The Art and Science of Digital Visual Media Forensics

Raahat Devender Singh*
Department of Computer Science and Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh, India

Abstract

Ever since the invention of photography, digital visual media has played a pivotal role in shaping our opinions and beliefs by continually affecting our perception of the world and the events that occur therein. For quite some time now, photographs and videos have been providing investigative benefits by serving as evidence repositories that can be used for post-event analysis and decision-making. But while the visual information provided by photographs and videos has incontrovertible forensic utility in today’s world, the widespread availability of powerful and user-friendly content editing software is causing us to become increasingly skeptical of the legitimacy of this information. Therefore, in matters where digital visual content constitutes potential evidence, it becomes paramount to first validate the integrity and authenticity of this content, a task that is performed using the chain of scientific inquiry known as digital visual media forensics.

Introduction

One eye-witness weighs more than ten hearsays. Seeing is believing, all the world over.- Plautus

Mankind has a long history of using pictorial representations of information as a way of disseminating and assimilating knowledge and among all pictorial representations; nothing bears more verisimilitude than photographs and videos. As direct expressions of reality, photographs and videos present an image of the world that influences human opinion the same way real-world stimuli do.

Photographs and videos are regarded as the most accurate ways of documenting people and objects, and the visual information they provide acts as an unbiased and universal record of occurrence of events, which is often used as evidence for making critical decisions and judgments in highly sensitive areas such as journalism, politics, civil litigations, criminal trials, defense planning, surveillance and intelligence operations. The forensic applications of photographs were actually recognized almost as soon as photography was invented.

Forensic Application of Photographs: Origin and History

After the invention of heliography and the creation of the first permanent photographic image in 1826 or 1827 by French inventor Joseph Nicéphore Niépce, several attempts were made to commercialize photography. But it wasn’t until the invention of daguerreotype cameras by French artist and photographer Louis-Jaques-Mandé Daguerre, that camera manufacturing became an industrial procedure [1]. When these cameras became available to the general public in 1839, photography’s potential for identification and documentation of the criminal classes was recognized, making photographs a widely acceptable forensic means of identification. The earliest evidence of photographic documentation of prisoners dates back to 1843 in Belgium. By 1848, police in Liverpool and Birmingham were photographing criminals and by the mid 1850s, English and French authorities had begun encouraging their law enforcement agencies to photograph prisoners, mostly to prevent escapes and document recidivism [2]. The forensic photographic process was standardized in 1888 when Alphonse Bertillon, a French police officer and biometrics researcher, suggested anthropological studies of profiles and full-face shots (which later came to known as mug shots) to identify criminals [3].

Videos as Forensic Evidence and the Rise of CCTV

A photograph captures a single moment in time and is essentially an arrested representation of reality. A video however, presents a record of an event over a period of time, making it a much more potent forensic evidence that still photographs.

In a world with a plethora of multimedia capture devices and ubiquitous surveillance, video data has emerged as a truly indispensable source of information. During a criminal trial for instance, video evidence, if available, is considered particularly inculpatory. Unlike other kinds of forensic evidence like DNA and fingerprints, which are circumstantial in nature and require further inference, videos provide a first-hand eyewitness account of an event.

The first documented instance of use of video footage during a criminal trial was the Rodney King case of 1991. King, a taxi driver and use of excessive force. Despite the video evidence, the officers were acquitted and within hours of the acquittals, the six-day 1992 speed car chase on March 3, 1991. A witness named George Holliday videotaped much of the beating from his balcony and sent the footage to a local news station [4] (Figure 1).

Holliday’s footage was used as evidence during the ensuing trial where the officers were charged with assault with a deadly weapon and use of excessive force. Despite the video evidence, the officers were acquitted and within hours of the acquittals, the six-day 1992...
LA riots broke out, in which 63 people lost their lives and over 2,000 received non-fatal injuries [4].

While the Rodney King case demonstrated an instance of use of a home video as evidence, these days, most of the video evidence admitted in courts comes from CCTV cameras and while CCTV surveillance had begun as early as the 1950s in minor capacity in the UK and US, it wasn’t until the Bulger case of 1993 that CCTV’s investigative benefits were truly recognized.

On February 12, 1993, James Patrick Bulger, a two-year-old boy from Merseyside, England, disappeared from the New Strand Shopping Centre in Bootle, where he had been shopping with his mother. After he was reported missing, tapes from CCTV cameras around the crime scene were examined, one of which showed Bulger being led out of the shopping centre by two ten-year-old boys, Robert Thompson and Jon Venables, who murdered him later that day [5] (Figure 2).

There have been numerous cases over the past years where video evidence has led to successful convictions and has brought justice to those in need. From the David Copeland (London Nail Bomber) case of 1999 [6] to the Manchester Arena bombing of 2017 [7], surveillance videos have demonstrated their forensic utility time and again.

Video evidence has been equally beneficial for the exposition of social injustice. Whether it’s a video of child soldiers or police torture of prisoners, or footage of human rights violations or wildlife abuse, video evidence can accomplish a lot from initiating investigations to expediting cases in courts. For instance, footage of the 2012 Houla massacre in Syria called attention to the incident and led the UN to call for a special investigation [8]. Video evidence played a critical role in the Enderoris Welfare Council’s human rights violation case against the Kenyan government in 2009 [9]. Similarly, videos documenting human rights violations and war crimes during the Sri Lankan Civil War in 2014 were used by the UN High Commission to initiate several investigations into the matter [10].

All these and numerous other cases are demonstrative of the fact that when used in an evidentiary capacity, photographs and videos can be remarkably influential, and in a world where visual evidence holds such power over our decision-making faculties, we must ascertain that this evidence is in actuality what it purports to be, because as compelling as photographs and videos are, the picture they paint may not always be accurate.

Digital Visual Content and Misrepresentation of Facts

Digital content is inherently susceptible to manipulation, and the possibility of tampering is especially worrisome in situations where this content is being used as evidence to make decisions and judgments that have far-reaching consequences.

Content tampering is not a recent trend; earliest known instances of photo tampering can be traced back to 1860, i.e., within half a century of the invention of photography itself (Figures 3 & 4).

These examples illustrate pre-digital era tampering. When photos were recorded on film, manipulation was a task usually performed by skilled photographers in specialized photo labs. In the digital era, the widespread availability of inexpensive yet sophisticated image editing software like Adobe Photoshop, Photo Scape, and Phantasmagoria, has made content manipulation even easier to perform, even by novice individuals with minimal skills. Figures 5 and 6 depict some instances of tampered digital content that surfaced in the media world in the more recent times.
All these examples of content manipulation bear witness to the fact that while a picture may be worth a thousand words, those words may not necessarily be true. Moreover, this issue of content infidelity is not restricted to digital images alone; integrity of video data too cannot be taken for granted any more. In the past, we had fewer reservations about accepting videos as truthful representations of reality, partly due to the inherent complexity of video processing and partly because of the general lack of readily available high-tech video processing tools. But lately, the proliferation of easily attainable and powerful video editing software like Adobe Premier, Light works and Video Edit Magic, has made it possible for digital videos to be altered without any significant effort, even by non-darkroom experts.

The missing footage cases of July 2005 police shooting of Jean Charles de Menezes in London [11] and Kendrick Johnson murder case of January 2013 in Georgia, US [12], to the cases of footage tampering as witnessed during the trial of Bosnian war criminal Radovan Karadžić in 2006 [13] and in the Sandra Bland case of 2015 [14], bring to pass the somber realization that video evidence, like photographic evidence, is not infallible.

Visual Media Evidence: Vulnerable yet Indispensable

In this day and age, use of visual evidence in one capacity or another is simply inevitable, and the susceptibility of this content to manipulations does not preclude it from being admitted as evidence. The indispensability of digital visual evidence is best exemplified by two rulings: “Merely raising the possibility of tampering is insufficient to render evidence inadmissible” (US v. Allen 106 F.3d 695, 700 - 6th Cir. 1997) and “The fact that it is possible to alter data contained in a computer is plainly insufficient to establish untrustworthiness” (US v. Bonallo, 858 F. 2d 1427, 1436 - 9th Cir. 1998).

Thus, in situations where dependence on photographic or video evidence is unavoidable and where reliance on tampered evidence could be detrimental, it becomes paramount to first examine the trustworthiness of the given evidence, before placing complete faith in the fidelity of its contents. And since subjective inspection fails to provide the desired degree of conviction regarding content authenticity, specialized digital forensic tools and techniques have to be relied upon, which are provided by the research field known as digital visual media forensics.

Digital Visual Media Forensics

Digital Visual Media Forensics (DVMF) is a branch of digital forensics that deals with the recovery, scientific examination and investigation, comparison and in-depth evaluation of image/video evidence, mostly in (civil or criminal) legal matters, so as to ascertain the authenticity of the evidence with a high degree of scientific certainty.

Broadly speaking, digital visual media forensics aims at the accomplishment of the following tasks:

Evidence collection

Forensic analysis of a digital photograph or video begins with the collection of the evidence from the storage media (such as an optical disk, flash drive or a magnetic hard disk drive) of a source device (which could be a surveillance camera, a police dash cam or body...
cam, a personal video camera or camcorder or a cell phone). In case the source media or device is damaged, the evidence may need to be properly recovered and/or repaired to ensure preservation of its integrity. The primary reasons that cause damage to a storage media or source device are heat, misuse, everyday wear and tear, environmental conditions prevalent at the time of recording or deliberate sabotage by the offender; the success of recovery/repair of the evidence depends on the exact circumstances and extent of damage to the storage media or device.

Forensic evidence enhancement

Evidence enhancement is perhaps one of the most frequently performed tasks during forensic analysis. More often than not, high quality photographs or video recordings of the event in question are not available, in which case, it becomes vital that the quality of the available evidence be improved, so that the details become clear and the events exhibited by the photograph or video become more apparent to the investigators, attorneys, jurors and judges. Most commonly employed content enhancement techniques, which are generally authorized by the court, include contrast and brightness adjustments, color correction, noise reduction, harsh lighting enhancement, sharpening, video stabilization, frame rate adjustment, masking or facial blur, zooming/cropping, image reconstruction to counteract the effect of motion blur and sub-titling and time-coding.

Content enhancement operations are performed using non-destructive techniques, which ensure that the integrity of the given content is preserved at all times. The extent of achievable enhancement is affected by various factors, such as the initial quality of the photograph or video, technical parameters of the recording device, environmental conditions prevalent at the time of recording, and amount of compression involved.

Evidence authentication and the concept of identifying traces

Visual evidence is not above reproach but the testimony it provides can be difficult to impeach. Therefore, whenever visual evidence of any kind is used as a means to convey important information, it is crucial to establish the trustworthiness of this information, a task that is performed by finding the answers to the following questions:

a. Where is the photograph or video coming from?
b. (How) Has the photograph or video been processed after acquisition?

Question (a) relates to ‘Evidence Provenance’ whereas question (b) relates to ‘Content Authentication’. In order to validate visual evidence, we must first establish its provenance and then authenticate its contents and the key to doing both is the Locard’s Exchange Principle, that is, “Every contact leaves a trace”. Formulated by Dr. Edmond Locard in 1910 [15], this exchange principle is one of the cornerstone principles of all disciplines of forensic science.

Every operation that is applied to a digital photograph or video and which alters its composition in some way, leaves behind subtle evidence of its occurrence. This evidence, generally referred to as a forensic artifact or a forensic fingerprint, is unique to every content processing operation (innocuous or otherwise) and thus serves as an identifying trace for that operation. Identifying traces enable both evidence provenance and content authentication.

Evidence provenance:

Evidence provenance refers to the process of establishing whether or not the photograph or video under consideration was recorded using the device it is claimed to have been recorded with, and that it did not undergo any unauthorized transfers from one media to another, except those that the investigators and forensic analysts were privy to. The forensic process that links the given content to a particular acquisition device is known as Source Camera Identification (SCI).

The identifying traces used by various SCI techniques are furnished by the components of the content generation process itself. Each component of an image/video generation process affect the characteristics of the resulting content in a unique manner, which implies, for instance, that a video captured by a CCTV camera will demonstrate characteristics different from that recorded by a camcorder. Moreover, the characteristics of videos recorded by different kinds of CCTV cameras will also be dissimilar.

Variations in the manner by which a particular image/video generation process impacts the characteristics of the resultant content occur because of the presence (or absence) of certain components in the recording device and because of the differences in the manufacturing and implementation of these components. Careful examination of these variations helps infer particulars of the acquisition device and the content generation process, thereby enabling evidence provenance. An example of such variations is sensor noise, which is a unique kind of noise that every camera introduces in every image or video it records. Sensor noise patterns vary from camera to camera but remain consistent across all the content recorded by a particular camera.

Furthermore, since the operation of various components of the content generation process is dependent on one another, each component can interfere with or even wipe out the traces introduced by the previous component, which implies that characteristics of an earlier stage may not be present in the final content. The forensic analyst can assess provenance of the given content based not only on the presence of identifying traces but also based on the non-existence of expected traces.
Visual Content Authentication:

Once the origin of the image/video is established, the next step is to validate its contents and ensure that they haven’t been tampered with from the time the content was recorded to the time it was presented for authentication. Forensic techniques that help detect the presence of semantic manipulations in digital images and videos are collectively referred to as Tamper or Forgery Detection Techniques.

Every operation that is applied to a digital image or video after it has been generated is considered to be a post-production operation. All such operations affect the underlying characteristics of the digital content and alter its existing configuration, thereby causing the attributes of the affected image/video to exhibit behavior that deviates from the normal behavior displayed by the attributes of an unmodified image/video. These deviations and uncharacteristic behavior are considered to be the identifying traces that enable tamper detection techniques to distinguish authentic content from content that has been modified post-production.

Content enhancement versus content tampering:

It is important to understand the distinction between content enhancement and content tampering from a forensic standpoint. While the primary purpose of content enhancement is to improve the quality of the image/video and highlight important scene details so as to make them easier to comprehend, tampering is a malicious operation that alters the events depicted in the image/video and the meaning conveyed by it, thus rendering it harmful for decision-making purposes. Common content tampering operations include copy-paste forgeries (where certain object are inserted into or removed from an image or video frame) and frame-based forgeries (where a set of frames are removed from a video sequence or the arrangement of frames is altered in some way).

Variants of Digital Visual Media Forensics:

So far, we have considered digital visual media forensics in a general setting without making any assumptions about the kind of access (if any) the forensic analyst has to the components of the image/video generation process, or the availability of any a priori knowledge vis-à-vis the processing history of the given content. Depending on the characteristic features exhibited by a particular forensic scenario, digital visual media forensics can be categorized according to the scheme presented in figure 8.

Active and Passive Forensics:

• Active Forensics: In active forensics, certain identifying traces are either either attached to the content in the form of metadata (such as a hash or signature), or are inserted directly into the content (such as a watermark), at an early stage of its generation. Active forensic methods are implemented directly in the acquisition device and the components of the content generation process cannot be inferred until the identifying traces are inserted into the content. This implies that active forensic methods do not allow for the evaluation of the trustworthiness of arbitrary photographs/videos of unknown origin.

• Passive Forensics: In passive forensics, the analyst has no control over the image/video creation process or the type and/or appearance of the identifying traces. The analyst is also unaware of the specifics of the content generation process and its processing history, and is confined to authenticating the content by inspecting the characteristics it displays. Passive forensic methods rely on two kinds of identifying traces: device characteristics and processing artifacts.

Device characteristics refer to the intrinsic variations in different acquisition devices, which can occur for a variety of reasons, such as different camera manufacturers using different components in their devices, or due to them adjusting the parameter settings of their devices in different ways. Variations can also occur due to unwanted technological imperfections such as sensor defects. Consequently, every acquisition device leaves unique identifying traces on the content it generates, and by studying these traces, inferences can be made about the device itself.

Processing artifacts refer to those traces that are introduced by the various processing operations that an image/video undergoes after it has been generated. Such artifacts are unique to each processing operation, and thus serve as identifying traces for the discernment of that particular operation.

It is important to understand that while both active and passive forensics rely on the detection of identifying traces, in case of active forensics, these traces are embedded or attached to the data intentionally, whereas in case of passive forensics, these traces are introduced by the inherent characteristics of the content acquisition process or the various post-production operation it undergoes during its lifetime.

Blind and non-blind forensics

• Blind forensics: A digital forensic process is called blind if the analyst works in the complete absence of any a priori knowledge regarding the recording device, the content generation process, the original scene captured by the image/video, or any of the processing operations that may have been applied to it. Blind forensic schemes examine the identifying traces present (or absent) in the given content and try to infer the identity of the device that may have been used to capture it or the post-production operations it may have been subjected to.

• Non-blind forensics: Unlike blind forensics, non-blind forensic methods use additional information about the characteristics of
the content generation and processing operations. Such additional information may include knowledge about the identity of the recording device or information regarding the processing history of the content. Identifying traces embedded at the time of content creation are also utilized during non-blind forensics.

Though non-blind approaches help mitigate some of the uncertainty the analyst may face (regarding whether or not the image/video has undergone any post-production processing), they are often infeasible in practical situations. For instance, a non-blind approach cannot help determine the origin of an image/video of unknown origin, because for such an image/video, there is no additional information available aside from the contents of the image/video itself.

Identification and interpretation

The final major task of visual media forensics is the thorough examination of the scene elements depicted in the image/video, so as to identify the objects it contains, and interpret their meaning in relation to the scene. During this phase of forensic analysis, operation like object detection, tracking, and highlighting are performed.

Preliminary object detection is followed by specialized examinations, which are performed in an attempt to make positive identification of people (victims, witnesses, or suspects) or inanimate objects (nametags, license plates, names of streets and buildings, house numbers etc.) depicted in the image/video. Common object identification techniques are facial mapping (i.e., comparing one facial image to other facial images), video grammetry (which is useful for taking measurements of objects, for instance, to estimate the height of the perpetrator), identification of other characteristic features, such as scars and tattoos and subjective inspection methods such as forensic gait analysis and behavioral pattern analysis.

Accurate interpretation of the meaning conveyed by the given evidence is absolutely imperative. During a trial, visual evidence is often left to speak for itself, since it is expected to deliver to the court all the facts it contains within itself. This tendency of letting the evidence “explain itself” renders it vulnerable to misinterpretation. Consider for instance, a case from October 2003 in Florida, where Claudia Muro, a nanny, was accused of aggravated child abuse by her employers [16]. During the trial, a hidden nanny cam video was produced as evidence, which showed Muro violently shaking the 5-year-old baby girl. The jury declared the evidence to be “clear as day” and despite Muro’s repeated claims of innocence, she was convicted.

In March 2006, the case was re-opened and the video evidence was subjected to forensic analysis, which revealed that the evidence was in fact misleading. The nanny cam had actually recorded around 5.5 frames per second, which made gentle motions seem violent when the video was played back at normal frame rate. This discovery exonerated Muro, who, despite being innocent, had to spend 29 months in jail.

This case teaches us an important lesson about the cognitive impact of photographic and video evidence, and how even the slightest misinterpretation of the contents of this evidence can lead to terrible miscarriage of justice.

Conclusion

As documented representations of reality, photographs and videos facilitate post-event analysis by offering potent forensic evidence, which we are becoming increasingly dependent on for making sensitive and momentous decisions in politics, journalism, surveillance, intelligence, civil and criminal litigations. Photographs and videos are especially influential because they provide visual evidence of occurrence of a real world event, which has a kind of authority over our perception that is surpassed only by firsthand real-time assimilation of that event. In a world where visual evidence holds such significance, it is imperative that we be certain of its integrity and trustworthiness before accepting it as an accurate portrayal of reality.

Digital visual media forensics is a chain of scientific inquiry that supports digital image and video authentication and integrity verification, the fundamental goal of which is to preserve the given content in its most original form, all the while conducting a structured investigation that enables validation and interpretation of the information it conveys.

References

10. UN News Centre (2010) Deeming Sri Lanka execution video authentic, UN expert calls for war crimes probe. UN News Centre, USA.