DOTTORATO DI RICERCA IN INFORMATICA
IX CICLO
UNIVERSITA' DEGLI STUDI DI SALERNO

Forensic Analysis for Digital Images

Maurizio Cembalo

November, 2010

PhD Program Chair
Prof.ssa
Margherita Napoli

Supervisor
Prof.
Alfredo De Santis

1. PhD Program Chair: Prof.ssa Margherita Napoli

2. Ph.D Committee: Prof. Alfredo De Santis, Prof. Marco Faella, Prof. Domenico Talia

3. Supervisor: Prof. Alfredo De Santis

Day of the defense: April 29th, 2011

Abstract

Nowadays, taking and sharing digital pictures is becoming a very popular activity. This is witnessed by the explosive growth of the digital cameras market: e.g., more than one billion of digital cameras have been produced and shipped in 2010. A consequence of this trend is that also the number of crimes involving digital pictures increases, either because pictures are part of the crime (e.g., exchanging pedopornographic pictures) or because their analysis may reveal some important clue about the author of the crime.

The highly technical nature of computer crimes facilitated a wholly new branch of forensic science called digital forensics. The Digital Forensic Science involves processes such as acquisition of data from an electronic source, analysis of the acquired data, extraction of evidence from the data, and the preservation and presentation of the evidence. Digital Imaging Forensics is a specialization of the Digital Forensics which deals with digital images. One of the many issues that the Digital Imaging Forensics tries to deal with is the *source camera identification problem*, i.e., establish if a given image has been taken by a given digital camera. Today this is a practical and important problem aiming to identify reliably the imaging device that acquired a particular digital image. Techniques to authenticate an electronic image are especially important in court. For example, identifying the source device could establish the origin of images presented as evidence. In a prosecution for child pornography, for example, it could be desirable that one could prove that certain imagery was obtained with a specific camera and is thus not an image generated by a computer, given that "virutal images" are not considered offense. As electronic images and digital video replace their analog counterparts, the importance of reliable, inexpensive, and fast identification of the origin of a particular image will increase.

The identification of a source camera of an image is a complex issue which requires the understanding of the several steps involved in the creation of the digital photographic representation of a real scene. In particular, it

is necessary to understand how the digital images are created, which are the processes which create (and therefore affect) the creation of the digital data, starting from the real scene. Moreover, it is necessary to point out the factors which can be used to support the camera identification and, may be even more important, which are the factors which can tamper the photos and prevent (maliciously or not) the camera identification.

Many identification techniques have been proposed so far in literature. All these techniques generally work by using the sensor noise (an unexpected variation of the digital signal) left by a digital sensor when taking a picture as a fingerprint for identifying the sensor. These studies are generally accompanied with tests proving the effectiveness of these techniques, both in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Unfortunately, most of these contributions do not take into consideration that, in practice, the images that are shared and exchanged over the Internet have often been pre-processed. Instead, it is a common practice to assume that the images to be examined are unmodified or, at most, to ignore the effects of the pre-processing.

Even without considering the case of malicious users that could intentionally process a picture in order to fool the existing identification techniques, this assumption is unrealistic for at least two reasons. The first is that, as previously mentioned, almost all current photo-managing software offers several functions for adjusting, sometimes in a "magic" way (see the "I'm feeling lucky" function on Google Picasa) different characteristics of a picture. The second reason can be found in the way the images are managed by some of the most important online social network (OSN) and online photo sharing (OPS) sites. These services usually make several modifications to the original photos before publishing them in order to either improve their appearance or reduce their size.

In this thesis we have first implemented the most prominent source camera identification technique, proposed by Lukáš *et al.* and based on the Photo-Response Non-Uniformity. Then, we present a new identification technique that use a SVM (Support Vector Macchine) classifier to associate photos to the right camera. Both our implementation of Lukáš *et al.* technique and our SVM technique have been extensively tested on a test-sample of nearly 2500 images taken from 8 different cameras. The main purpose

of the experiments conducted is to see how these techniques performs in presence of pre-processed images, either explicit modified by a user with photo management tools or modified by OSNs and OPSs services without user awareness.

The results confirm that, in several cases, the method by Lukáš *et al.* and our SVM technique is resilient to the modifications introduced by the considered image-processing functions. However, in the experiments it has been possible to identify several cases where the quality of the identification process was deteriorated because of the noise introduced by the image-processing. In addition, when dealing with Online Social Networks and Online Photo Sharing services, it has been noted that some of them process and modify the uploaded pictures. These modifications make ineffective, in many cases, the method by Lukáš *et al.* while SVM technique performs slightly better.