Office of Graduate Studies
Dissertation / Thesis Approval Form

This form is for use by all doctoral and master's students with a dissertation/thesis requirement. Please print clearly as the library will bind a copy of this form with each copy of the dissertation/thesis. All doctoral dissertations must conform to university format requirements, which is the responsibility of the student and supervising professor. Students should obtain a copy of the Thesis Manual located on the library website.

Dissertation/Thesis Title: New Methods for Detecting Frame Deletion in Modern Video

Author: Hunter Kippen

This dissertation/thesis is hereby accepted and approved.

Signatures:

Examining Committee

Chair   _____

Members   _____

                          _____

                          _____

                          _____

                          _____

Academic Advisor   _____

Department Head   _____

New Methods for Detecting Frame Deletion in Modern Video

A Thesis

Submitted to the Faculty

of

Drexel University

by

Hunter Kippen

in partial fulfillment of the

requirements for the degree

of

Master of Science

May 2019

Dedications

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgments

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Table of Contents

List of Tables

List of Figures

Abstract
New Methods for Detecting Frame Deletion in Modern Video
Hunter Kippen
Dr. Matthew Stamm, Ph.D.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 1: Proposed Approach

1.1   Prediction Error Sequence Extraction

In previous work on frame deletion detection in MPEG-2, the prediction error sequence was extracted directly from the video decoder using the DCT coefficients of the prediction error residuals located in the compressed video file. The prediction error was averaged over all macroblocks in a frame. This prediction error was then stored as a sequence. Due to the nature of the correlation between P-frame prediction errors across a single GOP, any prediction made across GOP boundaries would result in increased prediction error. Wang and Farid showed that for fixed GOP video, the increase in average prediction error is periodic with respect to the number of frames deleted from the video. Stamm's work expands the idea of the prediction error trace by introducing the formulation of a fingerprint signal defined like so:

This fingerprint signal can be used to build a detector for both variable and fixed GOP videos. Despite this advance in detection, the underlying method for extracting the prediction error sequence is the same.

In MPEG-2, a P-frame is encoded by searching the previous anchor frame for the macroblock which incurs the least error. This means that the average prediction error for a single P-frame is only asssociated with the previous I or P-frame. H.264 expands the capabilities of its motion compensation and estimation system by allowing prediction from multiple previous frames (and subsequent frames in the case of B-frames). If the prediction error trace is extracted via the codec for H.264, the average prediction error associated with one frame is comprised of a linear combination of the average prediction error associated with motion vectors that map to the different anchor frames used in the motion estimation and compensation process. Thus, cross GOP predictions are smoothed out in such a way that it makes the fingerprint energy detector in Stamm's paper perform inadequately.

As shown in Fig. **??** the probability of detection using the methodology from MPEG-2 videos

does not translate to H.264. Instead, we propose the following methodology for extracting the prediction error sequence in H.264.

1.1.1   Proposed Prediction Error Sequence Extractor for H.264

The goal of the proposed extraction algorithm is to maximize the probability that should frame deletion exist, a given measurement of the prediction error comes from a cross-GOP prediction. To this end, instead of directly measuring the prediction error from the DCT coefficients from the decoder, we decode the frame of interest and store the motion vectors associated with said frame. For each macroblock in the current frame, we use its associated motion vector to find the x and y coordinates defining the source macroblock which provides the least error mapping from a previous anchor frame. Then for a previous anchor frame, we subtract the pixels in the source macroblock from the destination macroblock in the current frame. This leaves us with a prediction error residual associated with the macroblock. We then calculate the average absolute value of this residual.

We repeat this process for each of D previous anchor frames. Then we store these prediction error values in a matrix M, where each row of the matrix corresponds to the errors associated with a single motion vector, and the columns are the errors associated with each of the previous anchor frames.

After obtaining the M matrix for the current frame, we create the matrix $\tilde{M}$ defined like so:

$$\tilde{M}_{i,j} = \mathbb{1}\left(j == \operatorname*{argmin}_{l}\left(M_{i,l}\right)\right) * M_{i,j} \tag{1.1}$$

Thus $\tilde{M}$ is a copy of M where the non-zero entries in each row correspond to the minimum average error associated with the macroblock and all other elements are zero. Effectively, the column index of the non-zero entry is an estimation of which previous frame the motion vector associated with the macroblock maps to in the decoding process.

Further processing is done on $\tilde{M}$ to output only a single prediction error value. First, $\tilde{M}$ is

reduced into a vector P, such that:

$$P_j = \frac{1}{N_j} \sum_i \tilde{M}_{i,j} \qquad (1.2)$$

Where N_j is the number of non-zero elements in the jth column of $\tilde{M}$. Then, the reported prediction error e is set as the maximum value of P. This entire process is repeated for every P frame in the video. This method of error extraction estimates which previous anchor frame contributes the maximum error per macroblock to the overall prediction error residual obtained by the codec for a given P frame. Since prediction across GOP boundaries results in spikes in the prediction error, the anchor frame that contributes the most error is most likely to be from a different original GOP. In this manner, we obtain a trace that is resilient to advances made in the motion compensation and estimation process in modern codecs as well as robust to variable frame rates and dynamic GOP structures.

1.2 Proposed Detection Algorithm

In addition to the new methods for prediction error extraction, we propose an expanded detection algorithm to better capture the statistical differences between videos. In fact, depending on scene content, video capture settings, and the amount of motion captured in a single recording, the prediction error sequence and fingerprint signal exhibit different structural behavior. This is true even for videos captured from a single camera model. Figure **??** shows this clearly. The two videos were captured from a *insert phone here* using the high quality 1080p capture mode. Both videos were shot using similar scene content, but the amount of motion in the video is different. The top row shows the different prediction error sequences, while the bottom row shows the fingerprint signals.

Thus there is a need to construct a parametric model of the fingerprint signal and the prediction error sequence to capture these statistical variations. We propose constructing an autoregressive (AR) model of both the fingerprint signal and the prediction error sequence. Over the time period of one GOP it can be said that the signals are mostly stationary. The parameters of the AR models are added to a feature vector along with the fingerprint energy described above. In order to capture

the degree to which this is true, the error variance of each AR model is also included. In addition, some basic statistical features are included to scale the overall decision surface. We propose including the mean and variance of both the prediction error sequence and fingerprint signal to the feature vector as well.

Note that after creating a feature vector for each video, the feature vector is quite large. It is inadvisable to create a probabilistic model of the feature vector for classification. Instead, we propose using a discriminative function to map an incoming feature vector directly to the set of natural videos or the set of videos altered by frame deletion. As such, we propose using a Support Vector Machine (SVM) classifier with a Radial Basis Kernel function for classification.

Appendix A: Some Appendix Heading

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at

all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Appendix B: Another Appendix Heading

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like

"Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Vita

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.