Chapter XIV
Digital Camera Source Identification Through JPEG Quantisation

Matthew James Sorrell
University of Adelaide, Australia

ABSTRACT

We propose that the implementation of the JPEG compression algorithm represents a manufacturer and model-series specific means of identification of the source camera of a digital photographic image. Experimental results based on a database of over 5,000 photographs from 27 camera models by 10 brands shows that the choice of JPEG quantisation table, in particular, acts as an effective discriminator between model series with a high level of differentiation. Furthermore, we demonstrate that even after recompression of an image, residual artefacts of double quantisation continue to provide limited means of source camera identification, provided that certain conditions are met. Other common techniques for source camera identification are also introduced, and their strengths and weaknesses are discussed.

INTRODUCTION

In a forensic context, digital photographs are becoming more common as sources of evidence in criminal and civil matters. Questions that arise include identifying the make and model of a camera to assist in the gathering of physical evidence; matching photographs to a particular camera through the camera's unique characteristics; and determining the integrity of a digital image, including whether the image contains steganographic (hidden message) information.

From a digital file perspective, there is also the question of whether metadata has been deliberately modified to mislead the investigator, and in the case of multiple images, whether a timeline can be established from the various timestamps within the file, imposed by the operating system or determined by other image characteristics.

This chapter is concerned specifically with techniques to identify the make, model series, and particular source camera model given a digital image. We exploit particular characteristics of the camera's JPEG coder to demonstrate that such identification is possible, and that even when an

image has subsequently been reprocessed, there are sufficient residual characteristics of the original coding to at least narrow down the possible camera models of interest in some cases.

BACKGROUND

In general, there are four sets of techniques for camera identification. The first uses information specifically embedded by the camera to identify itself (metadata). Metadata is usually specific to a make and model of camera, but not to a specific camera. The second set of techniques can be used to identify a specific camera, either by identifying specific characteristics of the camera (commonly referred to as bullet scratches or fingerprinting). These techniques generally require a candidate source camera for comparison. The third set of techniques relies on characteristics specific to a manufacturer and possibly a series of cameras, particularly the choice of coding parameters, interpolation, and filtering. These techniques are useful for checking consistency with other evidence and can aid the investigation when metadata has been removed. Finally, a wide range of steganographic (watermarking) techniques have been proposed, but these are really only useful for proving ownership of a copyright work and would almost certainly not be deliberately embedded in an image from a deliberately anonymous source. While watermarking might be introduced in a future generation of cameras, this is of no help in tracking the sources of the existing images of interest in the digital domain.

Metadata

The simplest technique for identifying the source camera is to use metadata embedded by the source camera, of which the Exif metadata standard, published by the Japan Electronics and Information Technology Industries Association (JEITA, 2002), is almost ubiquitously supported. In many cases this is in fact sufficient because it is often beyond the skills of camera users to manipulate or remove the Exif metadata header. Two key forensic fields in the Exif metadata are the *make* and *model*. The metadata is easily extracted using a range of photography tools including recent versions of Adobe Photoshop and such shareware applications as IrfanView.

On the other hand, any savvy photographer who wishes to remove or modify Exif metadata will find a range of easy-to-use tools at their disposal on the Internet. Older versions of image manipulation software, such as early versions of Adobe Photoshop, do not recognise the Exif metadata standard and will strip such identification from the file. More sophisticated users can develop their own techniques to edit Exif. It should be noted in particular that Exif metadata is encoded in clear ASCII text and is both easy to read directly in the binary file, and easy to change.

For this reason, it is unreasonable to expect metadata to be present, and if it is present, to be a reliable indicator of the image source. It is true, however, that many photographers of interest are not aware of the existence of metadata, and the author's experience in criminal cases suggests that where metadata is not present this is more often due to inadvertent erasure rather than deliberate action.

Even if Exif metadata is present, other indicators of the source camera identity are useful to establish whether that metadata has been tampered with. For example, the image size should be consistent with the capabilities of the candidate camera. Also, many camera manufacturers define their own proprietary extensions to Exif, or use their own metadata protocol, which can also provide a level of confidence in the metadata present in the file.