Chapter 5
Digital Camera Photographic Provenance

Matthew Sorell
University of Adelaide, Australia

ABSTRACT

Whether investigating individual photographs or a large repository of images, it is often critical to establish some history of the generation, manipulation and/or distribution of the images, which is to say the provenance. The applications of image provenance are wide, including the detection of steganographic messages and image tampering, the clustering of images with like provenance, and the gathering of evidence which establishes (or refutes) a hypothetical source. This chapter considers published research and identifies research gaps which address the general challenges of digital image provenance with an explicit emphasis on evidence related to the camera or other digital source.

INTRODUCTION

The term *provenance* is traditionally applied to works of art, referring to documentation which relates to the ownership and public visibility of a particular work, but also includes documentation of production, restoration, thefts, expert opinions on condition and valuations, and any other records which help to assess its integrity.

In the realm of criminal evidence, the term *chain-of-evidence* carries a similar meaning, although the context is usually limited to establishing the source of a particular piece of evidence and its *chain-of-custody*, an even more specific term which refers to the documentation and handling of a piece of evidence once it has been received by investigators.

The difficulty with using the *chain-of-evidence* term for digital media is that the term implies a single version from a particular source, which survives as a single object in custody. Where the chain is

DOI: 10.4018/978-1-60566-836-9.ch005

broken (a continuous record cannot be established), the chain-of-evidence breaks down. When it comes to digital media, however, it is entirely feasible that the file contents might be widely distributed, exist in multiple transcoded forms, include overt or covert tampering, be made available to investigators via multiple sources, and that elements of the content might well come from multiple sources.

Provenance is a term which does not carry this linguistic baggage, at least not to the same extent. It is therefore timely to propose an updated definition for *provenance* in the context of digital media. In this work, *provenance* is defined to mean:

The record of evidence of creation, processing, compression, transmission, transcoding, manipulation, referencing, plagiarism, storage, distribution and other electronic transactions of a digital file or, where appropriate, of multiple digital files whose record can be traced to common ancestors.

The applications of tracking the provenance of digital files are very wide indeed, including:

- The identification of leaked information from confidential sources,
- The detection of plagiarism in academic and other allegedly original works,
- The detection of forgeries, through the manipulation of content to produce misleading evidence,
- The detection of hidden tampering, such as steganographic messages and malicious software,
- The control of distribution and copying of electronic files,
- The establishment of a robust record of forensic discovery, seizure, custody, control, transfer, analysis and disposition of digital evidence,
- The determination of a source (person, location, organisation or other reasonable definition) of a particular piece of digital information, or
- The determination of parties who have been in possession of a piece of digital information and who might have accessed, stored, manipulated, transcoded, distributed or otherwise contributed to the provenance record of that information.

Of course, the challenge of establishing provenance is not easy. It is common for there to be no evidence in existence of a particular stage in the provenance of digital media, and even more common for such evidence not to be available to an investigator (even if it exists). In other words, the file itself often does not contain an inherent history. More commonly, any relevant history which is stored in separate digital files might well be only available on a particular computer which is not available to the investigator. Furthermore, even if such provenancial evidence is available through networked information systems, such as logs of mailing lists or other records, it is often intractable to discover such records or establish their relevance to the evidence under consideration. The search for such evidence becomes not a *needle in a haystack*, but a *needle in a needle stack*: the evidence might exist, but it looks the same as a large volume of quite unrelated information.

Having established the general motivation for the terminology in this work, we illustrate a specific aspect of provenance as related to digital still photographs, with a specific (but not exclusive) focus on establishing some key provenance stages in the creation of a still digital photograph at the camera. Similar arguments can be, and indeed are, applicable to other media such as text, software applications, spreadsheets, databases, video, audio and other multimedia. However, there is currently a strong focus on digital still images in the academic literature, in part because of the prevalence of such images available on the Internet, in part because photographs are often considered to be compelling evidence when