
Development of a technique to identify advertisements in a video signal

Dissertation submitted in fulfilment of the requirements for the degree
Master of Engineering in Computer Engineering at the
Potchefstroom campus of the North-West University

R. Moolman

20673892

Supervisor: Prof. W.C. Venter

November 2012

Declaration

I, Ruan Moolman, hereby declare that the dissertation entitled “**Development of a technique to identify advertisements in a video signal**” is my own original work and has not already been submitted to any other university or institution for examination.

R. Moolman

Student number: 20673892

Signed on the 16th day of November 2012 at Potchefstroom.

Acknowledgements

Thank you to my study leader, prof. W.C. Venter, for giving guidance and helping me complete my work. I also want to thank him for checking my program code and proof reading this dissertation.

Thank you to my colleague, mr. H.A. van Nieuwenhuizen, who helped me with his knowledge of content based identification and steering me in the right direction with my work.

I would like to thank Telkom for their financial support from the bursary they gave me.

The TeleNet research group, for the support every single person gave me.

I also want to thank all my friends, for everything they mean to me and for keeping my life balanced these past two years.

Abstract

In recent years Content Based Information Retrieval (CBIR) has received a lot of research attention, starting with audio, followed by images and video. Video fingerprinting is a CBIR technique that creates a digital descriptor, also known as a fingerprint, for videos based on its content. These fingerprints are then saved to a database and used to detect unknown videos by comparing the unknown video's fingerprint to the fingerprints in the database to get a match. Many techniques have already been proposed with various levels of success, but most of the existing techniques focus mainly on robustness and neglect the speed of implementation.

In this dissertation a novel video fingerprinting technique will be developed with the main focus on detecting advertisements in a television broadcast. Therefore the system must be able to process the incoming video stream in real-time and detect all the advertisements that are present. Even though the algorithm has to be fast, it still has to be robust enough to handle a moderate amount of distortions.

These days video fingerprinting still holds many challenges as it involves characterizing videos, made up of sequences of images, effectively. This means the algorithm must somehow imitate the inherent ability of humans to recognize a video almost instantly. The technique uses the content of the video to derive a fingerprint, thus the features used by the fingerprinting algorithm should be robust to distortions that don't affect content according to humans.

Keywords: *video fingerprinting; advertisement tracking; broadcast monitoring; copy detection; automatic video recognition; perceptual frame hashing; content-based video identification; robust matching*

Contents

List of Figures	x
List of Tables	xii
List of Acronyms	xiii
1 Introduction	1
1.1 Background	1
1.2 Research Question	2
1.3 Objectives	4
1.4 Research Methodology	5
1.4.1 Plan of action for dissertation	5
1.4.2 Study video fingerprinting and related work	5
1.4.3 Design a novel video fingerprinting system	5
1.4.4 Select development environment	6
1.4.5 Experimental set-up	6
1.4.6 Acquire results from tests	6
1.4.7 Draw a conclusion	6
1.5 Dissertation Outline	6
2 Background: Video Fingerprinting	8

2.1	Basic Idea	8
2.2	Uses for Video Fingerprinting	9
2.2.1	Copyright management	9
2.2.2	Advertisement tracking	9
2.2.3	Media tagging	9
2.3	Video Background	10
2.3.1	Basic properties	10
2.3.2	Wrappers and codecs	11
2.4	Important Aspects of Video Fingerprinting	11
2.4.1	Robustness	12
2.4.2	Accuracy	12
2.4.3	Speed	13
2.4.4	Efficiency	13
2.5	Existing Algorithms and Techniques	13
2.5.1	Radial Projection of Key Frames	14
2.5.2	Centroids of Gradient Orientations (CGO)	15
2.5.3	Visual Attention Areas (VAA)	16
2.5.4	Average Luminance over Time	16
2.5.5	Colour Histograms	17
2.5.6	Linear Local Embedding (LLE)	17
2.5.7	Frame vs. Sequence fingerprinting	18
3	Techniques	20
3.1	Introduction	20
3.2	Scale Invariant Feature Transform (SIFT)	22
3.2.1	Algorithm	22

3.2.2	Speeded Up Robust Features (SURF)	26
3.3	Shazam	27
3.4	Key Frame Detectors (KFDs)	29
3.4.1	Jensen Shannon Divergence (JSD)	30
4	Implementation	32
4.1	Frame Fingerprinting	32
4.1.1	Algorithm	33
4.1.2	Database	36
4.1.3	Searching	37
4.2	Video Fingerprinting	38
4.2.1	Key Frame Detector	39
4.2.2	Video detection	41
4.3	Matlab	41
4.4	Visual Basic (VB)	42
5	Tests and Results	43
5.1	Experimental Setup: Frame Fingerprinting	43
5.2	Parameter Optimization	46
5.2.1	Bits per hash	46
5.2.2	Hashes per frame and detection threshold	49
5.3	Hashing Method Optimization	53
5.3.1	Relative or fixed distance	53
5.3.2	Sectional magnitude ratio	54
5.3.3	Alternate hashing method	55
5.4	Codecs and Distortion	57

5.4.1	Codecs	57
5.4.2	Distortion	58
5.5	SIFT vs. SURF	61
5.6	Database Saturation	61
5.7	Key Frame Detector (KFD)	63
5.8	Video Detection	64
5.8.1	Experimental setup	64
5.8.2	Results	64
5.9	Real-time System	65
5.10	Verification and Validation	67
5.10.1	Verification	68
5.10.2	Validation	69
6	Conclusion and Future Work	71
6.1	Conclusion	71
6.1.1	Objectives achieved	72
6.1.2	Contribution	72
6.1.3	Interpreting the results	73
6.2	Recommendations for Future Work	73
6.2.1	Graphics Processing Unit (GPU) implementation	74
6.2.2	Affine invariant key points	74
6.2.3	Fingerprint based on constellation	74
6.2.4	High dimensionality reduction	75
6.2.5	Improved real-time KFD	75
6.2.6	Sectional or sequence fingerprinting	75

Bibliography	76
Appendices	81
A Visual Basic Code	81
B Conference Papers	86

List of Figures

1.1	Diagram of basic video fingerprinting system.	3
3.1	Example of a Gaussian pyramid.	23
3.2	A set of scale space images is created for each level of the Gaussian pyramid of the image as shown on the left. The Difference of Gaussian (DoG) is then calculated by subtracting adjacent scale space images from each other [1].	24
3.3	A peak is found if the DoG point's value is bigger or smaller than all its neighbouring points [1].	25
3.4	Images with SIFT key points displayed on them [2].	26
3.5	Second-order Gaussian partial derivatives on the left and the box filters used by SURF on the right [3].	26
3.6	SURF keeps the image a constant size and changes the filter's size (right) as opposed to changing the image's size and keeping the filter size constant (left) [3].	27
3.7	Shazam algorithm diagrams [4].	28
4.1	Diagram of two matched key points.	35
4.2	Diagram of database structure.	37
4.3	Diagram of the video fingerprinting system.	39
4.4	Example of JSD values containing a spike (shot boundary).	40
5.1	Example frequency histogram of results.	46
5.2	Results for <i>bits per hash</i> test.	47

5.3	Results for <i>hashes per frame</i> test.	50
5.4	ROC curves for tests done with 50 <i>hashes per frame</i> and 16 <i>bits per hash</i> . .	52
5.5	Results for the fixed and relative hashing methods using scale distorted <i>test videos</i>	54
5.6	Results for the sectional magnitude ratio test.	54
5.7	Results for alternate hashing test.	55
5.8	ROC curves for the <i>distortion results</i> of the alternate hashing method. . .	57
5.9	Results for codecs tests with MP4 <i>test video</i> database.	58
5.10	Results for codecs tests with AVI <i>test video</i> database.	58
5.11	Results with Gaussian blur applied to the test frames. The numbers 9, 13 and 17 refer to Gaussian filter sizes of 9×9 , 13×13 and 17×17 used to filter the image.	58
5.12	Results for the test frames with different angles of rotation.	59
5.13	Results for linear scaling/resizing of the test frames.	59
5.14	SIFT key points are displayed on a frame and distorted versions of the frame.	60
5.15	Results for the SIFT and SURF tests.	61
5.16	A frequency histogram of the number of hashes containing a Video Frame Identification (VFID) in the corresponding spot.	62
5.17	Distribution histograms showing how the different hash values are distributed among all the hashes.	62

List of Tables

1.1	Cost (excluding VAT) of 30 second advertisements on South African television stations [5].	4
5.1	Table of video wrapper and codec combinations used in the tests.	44
5.2	Quantitive results for <i>bits per hash</i> test.	48
5.3	Quantitive results for <i>hashes per frame</i> test.	51
5.4	Quantitive results for <i>alternate hashing</i> test.	56
5.5	Key Frame Detector (KFD) test results.	63
5.6	Results of advertisement tracking/broadcast monitoring using video fingerprinting.	65
5.7	Adding and detection times for videos used during testing.	65
5.8	Detection times for SABC 2 video.	66
5.9	Frame detect time (in milliseconds) using MP4 and FLV videos	67

List of Acronyms

AVI Audio Video Interleave

CBIR Content Based Information Retrieval

CGO Centroids of Gradient Orientations

DCT Discreet Cosine Transform

DoG Difference of Gaussian

FDM Frame Difference Measurement

FLV Flash Video

FPR False Positive Rate

GoF Group of Frames

GPU Graphics Processing Unit

GUI Graphical User Interface

JSD Jensen Shannon Divergence

KFD Key Frame Detector

LLE Linear Local Embedding

MPEG Motion Picture Experts Group

MP4 MPEG-4 Part 14

RGB Red Green Blue

ROC Receiver Operator Characteristics

SABC South African Broadcasting Corporation

SBD Shot Boundary Detector

SIFT Scale Invariant Feature Transform

SURF Speeded Up Robust Features

SQL Structured Query Language

TPR True Positive Rate

VAA Visual Attention Areas

VAT Value Added Tax

VB Visual Basic

VFID Video Frame Identification