the video content. Another method is to use content itself. This is known as Content-Based Copy Detection (CBCD). CBCD methods are called fingerprinting in analogy to human fingerprints that can uniquely identify a human being. In this context, a video fingerprint can uniquely identify a piece of video content. Thus, the process of extracting a fingerprint from the video content is referred as video fingerprinting. This kind of fingerprinting is additionally called passive fingerprinting to differentiate it from the watermarking approach.

Compared to watermarking, fingerprinting permit the identification of multimedia content without changing the media content. The implanted data in watermarking is in fingerprinting extracted from the media content itself. Thus, fingerprinting allow the identification of legacy content without requiring any preprocessing before use. Fingerprinting is less vulnerable against attacks and distortions compared to watermarking, since fingerprints are derived from the most applicable parts of the media content. As detriment, the computational complexity of fingerprinting is higher than watermarking. Moreover, vice versa to watermarking, the message (unique mark) is not independent from the media content. The fingerprinting system can just recognize perceptually comparable duplicates of information on account of its stochastic nature. Also as a consequence of content dependency, fingerprints can't give confirmation of rights data to publishers. Figures 1.1 showed analogy of watermark and fingerprint.

Figure 1.1: Example of Watermark (left) and Video Fingerprint (right)

Many attack already been submitted on Video Fingerprint case, such as blurring, the zooming case, put a logos inside, changes the color video into only gray-level, changes the background, etc. Those attacks are mostly done on the content and temporarily like video trimming.

1.3 Statement of the Problem

During this time, the modification on the video was still a modification on the content, in the form of changes in brightness, blurring the video, adding a text, zooming, picture in picture, change the color video into grayscale or in temporal, such as video trimming. Figure 1.2 illustrates examples of such modifications.

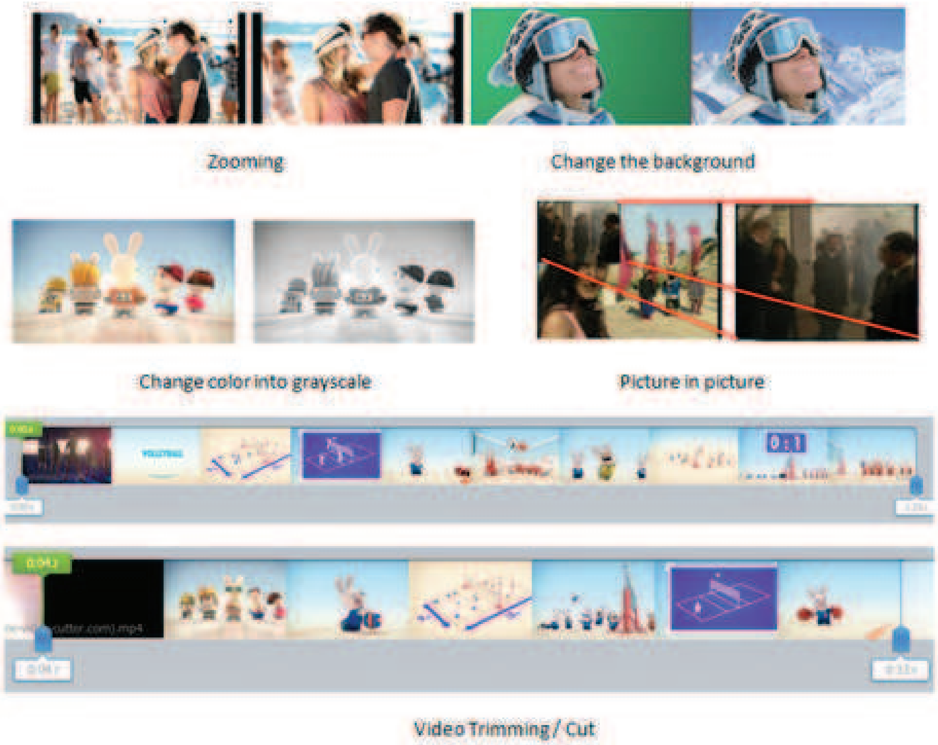

Figure 1.2: Example of modified video in content and temporal

All the above modifications had been done in the content and temporally domains, while spatial modifications have not been handled yet. This modification has been mentioned by professor Pierre Moulin[5]. The statement has motivated us to study the spatial attack problem in video fingerprinting and proposed a method to solve the problem.

1.4 Objective

We need a method to detect the spatially modified video. Firstly, we need how many number of frames to be compared between the modified video and the number of videos in the database. Second we need a feature extraction that can represent a partially attacked part from a video. Third, we must determine the most appropriate comparative way to matching the video with notice the accuracy and timing of the process.

1.5 Hypotheses

As has been mentioned above, this work tried to propose a method of building video fingerprinting against spatial attack. In this case, if a video is cut in half through the whole frames and we must still recognize the video, it is necessary to develop a feature taken from independent positions within the frame, so that if there is an attempt to cut the video spatially, only the feature on the piece is drawn, but the feature on the remaining piece can still be read and identified. The smallest unit of the feature is the pixel itself. But it would take tremendous resources to store and to recognize in pixel level, while on the other side, feature taken from the whole frame is vulnerable to the spatial modification. The optimal one is the feature taken from the area between them.Therefore, each frame is divided into blocks with the same size and the features are extracted from each block.

In this work, we used Discrete Cosine Transform (DCT) coefficients [2, 8, 23] for the feature. The DCT is related to the Discrete Fourier Transform but the DCT does not involve complex computations, so that the cost is simpler. Since each coefficient represents one of many wave behaviors of the whole content of block, we dont have to use all coefficients, but as few as necessary.

To reduce the amount of data which should be stored for a single video clip, the features were not extracted from all frames, but only from certain frames (called key frames) which were selected based on similarity with consecutive frames. The detection process was based on features comparison between key frames with those stored in database.

1.6 Conceptual Framework

Figure 1.3 illustrates the conceptual framework of the proposed system. First the video is broken down into frames, and divide each frame into blocks. Then we took the feature from each block in frame, finding the key frame from frames and compare it with the video in database. If there is a suitable video from the database then the system will tell us that the Modified video comes from the video in the database otherwise the modified video does not come from the video in the database.

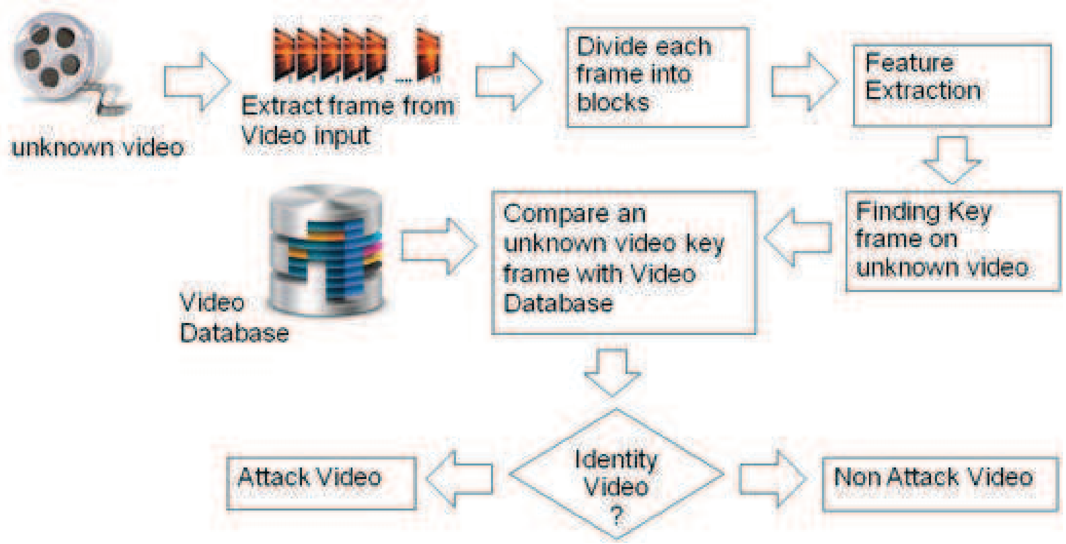
unknown video
Extract frame from Video input
Divide each frame into blocks
Feature Extraction
Finding Key frame on unknown video
Compare an unknown video key frame with Video Database
Video Database
Identity Video ?
Attack Video
Non Attack Video

Figure 1.3: The proposed system to detect spatial video modification

1.7 Scope and Delimitation

In this research, the scope and delimitation of this study are:

1. The video clips used in this study have resolutions up to 640 pixels wide and 360 pixels high.

2. The spatially modified video is a video that can be cut in right/left or up/down direction, and still contains a corner of each frame.

1.8 Importance of the Study

In this works we try to solve modification of video in spatial domain by using the DCT feature to represent the frame from video. So our research can provide added value in the related multimedia fields in the video search process. For social side our works can help people find out about watched movies, video clips or movie titles. And for industrial side it can help movie/advertisement/music producers to find out whether the video they made was the first and original and not a plagiarism.