1 Introduction

The importance of (image) databases in forensic science has long been recognized. For example, the utility of databases of fingerprints is well known. Over the past four years, DNA databases have received particular attention and have been featured on the front page of newspapers. These databases have proven extremely useful in verifying or falsifying the involvement of a person or an object as a source of evidence material in a crime, and have led to the resolution of old cases. DNA databases are also playing an increasingly prominent role in the forensic literature. However, a variety of other databases are also crucial to forensic casework, such as databases of faces, and databases of bullets and cartridge cases of firearms. Research into forensic image databases is a rapidly expanding field of scientific endeavor that has a direct impact on the number of criminal cases solved.

Throughout the 20th century, many databases were available in the form of paper files or photographs (e.g., cartridge cases, fingerprints, shoe marks). Fingerprint databases were computerized in the 1980s, and became the first databases to be widely used in networks. In addition, the Bundes Kriminal Amt attempted to store images of handwriting in a database. These databases were all in binary image format. At the beginning of the 1990s, computer databases of shoe marks, tool marks and striation marks on cartridge cases and bullets became available. Improvements in image acquisition and storage made it economically feasible to compile these databases in gray-value or color format using off-the-shelf computers.

Some forensic databases contain several millions of images, as is the case with fingerprints. If databases are large, the forensic examiner needs a method for selecting a small number of relevant items from a database, because if this cannot be achieved the investigation becomes time consuming and therefore either expensive or impossible. The retrieval of similar images from a database based on the contents of each image requires an automatic comparison algorithm that is fast, accurate, and reliable. To formulate such an algorithm, one must first identify which parts or features of the images are suitable for finding correspondences. The development of the retrieval system then requires a multidisciplinary approach with knowledge of multimedia database organization, pattern recognition, image analysis and user interfaces.
1. Introduction

Forensic image databases often contain one or more sub-databases:

- Images that are collected from the scene of crime (e.g., shoe marks) ; with this database it is possible to link cases to each other

- Images of marks that are collected from the suspect (e.g., shoe marks that are made with shoes of a suspect) ; with this database in combination with the database of images that are collected from the scene of crime, it is possible to link suspects with cases

- Reference images (e.g., shoe marks from shoes that are commercially available, that can be used to determine which brand and make of shoe a certain shoe mark is from)

Forensic image databases serve two potential goals: identification or recognition. Recognition aims at distinguishing a particular individual from a limited number of people whose biometric data are known. An example is a system for airport access with face recognition for access control. There are several known faces of terrorists in the database, and if the system gives signals that it recognizes someone as a terrorist (positive), human intervention is needed to verify the conclusion of the system. In these biometric systems false positives are acceptable.

Identification is much more difficult to achieve than recognition, because false positives are unacceptable. Forensic identification is an act of identifying a trace, mark or image with a person or an object. A qualified examiner will testify in court the opinion that a certain trace mark or image is from the person or the object. An example is fingerprint identification.

The compiling of large-scale forensic image databases has made available statistical information regarding the uniqueness of certain features for forensic identification. For example, at the beginning of the 20th century there have been arguments about the number of matching points that are needed for concluding that a fingerprint matches. (depending on the country this could be 8-16 points). Nowadays, however, statistical ranking in fingerprint databases provides more information regarding the uniqueness of fingerprints and the number of points versus the statistical relevancy. Up until very recently, most forensic conclusions were drawn on
experience of the investigator rather than on real statistics. The statistical information now available from databases should result in forensic investigation that is more objective. This is necessary since courts and lawyers are asking questions that are more critical about forensic investigation and conclusions that are based on experience of the investigator instead of real statistics*.

1.1 Main Contributions

The primary objective of this study is to investigate the applicability of image matching algorithms in forensic image database retrieval.

To accomplish these aims we developed, implemented, and evaluated a wide variety of algorithms for measuring correlations in image databases that are used in forensic science. In the evaluation step, we concentrated on the feasibility and accuracy of matching; computational efficiency was not considered.

The research described in this thesis was carried out at The Netherlands Forensic Institute of the Ministry of Justice over a period of ten years. In view of the fact that the research described was performed up to a decade ago, some of the case studies have witnessed the publication of new approaches, sometimes extending the results obtained by us.

In the final discussion an overview is given of the feasibility and usefulness of the different forensic image databases.

* On January 9th 2002 U.S. District Court Judge Louis H. Pollak in Philadelphia, ruled that fingerprint evidence does not meet standards of scientific scrutiny established by the U.S. Supreme Court, and said fingerprint examiners cannot testify at trial that a suspect's fingerprints "match" those found at a crime scene.
1.2 Outline of the thesis

Here, we briefly describe the outline of the thesis and the contents of each of the chapters.

This thesis is based on six peer-reviewed publications and six conference papers:

Chapter 2, State-of-the-Art is based on:

- Geradts Z; Bijhold J; “Content Based Information Retrieval in Forensic Image Databases”; Journal of Forensic Science, 2002, 47(2), pp.40-47 and

This chapter provides an overview of image matching methods and the extraction of features such as color, shape, and structure. The measurement of database performance is considered and several existing databases are discussed.

Chapter 3, Databases of Tool marks is based on:


This chapter describes a database for tool marks we developed. This database (named TRAX) was developed for use on a PC. The database is filled with video-images and administrative data about the tool marks (e.g., width, type of tool mark, etc.). TRAX provides a comparison screen that makes it possible to compare images of tool marks. An new algorithm for the automatic comparison of digitized striation patterns is described and evaluated in this chapter. Results of new experiments with 3D surface scanners are presented and discussed.
1. Introduction

Chapter 4, **Databases of Cartridge Cases** is based on:


This chapter discusses several existing systems for collecting spent ammunition data for forensic investigation. These databases store images of cartridge cases and the marks on them. Image matching is used to create a hit list of the cartridges in the database that have marks that are most similar to the marks on the cartridge case under investigation. In this research the different methods of feature selection and pattern recognition have been described that can be used to optimize the results of image matching.

Chapter 5, **Databases of Shoe marks** is based on:


This chapter describes methods for recognition of shapes in shoe profiles, with a focus on the database REBEZO that contains video-images of footwear designs. Efforts aimed at the automatic classification of outsole patterns are discussed. The algorithm for automatic classification first segments the separate shapes in a shoe profile, after which Fourier-features are selected for the separate elements and classified with a neural network.
Chapter 6, **Databases of Logos of Drugs Tablets** is based on:

- Geradts, Z; Bijhold, J; Poortman, A; Hardy, H; “Databases of logos of drugs tablets” submitted for publication.

In this chapter we evaluate the different approaches to content based image retrieval of the logos of drug tablets. The retrieval methods are compared using a database of 432 illicitly produced tablets. The results for this database are then compared to the MPEG-7 shape description methods, which comprise the contour-shape, bounding box and region-based shape methods. In addition, we compared the log polar method with the MPEG-7 shape-description methods.

Chapter 7: **Summary and Discussion**

This last chapter discusses the results presented in this thesis in the light of recent developments and the value of the different image databases for forensic investigations. Updates are given where appropriate and other approaches to pattern recognition in forensic databases are discussed.