Search optimization for JPEG quantization table

Sharon A. Gieske
6167667

Master Reseach Project #2
Credits: 18 EC

Master Program System and Network Engineering

University of Amsterdam
Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervisor
dr. Z. Geradts

Netherlands Forensic Institute
Laan van Ypenburg 6
2497 GB DEN HAAG

June 27th, 2014

CONTENTS

1 INTRODUCTION

1.1 MOTIVATION

Taking pictures is very easy and popular in this digital age. The demand for digital cameras was forecast to be 86 million units for 2013 by Futuresource Consulting[5]. And even though a decline in market share is present for digital cameras, due to the proliferation of smartphones (nowadays all equipped with a camera function) the total number of digital images taken each year is very high. Social media sites which have photo upload functions, such as Facebook and Instagram, report significantly huge numbers on the total upload of images. Facebook alone reported in a white paper [1] that more than 250 billion photos are uploaded to their site, with on average a total upload of more than 350 million photos every day. Statistics on Instagram[1] show a total of 20 billion photos shared on Instagram.

Due to this popularity, digital images are often recovered in forensic investigation. For example, in child pornography cases many digital images are present and are important evidence for the investigation. In such a case it can be very important to identify the origin of images to a specific camera or identify images that come from a common source. This can be done by uncovering traces on pictures that are distinguishable for camera models. One of these traces is the JPEG quantization table, which is specified as a set of 8×8 (integer) values. Separate quantization tables are employed for luminance and chrominance data, where some implementations include two chrominance quantization tables for chrominance-red and chrominance-blue.

In order to match JPEG quantization tables a comparison between 128 values, or 192 when two chrominance quantization tables are present, is made. With over a dozen different camera brands, each developing different models over the years, the number of camera models (and consequently the number of JPEG quantization tables) to be matched against is significantly high. The matching of large databases of images against these camera models will be time costly as for every matching 128 or more integer comparisons are made. This matching process needs to be minimized since time is often limited in forensic investigations. This research will focus on optimizing search through the image databases regarding JPEG quantization tables.

1.2 FOCUS OF RESEARCH

The research question on which is focused is set as: *'How can searching through JPEG quantization tables be optimized?*

In order to answer the research question, this research will focus on the following subquestions:

1. What are identifiable parameters of JPEG quantization tables?

1 http://instagram.com/press/ accessed 03-06-2014

2. How can we quickly decrease the search space for JPEG quantization table matches?

In the following chapters the answers to these questions will be given.

2

RELATED WORK

Research on digital image forensics is a research growing field. It focuses on two main interests, namely source identification and forgery detection. Van Lanh et al. [8] created a survey on digital camera forensics, which describes several techniques in these two fields. Their survey shows the use of intrinsic features of camera hardware and software for camera identification and concludes that hardware features give more reliable and better result. To distinguish between cameras of the same model imperfections of camera the use of hardware features seems to be the best method. Methods for forgery detection also rely on hardware-dependent characteristics but show a lower accuracy rates compared to camera identification methods. In another survey, Weiqi et al. [7] describe methods for passive technology for digital image forensics. They state that in most cases passive forensics can be converted to a problem of pattern recognition.

In forgery detection methods to identify JPEG quantization tables are often used. In reseach by Kornblum[6] quantization tables used by several image software are identified. A software library called Calvin is developed to identify those images who cannot be guaranteed to have been created by a real camera. Reseach by Farid[4] shows a technique for detecting tampering in low-quality JPEG images by identifying a cumulative effect of quantization.

JPEG quantization tables can also be used for source identification. Farid has performed research[2][3] on source identification with the use of JPEG quantization tables. This research states that a sort of camera signature is embedded within each JPEG image due to the used JPEG quantization tables since they differ between manufacturers. Although the JPEG quantization is not perfectly unique, the majority of cases where the same tables are found it is cameras from the same manufacturer that share the same quantization table. It states that (the use of JPEG quantization tables) "*is reasonably effective at narrowing the source of an image to a single camera make and model or to a small set of possible cameras.*" (p. 3)

There exist several projects where JPEG quantization tables are used as camera signatures. For example, the JPEGsnoop[1] project reports a huge amount of information to expose hidden information in images. Another project is the (discontinued) commercial FourMatch[2], which was focused on forgery detection. These projects are not focused on matching large sets of images against a large camera database. In contrast, this research hopes to contribute by creating a decision tree model in order to decrease the search space for large datasets and which can easily be combined further with other (more accurate) source identification techniques.

1 http://www.impulseadventure.com/photo/jpeg-snoop.html
2 http://www.fourandsix.com/fourmatch

3 METHOD

In order to optimize search through JPEG quantization tables the search space needs to be decreased. This reduction in search space can be performed by creating a decision tree model. This model maps observations about an item (specific features of the quantization table) to conclusions about the item's target value (camera model). Decision tree learning is used, which is the construction of a decision tree from class-labelled training tuples, to identify important parameters and their position in the decision tree model. The matching with the use of decision tree model parameters and the matching between full JPEG quantization tables are both benchmarked for time to see whether the search time is accelerated.

The following steps are taken:

1. Gather dataset of JPEG quantization tables. Dataset of pictures and their JPEG quantization table and for JPEG quantization for camera models are needed.

2. Create numerous possible parameters to identify these tables. Rewrite JPEG quantization table as collection of these parameter values.

3. Create training and test set for decision tree learning.

4. Perform decision tree learning to create decision tree model

5. Perform benchmarks: matching with the decision tree model parameters and matching full JPEG quantization tables

3.1 JPEG QUANTIZATION TABLE

EXPLAIN BASICS

4 EXPERIMENTAL SETUP

In this chapter the implementation of the algorithms for predicting popular activities is described.

5 RESULTS

In this chapter the results of this research will be discussed.

6 CONCLUSION

This chapter describes the conclusions that can be drawn from this research. In **??** the results gained in this research will be discussed and in **??** future work for improvement of this research are described.

[1] *A Focus on Efficiency*. White Paper, September 2013. https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-ash3/851560_196423357203561_929747697_n.pdf.

[2] H. Farid. Digital image ballistics from jpeg quantization. Technical report, Dartmouth College, Department of Computer Science, 2006.

[3] H. Farid. Digital image ballistics from jpeg quantization: A followup study. *Department of Computer Science, Dartmouth College, Tech. Rep. TR2008-638*, 2008.

[4] H. Farid. Exposing digital forgeries from jpeg ghosts. *Information Forensics and Security, IEEE Transactions on*, 4(1):154–160, March 2009.

[5] Futuresource Consulting. *Digital cameras in decline, though interchangeable lens segment sees growth*. Press Release, November 2013. http://www.futuresource-consulting.com/2013-11Cameraspressrelease.html.

[6] J. D. Kornblum. Using jpeg quantization tables to identify imagery processed by software. *digital investigation*, 5:S21–S25, 2008.

[7] W. Luo, Z. Qu, F. Pan, and J. Huang. A survey of passive technology for digital image forensics. *Frontiers of Computer Science in China*, 1(2):166–179, 2007.

[8] T. Van Lanh, K.-S. Chong, S. Emmanuel, and M. S. Kankanhalli. A survey on digital camera image forensic methods. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 16–19. IEEE, 2007.