**PhD Dissertation**


**International Doctorate School in Information and Communication Technologies**

DISI - University of Trento

# Active and Passive Multimedia Forensics

Valentina Conotter

Advisor:

Prof. Giulia Boato

University of Trento, Italy

Co-Advisor:

Prof. Hany Farid

Dartmouth College, USA

April 2011

*"The future belongs to those who believe in the beauty of their dreams"*
*E. Roosevelt*

*Ai miei nonni Arnaldo e Zita*

# Abstract

*Thanks to their huge expressive capability, coupled with the widespread use of the Internet and of affordable and high quality cameras and computers, digital multimedia represent nowadays one of the principal means of communication. Besides the many benefits, the wide proliferation of such contents has lead to problematic issues regarding their authenticity and security. To cope with such problems, the scientific community has focused its attention on digital forensic techniques.*

*The objective of this doctoral study is to actively contribute to this field of research, developing efficient techniques to protect digital contents and verify their integrity.*

*Digital Watermarking has been initially proposed as a valuable instrument to prove content ownership, protect copyright and verify integrity, by imperceptibly embedding a message into a documents. Such message can later be detected and used to disclose possible copyrights violations or manipulations. For specific applications, such as copyright protection, the watermark is required to be as robust as possible, surviving possible attack a malevolent user may be willing to apply. In light of this, we developed a novel watermarking benchmarking tool able to evaluate the robustness of watermarking techniques under the attack of multiple processing operators. On the other hand, for specific applications, such as forensic and medical, the robustness requirement is overtaken by integrity preservation. To cope with this aim, fragile watermarking has been developed, assuming that the watermark is modified whenever a tampering occurs, thus its absence can be taken as evidence of manipulation. Among this class of techniques, we developed a prediction-based reversible watermarking algorithm, which allows a perfect recovery of both the original content and the watermark.*

*More recently, passive forensics approaches, which work in absence of any watermark or special hardware, have been proposed for authentication purposes. The basic idea is that the manipulation of a digital media, if performed properly, may not leave any visual trace of its occurrence, but it alters the statistics of the content. Without any prior knowledge about the content, such alterations can be revealed and taken as evidence of forgery. We focused our study on geometric-based forensic techniques both for images and videos authentication. Firstly we proposed a method for authenticating text on signs and billboards, based on the assumption that text on a planar surface is imaged under perspective projection, but it is unlikely to satisfy such geometric mapping when manipulated. Finally, we proposed a novel geometric technique to detect physically implausible trajectories of objects in video sequences. This technique explicitly models the three-dimensional trajectory of objects in free-flight and the corresponding two-dimensional projection into the image plane. Deviations from this model provide evidence of manipulation.*

**Keywords:**
[Digital Image Watermarking, Digital Image Forensics, Digital Video Forensics]

# Contents

# Chapter 1

# Introduction

*This chapter overviews the research field investigated in this doctoral study. In particular, we describe active and passive digital forensic techniques, focusing on digital watermarking and geometric-based forensic methods. The main objectives and the novel contributions of this thesis are also presented. Finally, we describe the organization of this document.*

*"One picture is worth ten thousand words"*
*Chinese proverb*

In today's digital age, our daily life is permeated with digital multimedia content as one of the principal means for communication. As a matter of fact, such information can be created, stored, transmitted and processed in digital format in an extremely easy way, thanks to the wide spread of low-cost cameras and computers and user-friendly editing tools. Besides the economic and technical advantages, the digital information revolution has led to problematic issues concerning multimedia security and reliability. Therefore, it is more and more important to be able to automatically provide protection to digital contents in order to guarantee their truthfulness and security. The scientific community is very active in this field, coming up with sophisticated and accurate methods for authentication and protection.

Digital watermarking has been firstly proposed as a valuable mean to cope with these problems, by imperceptibly embedding a message into a documents. Such message can later be detected and/or retrieved and used to disclose possible copyrights violations or manipulations. This technology is said to be active, since it requires a known information to be embedded onto the content at the time of recording ( or a person to embed it at the time of sending) to make a forensic analysis possible. This may represent a limitation to digital watermarking techniques, requiring a special equipped hardware or a post-processing of the content. In 1993, the idea of a trustworthy camera had been proposed

[49]. This special camera is provided with a watermarking system on-board, in order to automatically embed a digital watermark on the image at the time of acquisition. However, the realization of the trustworthy camera idea implied problematic issues, both for the camera manufacturers and the conumers. Specifically, a common standard protocol is required for all camera manufacturers and the consumers need to accept the reduced image quality due to the embedding of a watermark. These issues limited the realization of this special equipped device. Moreover, the Secure Digital Music Initiative (SDMI) fiasco [27] pointed out serious worries for the security of digital watermarking. In particular, the proposed audio watermarking system was rapidly hacked by a group of cryptography and watermarking researchers form Princeton University, Xerox PARC and Rice University.

Given these limitations for active forensics techniques, the majority of contents available nowadays over the network is not protected, thus being subjected to unauthorized use and manipulation. In such a scenario, where digital watermarks or signatures are not available, passive (or blind) approaches have to be applied to protect and verify the integrity of multimedia contents. The basic idea of passive forensics techniques is that the alteration of a digital media, if performed properly, may not leave visual trace of its occurrence, but it alters the underlying statistics of the content. An accurate analysis can be carried out, without any prior knowledge about the content and alterations can be taken as evidence of forgery or help in tracing back the history of the content.

In the following, we present the research field of active and passive forensics, where this doctoral study took place.

## 1.1 Active Forensics

Active forensics techniques, such as digital watermarking, have been proposed as a valuable mean to prove the content ownership and authenticity and to track copyright violations [25, 38, 5]. Generally, a watermark (an imperceptible digital code) is embedded into a multimedia content before its delivering/sharing. By verifying the presence of such watermark, ownership protection and authentication may be demonstrated by comparing the extracted watermark with the original inserted code. Although the main application behind digital watermarking is copyright protection, this technology can be used for different purposes, such as multimedia indexing, enrichment, broadcast monitoring, fingerprinting, covert communication, quality assessment, error concealment, and others. For general purposes, three main requirements must be satisfied when designing a watermarking schemes:

- *Robustness* : the capacity to preserve the hidden data after processing.
- *Imperceptibility* : the fidelity of the watermarked data to the original one.
- *Capacity* : the amount of data that can be embedded.

It is important to find the right trade-off between such constraints, whose relative importance strictly depends on the application scenario they are designed for. For some

applications, such as copyright and intellectual property protection, robustness is highly desired and required. An illustrative example of such application is given by the case of Cinema Industry vs Sprague and Caridi [1]. Carmine Caridi, a member of the Academy of Motion Picture Arts and Science and Oscar voter, was given with an exclusive and promotional copy of a movie. Although not allowed, he provided Russel William Sprague with such movie. Unfortunately for him, the given film had been previously watermarked and, when copies of the movie were discovered to be illegally sold over the web, the presence of the watermark guided the FBI through a forensic investigation about the responsible of the crime. In particular, they were able to disclose that most of the movies were derived form Motion Picture Academy screeners and the embedded watermark uniquely identified Carmine Caridi as the owner of the original videos all the copies were originated from.

It becomes clear that, for such applications, the major constraint for a mark embedded into a cover work is robustness against manipulations, including a great variety of digital and analog processing operations, such as lossy compression, linear and non- linear filtering, scaling, noise addition, etc. However it is well known that designing an efficient watermarking algorithm is extremely challenging and the research is still in progress, proposing a great variety of solutions [89]. Therefore, the role of performance evaluation through the use of benchmarking frameworks has grown its importance to speed up the research in the field of digital watermarking and stimulate a continuous improvement of the existing techniques by identifying methods weaknesses and failings [101].

On the other hand, some applications, such as integrity proof and authentication, do not require the embedded watermark to be robust. Fragile and semi-fragile watermarks are assumed to be modified or even deleted after any tampering occurs to the content. So they fails to be detectable after any attempts of modification of the content, thus making any tamperings identifiable. The effectiveness of such watermarks are measured by considering imperceptibility, payload capacity, and computational complexity. Robustness is barely achieved due to the challenging integrity requirement. In this class of techniques, reversible watermarking schemes have been designed and largely studied, especially for medical, forensics, artwork, or military applications, which require the integrity of the data to be preserved. Their peculiarity is the possibility to perfectly restore the original host data after the detection or the extraction of the hidden data.

The major drawback of active forensics techniques is that they require a specific signature or watermark to be embedded at the time of recording or a person to embed it later at the time of sending. This limits the application of such methods to special equipped cameras or subsequent processing of the content. The need of such a-priori procedure leads to the definition of such techniques as being active, in contrast to passive techniques which work in absence of any watermark or special hardware.

[1] http://news.bbc.co.uk/2/hi/entertainment/4037901.stm

## 1.2 Passive Forensics

Passive-blind techniques are regarded as the new direction in digital multimedia security as they operate, in contrast of active forensics techniques, in absence of any any special equipped device and do not require the knowledge of any prior information about the content. The core assumption for this class of techniques is the assumption that original non-forged content owns some inherent statistical pattern introduced by the generative processing. Such patterns are always consistent in the un-forged content, but they are very likely to be altered after some tampering processes. Although visually imperceptible, such changes can be detecting by statistical analysis of the content itself, without the need of any a-priori information. Thus they are said to be passive and blind.

Passive forensics techniques can be primarily divided in three categories [140].

Image source identification aims at establishing a link between an image and the device it was generated from (i.e. camera, scanner or cell phone). The basic assumption is that digital pictures taken by the same device are overlaid by a specific pattern, that is a unique and intrinsic fingerprint of the acquisition device.

The second class of digital forensics techniques aims at discriminating between real and computer generated images, based on the assumption that computer and imaging technology are nowadays so sophisticated and accurate that the distinction between virtual and real images is increasingly difficult to be done at simple visual inspection due their high photorealism. This field of research has been intensively explored by the scientific community especially for the legal issues related to it. In 1996 Child Pornography Prevention Act (CPPA) included certain types of digital images in the law against child pornography [2]:

*"2256(8) child pornography means any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct, where: (B) such visual depiction is, or appears to be, of a minor engaging in sexually explicit conduct; (C) such visual depiction has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct."*.

In 2002 such law was struck down by the United States Supreme Court in the case Ashcroft vs Free Speech Coalition [3], declaring the possession and distribution of synthetically generated images constitutionally prohibited. This ruling brought an immediate need for algorithms able to automatically and reliably discriminate between natural and synthetical images in order to prosecute abusers. Many techniques have been proposed in the literature [140]. Generally, the goal is achieved via machine learning algorithms suitably tuned to classify natural and artificial images, exploiting either statistical information present in the natural images or the difference in the acquisition process of the two classes of images.

[2] http://www.cs.dartmouth.edu/farid/ publications/cppa96.html.

[3] The US Supreme Court Ruling in Ashcroft vs. Free Speech Coalition, No. 00-795.

The third class of passive forensics aims at uncovering tampering that possibly occurred in the content. Generally, when an image is forged, no visual artifacts are introduced in the digital image and it is hard to disclose the manipulation at simple visual inspection. However, the underlying image statistics are heavily affected, thus allowing forgery to be treacable. Plenty of techniques have been proposed in the literature, revealing the increasing interest of the scientific community in the field.

Many of the techniques proposed in the literature are very promising and innovative but they all have limitations. None of them offers by itself a definitive solution for tampering detection and authentication, but a general framework could be built to incorporate several, if not all, techniques which address different issues in digital forensics. However, the efforts of the scientific community are rapidly progressing and forensics technique will continue to make it harder and time-consuming, although never impossible, to circumvent such techniques and make a forgery undetectable.

## 1.3 Proposed Solutions and Innovation

In the doctoral study here presented two main areas of research are under investigation: the former finds room in the field of digital watermarking, focusing on the robustness requirement for such techniques, which, as previously stated in Section 1.1, strictly depends on the application scenario we are working in. The latter area of research concentrates on passive forensics techniques for digital multimedia authentication, in particular focusing on geometric-based techniques, both for images and videos.

As far as digital watermarking is concerned, our study started by considering the fact that in this field no general and standard solution has been reached so far, thus leading to a multitude of developed techniques. This lack of standard is due to several factors, among which the heterogeneity of requirements imposed by the application context.
For applications such as copyright protection and digital rights managements, the major constraint for a mark embedded into a cover work is robustness against manipulations, including a great variety of digital and analog processing operations. But robustness is inversely proportional to imperceptibility, i.e. the more robust the watermark, the higher the strength it has, the higher is its impact on the visual quality of the content. Thus designing an efficient watermarking algorithm is extremely challenging and the research is still in progress, proposing a great variety of solutions. As a consequence, the role of performance evaluation through the use of benchmarking frameworks has achieved more and more importance to speed up the research in the field of digital watermarking and stimulate a continuous improvement of the existing techniques by identifying the weaknesses of different methods.
On the other hand, some other applications, such as medical, forensics or military, require preserving the integrity of the data. To this aim reversible watermarking schemes have been largely studied. Their effectiveness is measured by considering imperceptibility,

payload capacity, and computational complexity, while robustness is barely achieved due to the challenging integrity requirement.

Given such robustness requirements, during this doctoral research we contributed in each application scenarios proposing the following approaches :

- **Robustness evaluation benchmarking tool for image watermarking**
  We presented a novel and flexible benchmarking tool, based on genetic algorithms (GA), to assess the robustness of any digital image watermarking system based the perceptual quality of un-marked images. Such quality is optimized through a stochastic approach, by finding the minimal degradation that needs to be introduced in a marked image in order to remove the embedded watermark.
  The major difference with the existing benchmarking tools lies in the possibility to test the selected algorithm under a combination of attacks (not single), and evaluate the relative performances in terms of perceptual degradation. Moreover, as far as we know, our present contribution is the first systematic attempt to apply GA as a benchmarking tool.

- **High capacity reversible watermarking**
  We developed a new reversible image watermarking technique, which applies prediction in the embedding procedure. It is based on the optimization of error prediction schemes by mean of a similarity matching approach. Neighboring pixels of the to-be-predicted value are combined and suitably weighted depending on their directional distances, assigning to similar pixels a greater role in the computation. The algorithm achieves high performances in terms of capacity, computational complexity and quality of the watermarked image.
  The proposed scheme outperforms some well-known state-of-the-art algorithms, both in terms of capacity gain and computational complexity.

Actually, the most of the overwhelming amount of images and videos we are daily dealing with do not contain a digital watermark, and it is likely to be like that for the foreseeable future. In fact, not many cameras are equipped on-chip with watermarking systems and the attitude of watermarking the content after the acquisition before publishing it is not common among average users. Therefore, in absence of standardization of such techniques, there is a strong need of techniques able to assess the integrity and authenticity of digital contents without any prior knowledge. Since a couple of decades, the scientific community started to focus its attention on passive forensics, and so did we. In particular, we primarily studied geometric-based techniques, which make use of measurements of objects in the world and their positions relative to the camera. We investigated both image and video authentication methods, exploiting knowledge of objects in the world and of the process of image formation (projection into the image plane).
One of the advantages of geometric techniques over techniques based on low-level image statistics is that the modeling and estimation of geometry is less sensitive to resolution and compression that can easily confound statistical properties of images and video.

Following, we briefly present our contribution in this field:

- **Geometric-based Image Forensics**
  We implemented a new forensic technique for authenticating text in photographs. Because it is relatively easy to digitally insert text into a photo in a visually compelling manner, it can be difficult to manually determine if text is authentic. Our forensic technique explicitly estimates the perspective projection of text onto a planar surface and detects deviations from this model. In particular, we have shown that inauthentic text often violates the rules of perspective projection and can therefore be detected.
  To the best of our knowledge, in the context of geometric-based forensic techniques, this is a first approach dealing with detecting manipulated text.

- **Geometric-based Video Forensics**
  We developed a forensic technique to detect physically implausible trajectory of objects in video sequences. It explicitly models the three-dimensional trajectory of objects in free-flight and the two-dimensional imaging of the trajectory by a static or moving camera. We have shown that the three-dimensional trajectory can be directly and reliably estimated from a video sequence. Deviations from this model provide evidence of manipulation. This forensic analysis makes minimal assumptions, requires limited user input, and is computationally efficient and effective also on videos obtained from video-sharing websites.
  Although addressed in the robotics and computer vision community, the analysis of projectiles in video has not previously been considered in the forensic community.

## 1.4 Structure of the Thesis

The thesis is organized in 8 chapters describing the investigated research filed together with the main objectives of this doctoral study.

Chapter 2 presents an overview on active forensics techniques (digital watermarking) and their characteristics. In particular, the research area of benchmarking of digital watermarking and reversible watermarking are deeply reviewed. In the following two chapters our major contributions in these fields are reported.

Specifically, in Chapter 3 we describe our new benchmarking framework, developed to assess the robustness of any watermarking systems. We outline the involved optimization procedure introducing genetic algorithms, provide details of the implementation and confirm the tool effectiveness with experimental results.

Chapter 4 presents in details the proposed reversible watermarking technique by describing the modified prediction algorithm, both embedding and detection phases and under/overflow control procedure. Experimental results are reported to prove the effectiveness of the algorithm.

Chapter 5 reviews passive forensic techniques and their principles, primarily focusing on the issue of image forgeries detection. We introduce our major contributions in such field, both for image and video authentication, which are described in the following two chapters.

In particular, Chapter 6 shows our novel forensic tool for detecting if text in an image obeys the rules of the expected perspective projection, deviations from which are used as evidence of tampering. Experimental results are reported to demonstrate the effectiveness of the approach.

In Chapter 7 we present our geometric forensic technique to authenticate projectiles in video sequences. We briefly review the physics law of a trajectory and explain in detail the mathematical foundations of the algorithm. Any detected forgeries are illustrated in an intuitive and geometric fashion. The efficacy of this analysis is demonstrated on videos of our own creation and on videos obtained from video-sharing websites.

Finally, Chapter 8 collects some concluding remarks and discusses the open issues related to active and passive forensics.

# Chapter 2

# Active Forensics

*This chapter presents a concise overview about digital watermarking, along with its characteristics and applications. We then focus our attention on benchmarking tools to measure robustness of digital watermarking schemes and on reversible techniques, research areas which our main contributions are mainly concerned with.*

In today's digital age, with the advent of the Internet, the creation and delivery of content in digital form results to be easy and affordable for everybody, even for non expert users. In such a scenario a huge amount of digital contents is available and subjected to transmission, reproduction and manipulation. Generally, the altered content is visually identical to the original and thus it has become challenging to be identified just at visual inspection, bringing up important issues regarding copyright and ownership protection as well as integrity verification [107, 123]. Unfortunately the current copyright laws are inadequate for dealing with digital data. This has led the scientific community to focus its interest towards developing new protection and authentication mechanisms to cope with such issues. Besides cryptography and steganography, a valuable effort in this direction is based on digital watermarking [159].

Digital watermarking is referred to as the procedure of embedding some information (watermark) into digital multimedia content (image, video or audio), possibly in an imperceptible way not to degrade the quality of the content. Such embedded information can be extracted or detected at any later stage for different purposes, including ownership proof, tamper detection and access control and the like [25, 38, 5].

Its origin and applications date back to ancient time, when first methodologies to create secret communications were first applied. Hidden messages were written so that

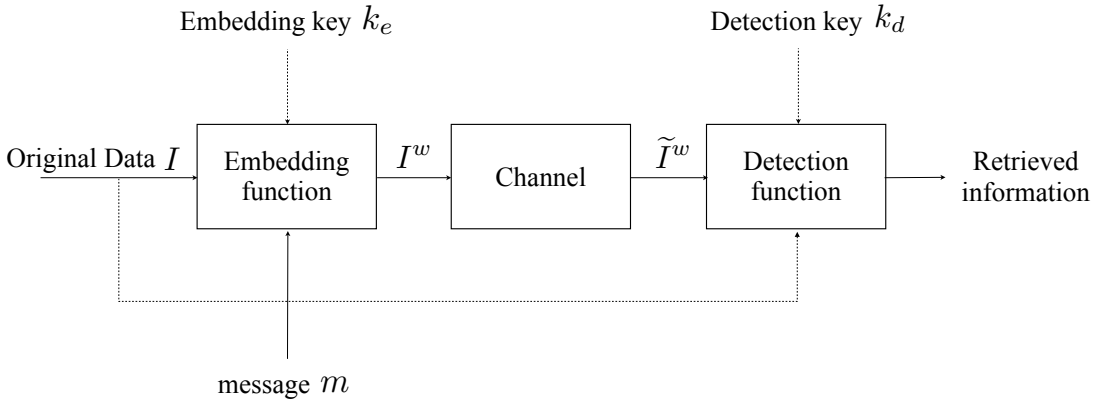



Figure 2.1: General watermarking scheme.

no one, but the sender and intended receiver, could even suspect the existence of the message. A well known example of early steganography (from greek, *'steganos'* meaning "covered or protected" and *graphein* meaning "to write") dates from the sixth century BC, when Herodotus described in his book *"The Histories"* how Histiaeus had tattooed a message upon the shaved head of a slave. Once his hair had grown back, he sent him to his son-in-law Aristagoras telling him to revolt.
First attempts of watermarks appeared in 1282 in Fabriano (Italy) in the production of paper, which bore a visible watermark in order to trace its origin.
Starting from the 1990's, thanks to the remarkable work by Cox et al [25], the digital watermarking applies similar concepts to digital multimedia contents, including images, audio or video files.

## 2.1 Digital Watermarking

According to a commonly accepted definition, a digital watermarking system can be sketched as a communication channel composed by three elements: the transmitter, the communication channel and the receiver [73]. They represent the watermark embedding procedure, the transmission during which any processing operators could be applied to the watermarked data and the watermark extraction (or detection phase), respectively (Fig. 2.1).

Given a message $m$ to be embedded and an embedding key $k_e$, a watermark signal is generated as:

$$w = f(m, k_e) \tag{2.1}$$

where $f(\cdot)$ is a function depending on several parameters, including owner id, image dimensions, random values. In the so called *informed watermarking* the creation of the

watermark could also depend on the original content $I$ (cover object):

$$w = f(m, k_e, I) \tag{2.2}$$

The embedding step consists of inserting the so generated watermark $w$ into the original content $I$ as following:

$$I^w = E(I, w) \tag{2.3}$$

where $I^w$ is the watermarked data and $E$ represents the watermark insertion function. Several embedding procedures have been explored in the literature. One first classification can be done based on the working domain the embedding takes place, e.g spatial domain [139], transformed domain such as Fourier domain [24] or Wavelet domain [176] ( for further reading please refer to [25, 126, 106]).

The so watermarked content is now ready to be delivered over the communication channel, where any kind of processing may happen to the content, producing an altered version of it $\tilde{I}^w$. At any later time the detection phase can be performed to verify the presence or integrity of the watermark:

$$\bar{m} = D(I, \tilde{I}^w, k_d) \tag{2.4}$$

where $\bar{m}$ is the recovered message, $D$ is the detection function and $k_d$ is the detection key. When dealing with *blind* schemes, the original image is not needed at the detector:

$$\bar{m} = D(\tilde{I}^w, k_d) \tag{2.5}$$

Since it is unlikely to have the original content always available, blind techniques are mostly adopted. Whenever the embedding key $k_e$ happens to be the same as the detection key $k_d$ we deal with *symmetric schemes*. Otherwise, if the required keys are different the method is defined *asymmetric* or *public*. These techniques are clearly more secure than the symmetric ones but implies higher complexity.
To be precise, the detector could either be able to extract the watermark, recovering the information it carries, or just act as a binary black-box which just verifies the presence of the mark without reading it [25].

A watermark exhibits many significant characteristics [108]. On the outline, when designing any digital watermarking, three main requirements must be taken into consideration:

- *Invisibility* : the ability of the watermark not to visually alter the content. When this requirement is reached, the original content $I$ and the watermarked one $I^w$ are approximately indistinguishable. However, visible watermark are available under the definition of logos.

- *Robustness* : the capability of the watermark to resist to attacks that may occur over the communication channel. Robust watermarks are able to resist to even

malicious attacks and they are adopted mainly in copyright protection applications. On the other hand, fragile watermarks are likely to be altered even at the slightest modification of the content, making them suitable for authentication and integrity verification.

- *Capacity* : the effective payload that can be successfully embedded. In other words, the number of bits the watermark encodes.

These requirements are inversely proportional with respect to each other. For example, to provide a robust watermark a high strength has to be impressed to it to resist to malicious attacks, but this would lead to a high degradation of the content, thus violating the invisibility requirements.
It is therefore challenging to design a complete watermarking systems and the fulfillment of these requirements strictly depends on the context of the entire system in which they are to be applied. We can summarize the applications for digital watermarking as following [25]:

- *Copyright protection* : the aim is to avoid unauthorized parties claiming content ownership. Such application requires the embedded information to be robust to possible attacks the unauthorized party could perform to remove or forge the watermark.

- *Fingerprinting* : in scenarios where multimedia content is electronically distributed over a network, the content owner would discourage unauthorized duplication and distribution. This can be done by inserting a distinct watermark (or fingerprint) in each copy of the data, thus making any violation traceable. Here the watermark needs to be invisible so not to alter the quality of the provided content and must be robust to deliberate attempts of forgery, removal or invalidation.

- *Copy prevention control* : in a closed system where special hardware is required to copy and/or view the digital content, a digital watermark could be inserted by the hardware itself to control and limit the the number of permitted copies. An example of such a system is the Digital Versatile Disc (DVD).

- *Authentication and tamper detection* : the integrity of the data is particularly important to be preserved especially in application such as medical, news reporting commercial transactions and legal purposes. Clearly a watermark used for authentication purposes should not affect the quality of an image and must be fragile. In particular, if the integrity must be preserved it is desirable that no processing occurs to the content. Fragile watermark is likely to be destroyed by any minimal attack, therefore its absence is evidence of alteration.

In this thesis, we started by focusing our attention on the robustness requirement, trying to understand what this constraint implies. The majority of application contexts requires a watermark to be robust against manipulations, including a great variety of digital and analog processing operations. As a consequence, the role of performance

evaluation through the use of benchmarking frameworks has achieved more and more importance to speed up the research in the field of digital watermarking and stimulate a continuous improvement of the existing techniques.

On the other hand, specific applications deal with sensitive imagery, such as military and forensic investigation, recognition and medical diagnosis. As such, they can not tolerate any alteration of the content. Fragile watermarking techniques have been developed for such contexts, aiming at proofing authentication and verify integrity. Reversible watermarking belongs to this class of techniques. Its peculiarity is the possibility to restore the original host data after the detection or the extraction of the hidden data. Thus, we have lossless data hiding techniques, which allow a perfect host recovery or near-lossless schemes which guarantee very high quality of the recovered data. Reversible watermarking effectiveness is measured by considering imperceptibility, payload capacity, and computational complexity. Robustness is barely achieved due to the challenging integrity requirement.

Following, we review the state-of-the-art of the digital watermark benchmarking and reversible watermarking, research areas where our innovative contributions take place.

## 2.2 Digital Watermark Benchmarking

When delivered over the communication channel the watermarked content may happen to be altered or manipulated by mean of several processing operators, which, depending on their nature, can be classified as *malicious* or *non-malicious*. The formers are common processing operators, such as compression, resizing, cropping or filtering, which do not aim at specifically removing the watermark but may alter the embedded information. The latters are performed by malicious attackers, whose goal is to impair the watermark in order to break the copyright or ownership of the content. According to [56] we can classify such attacks as following:

- *Simple attacks* : all those basic processing operators available with common processing tools, such as filtering, compression or resizing. These attacks are not ad-hoc, since they aim at destroying the watermark impairing the whole content, without any attempt to disclose and identify the embedded information.

- *Synchronization attacks* : all the geometric distortions, such as zooming, rotation, shearing or cropping, belong to this category. They do not really aim at destroying the mark, but at making the detection or extraction of the mark infeasible or even impossible. In particular the watermark is still present in the image and could be eventually still detected.

- *Ambiguity attacks* : attacks whose goal is to produce a fake watermark to confuse the original one. One example is the inversion attack by Craver et al. [26].

- *Removal attacks* : operations which attempt to analyze the watermarked data, consecutively separate the embedded information from the host content. In such a way

the un-marked content is available and usable. Examples are the collusion attack, denoising operators or compression. In this class we have also attacks derived from cryptography, as the brute force search of the embedding and/or detection key or the oracle attack (small variation of the content are applied until the watermark is not detectable anymore).

It is important to keep in mind that a malicious attacker may be willing to keep a high quality of the un-watermarked content while removing the mark for future use.
Generally, such attacks could be performed in combination, in order to reach the goal. However, robustness analysis under combination of attacks is a neglected issue, since no theoretical formulation for such a scenario has been set so far. In general, benchmarking tools are very important and useful for the evaluation of algorithm performances against various attacks, identifying in this way also their weaknesses and failings [101]. As widely known in cryptography, this analysis process can speed up the research in the field of digital watermarking and promote a continuous improvement of the existing techniques. Robustness analysis results important for developers to verify the robustness of their own method with respect to existing algorithms, thus stimulating the research to look for better and better solutions. On the other hand, from a users point of view, a benchmarking tool is useful in order to select the most appropriate watermarking algorithm for the intended application.

In the literature there are already several benchmarking tools, which standardize the process of evaluating a watermarking system on a large set of single attacks [87].

- **Stirmark** [121] : It is a C/C++ software package which applies a number of attacks (one at each time) to the given watermarked content and performs the detection process to check the presence of the mark. The average percentage of the correctly detected watermarks is used as a performance measure to compare different watermarking techniques. It contains three major category of attacks: *(*i) common signal processing (do not change the size of the content); *(ii)* geometric transformations (use of resampling algorithm), *(iii)* other special transformations [122].

  A similar version of this software has been developed for audio contents [86]. It modifies the audio signal by mean of more than 35 available attacks. Their parametrization can be tuned to control the strength of the attacks. As in Stirmark for images, a set of profiles are available, depending on the application scenario the technique is intended for [36].

- **Checkmark** [119] and **CertiMark** [2] : they belong to the so-called "second generation" of benchmarking tools [120]. These packages provide higher quality performance assessment of the watermarking techniques under test. The novel features of these benchmarking tools are the introduction of new types of attacks, the use of a perceptual quality metric to measure the introduced degradation, the possibility to distinguish between watermark detection and decoding, and finally an application driven evaluation.

- **Optimark** [144]: it is a benchmarking for image watermarking, which provides a friendly graphical interface and it implements the same attacks as Stirmark, but with the possibility to create different application-driven application of them. It supports the execution of multiple trials using images (automatically calculating the embedding strength that leads to the chosen image quality), attacks, keys, and messages selected by the user. As output it provides a set of performance indices and graphics characterizing robustness, payload, execution time, and breakdown limits of the under-test-technique. The main drawback of Optimark is the lack of possibility to expand the number of attacks. [145]

- **Mesh Benchmark** [161] : It s a benchmark for three-dimensional mesh watermarking. Common attacks and perceptual distortions are integrated in the software and the MRMS (Maximum Root Mean Square Error) is adopted to evaluate the quality of the un-marked mesh.

### 2.2.1 Innovative Contribution

We developed an innovative and flexible benchmarking tool to assess the robustness level of digital watermarking techniques, whose major difference with the existing approaches is the possibility to test the selected algorithm under a combination of attacks. Moreover we allow the evaluation of the relative performance to be in terms of visual degradation perceived by the Human Visual System (HVS), by introducing a novel robustness metric. In particular, given an application-driven set of image processing operators, it is possible to evaluate the method performance under any combination of them, finding out the most effective parametrization of such set of attacks that removes the watermark while ensuring maximal perceived quality. This problem results to be non linear and multidimensional, thus requiring a suitable optimization technique. We let Genetic Algorithms (GA) [52] support the stochastic search of the parameters to be assigned to the chosen operators, as well as the order they need to be applied in, while optimizing the Weightened Peak Signal to Noise Ratio (WPSNR) [157], the metric used for the visual quality estimation. Therefore, the recovered un-marked image turns out to be as close as possible to the watermarked one.
Up to now, in the field of watermarking, the application of GA has been limited to the embedding procedure, in particular for the selection of suitable parameters to achieve imperceptibility and robustness [182, 105, 65, 142, 17, 83, 143]. Recently, a robust steganographic system, based on GA, was introduced in [175]. In order to create stego-images able to break the inspection of steganalytic systems, the authors employ GA to adjust cover-image values and create the desired statistic features. Finally, in [141], GA are exploited in the context of relational databases watermarking to optimize the decoding threshold.
To the best of our knowledge, our present contribution is the first systematic attempt to apply GA as a benchmarking tool.

Chapter 3 presents a detailed description of such benchmarking framework.

## 2.3 Reversible Watermarking

Generally, it is desirable to keep a high visual quality of a watermarked content not to impair its future use, while satisfying capacity and robustness requirement. In any case, the watermarked content is somehow different from the original one, because embedding extra information always introduces even a slight modification. Moreover, such embedding is mostly an irreversible operation. In some particular application systems, this is not acceptable: medical diagnosis, forensic investigations, art-work, or military require preserving the integrity of the data. In particular, when dealing with sensitive imagery the end-user cannot risk to get wrong information from a distorted data (e.g. a radiologist analyzing a patient's X-ray).
In such a context, reversible watermarking has been proposed, which allows, besides the watermark extraction, also a perfect host recovery. Most of reversible techniques are fragile and they can be exploited for authenticity verification, since the inserted mark is supposed to disappear whenever any modifications to the host content occur. Therefore, the major constraints for a reversible watermarking are imperceptibility, payload capacity and computational complexity, while robustness is scarcely accomplished.

A pioneering work on reversible data hiding has been presented in [48] where Fridrich et al. propose to compress data features to gain space for hiding the watermark in an invertible way. Indeed, the lossless compressed original data as well as their location information are then embedded with the watermark. Following a similar approach, Celik et al. propose a generalized-LSB watermark embedding [16], while in [177] Xuan et al. describe a method based on integer wavelet transform. More recently, Lee et al. present a new proposal for LSB-substitution applied into an integer-to-integer wavelet transform domain [88].

An important class of methods is based on difference expansion. The first proposal is due to Tian, which exploits the redundancy among neighboring pixel values, and therefore pixel differences, to embed the secret data [152]. This method is expanded and modified in various ways and by different authors. For example, in [3] Alattar extends the pixel-pair difference approach to vectors, therefore reducing the location map while increasing the hiding capacity and the computational efficiency. Moreover, Lin et al. work on a three-pixel block [95], Thodi et al. improve neighboring pixels redundancy exploitation in [150], Kamstra et al. introduce sorting of expandable locations [74], and Hu et al. propose to exploit expandable differences both in vertical and horizontal directions [63].

Another significant class of algorithms is based on histogram shifting. Ni et al. propose the basic technique in [115], where the secret data is embedded through histogram modification. It exploits peak or zero points in the histogram to modify the pixel values ranging from the peak point to the zero point. Hwang et al. [66] extend the above algorithm, by using two zero points and one peak point of the histogram as triplet to embed the data.
The above mentioned classes of techniques were improved, in the last few years, by exploiting prediction error techniques in the embedding procedure, generally adopted for lossless compression. In particular, Thodi et al. propose a new technique combining

histogram shifting and the Median Edge Detector (MED) predictor, thus resulting in a higher embedding capacity [149]. In particular, the MED predictor allows to better use the correlation inherent in the neighborhood of a pixel than the difference expansion. Similarly, Hong et al. employ MED with histogram shifting for improving both capacity and image quality [58]. On the other hand, Fallahpour designs a novel scheme based on Gradient Adjusted Prediction (GAP) where prediction errors are first computed and then modified for embedding the watermark [39]. In [84] Kuribayashi et al. apply difference expansion of a generalized integer transform and the JPEG-LS predictor, achieving large capacity with improved image quality. Tsai et al. apply predictive coding and histogram shifting specifically for medical imaging [153]. Tseng et al. propose in [154] the use of various predictors according to embedding capacity given by the prediction error. Similarly, a new method improving the overflow location map and its compressibility is presented by Hu et al. in [64]. Finally, Sachnev et al. propose the sorting of prediction errors to reduce the size of the location map enable by the use of a rhombus prediction scheme [137].

### 2.3.1 Innovative Contribution

We implemented a high-capacity reversible watermarking framework to extend classical prediction-based schemes, exploiting image redundancy and pixel local dependency, with non-local similarity information. Instead of applying the prediction to evaluate the pixel value from a single neighborhood, we expand the prediction to other neighborhoods inside a region surrounding the to-be-predicted pixel in order to benefit from image content redundancy.
In the literature the use of non-local information has been proposed for different purposes, such as motion estimation [173], in-painting [174, 28], compression [129] and denoising [104, 13]. To the best of our knowledge this is the first attempt to apply such a concept to digital watermarking.
Experimental tests confirm that the proposed algorithm allows the complete recovery of the original host signal, and achieves high embedding capacity still providing good quality of the watermarked image. The effectiveness of the method is demonstrated also with respect to some state-of-the-art algorithms, whose performances are outperformed.

Chapter 4 describes in details the proposed reversible framework.

# Chapter 3

# GA-based Robustness Evaluation Method for Digital Watermarking

*This chapter presents a novel and flexible benchmarking tool based on genetic algorithms (GA) and designed to assess the robustness of any digital image watermarking systems. The main idea is to evaluate robustness in terms of perceptual quality, measured by weighted peak signal-to-noise ratio. Through a stochastic approach, we optimize this quality metric, by finding the minimal degradation that needs to be introduced in a marked image in order to remove the embedded watermark. Given a set of attacks, chosen according to the considered application scenario, GA support the optimization of the parameters to be assigned to each processing operation, in order to obtain an unmarked image with perceptual quality as high as possible.*

## Acknowledgment

I would like to deeply thank prof. Francesco De Natale and prof. Claudio Fontanari, University of Trento (Italy), for the inspiring collaboration and valuable comments. Moreover, a warm thank goes to Q. Li and I. J. Cox for kindly providing the code of their algorithm.

Parts of this Chapter appear in:

- G. Boato, V. Conotter, F.G.B. De Natale and C. Fontanari, "Watermarking Robustness Evaluation based on Perceptual Quality via Genetic Algorithms", *IEEE Transaction on Information Forensics & Security*, vol.4, pp. 207-216, June 2009.
- V. Conotter, G. Boato, C. Fontanari, F.G.B. De Natale "Comparison of Watermarking Algorithms via a GA-based Benchmarking Tool", *IEEE International Conference on Image Processing*, Cairo (Egypt), November 2009.
- V. Conotter, G. Boato, C. Fontanari and F. G. B. De Natale, "Robustness and Security Assessment of Image Watermarking Techniques by a Stochastic Approach", *Proc. of SPIE*, San Jose (CA), January 2009.
- G. Boato, V. Conotter and F.G.B. De Natale, GA-based robustness evaluation method for digital image watermarking. *Proc. of International Workshop on Digital watermarking (IWDW)*, Guangzhou (Cina), December 2007.

## 3.1 Introduction

In the age of information technology, it has become easier and easier to access and redistribute digital multimedia data. In this context, the scientific community started focusing on the growing problems related to copyright management and ownership proof. Digital watermarking techniques have been widely studied (see for instance [25, 38, 5], and the references therein) as an effective instrument against piracy, improper use or illegal alteration of contents [96]. Therefore, except for specific applications, the major constraint for a mark embedded into a cover work is robustness against manipulations, including a great variety of digital and analog processing operations, such as lossy compression, linear and nonlinear filtering, scaling and noise addition. Therefore, an accurate robustness analysis becomes important both for developers, to verify the robustness of the own method with respect to existing algorithms, thus stimulating the research to look for better and better solutions, and for an users point who can assess the robustness level and select the most appropriate watermarking algorithm for the intended application. In light of this, the role of benchmarking frameworks has become more and more important. In the literature, there are already several benchmarking tools, which standardize the process of evaluating a watermarking system on a large set of single attacks - Stirmark, CheckMark, CertiMark and OptiMark (see Section 2.2).

In this work, we developed an innovative and flexible tool suitable to assess the robustness of digital watermarking techniques, by introducing a novel metric based on the perceptual quality evaluation for unmarked images. A set of attacks is chosen depending on the application the under-test algorithm is intended for (e.g., copyright or medical applications). Then genetic algorithms (GA) perform the search of near-optimal parameters to be assigned to each image processing operator, as well as the order they need to be applied in, to remove the watermark from the content while keeping the perceptual quality of the resulting image as high as possible. The recovered unmarked image turns out to be as close as possible to the watermarked one in terms of the perceived quality, here measured by means of the weighted peak signal-to-noise ratio (WPSNR). We stress, however, that other metrics could be adopted as well. The major difference with the existing benchmarking tools consists of the possibility to test the selected algorithm under a combination of attacks, evaluating the relative performance in terms of visual degradation perceived by the human visual system (HVS). We point out that the combination of more attacks produces a gain of quality in the unmarked image com- pared to the degradation introduced by one single image processing operator to remove the watermark. On the other hand, taking into account the effect of more than one attack at one time makes this problem nonlinear and multidimensional. Therefore, a suitable optimization technique as GA is needed to converge to a optimal or near-optimal solution. As far as we know, our present contribution is indeed the first systematic attempt to apply GA as a benchmarking tool.

## 3.2 The proposed approach

### 3.2.1 Robustness Evaluation

Visual quality degradation due to the watermark embedding and the removing process is an important but often neglected issue to consider in order to design a fair watermarking benchmark. Given a pattern of possible attacks, the aim of this work is to find a near-optimal combination of them, which removes the mark minimizing the degradation perceived by the HVS [30]. Hence, we need to define a proper quality metric.

In general, several metrics can be used to evaluate the artifacts but the most popular one is the peak signal-to-noise ratio (PSNR) metric. The success of this measure is due to its simplicity but it is well known that it is not suitable to measure the quality perceived by Human Visual System (HSV) [170].

Since advanced watermarking techniques exploit the HVS, the use of the PSNR as a quantitative metric for the distortion caused by a watermarking process might result in a misleading quantitative distortion measurement. Recently, more and more research has been carried on about distortion metrics adapted to the HVS [90]. In [157], a modified version of PSNR, the so-called Weighted PSNR, is introduced: it takes into account that HVS is less sensitive to changes in highly textured areas and introduces an additional parameter, called the noise visibility function ($NVF$), which is a texture masking function:

$$\text{WPSNR} \;=\; 10\log_{10}\frac{I_{peak}^2}{\text{MSE}\times\text{NVF}^2}, \tag{3.1}$$

where $I_{peak}$ is the peak value of the input image $I$.

The NVF can be modeled as a Gaussian to estimate the local amount of texture in the image. The value of NVF ranges from approximately zero for extremely textured areas, and up to one for clear smooth areas of an image:

$$\text{NVF} \;=\; norm\left(\frac{1}{1+\delta_{block}^2}\right), \tag{3.2}$$

where $norm$ is a normalization function and $\delta_{block}^2$ is the luminance variance of the $8\times 8$ block. The NVF is inversely proportional to the local image energy defined by the local variance and identifies textured and edge areas where modifications are less visible. Therefore, for images with no high textured areas, WPSNR is almost equivalent to PSNR.

The main idea of this contribution is to evaluate the robustness of a watermarking system in terms of perceptual quality measured by WPSNR. Namely, fixed a set of admissible image processing operators, the robustness of a method is quantified as:

$$R(q) \;=\; \frac{Q}{M(q)}, \tag{3.3}$$

where $Q$ is a fixed quality threshold, $q$ is the perceptual quality of a watermarked image $I^w$ and $M(q)$ is the maximal perceptual quality of the unmarked image obtained by applying to $I^w$ any combination of the selected attacks.

| | | |
|---|---|---|
| $Q \geqslant M(q)$ | $R(q) \geqslant 1$ | ROBUST |
| $Q \leqslant M(q)$ | $R(q) \leqslant 1$ | NOT ROBUST |

Table 3.1: Robustness evaluation metric. Given $Q$, if the robustness index $R(q)$ is greater than 1, then it is possible to remove the mark from the given image only degrading its maximal perceptual quality $M(q)$ under $Q$, thus the algorithm is said to be robust. If $M(q)$ is greater than the threshold $Q$ (i.e., $R(q)$ is less than 1) the watermark is said to be not robust.

Given $Q$, chosen dependently on the application scenario, and the value of $M(q)$, found according to the process later described in Section 3.2.3, the robustness index $R(q)$ is evaluated according to Equation 3.3. If $R(q)$ is greater than 1, then it is possible to remove the mark from the given image only degrading its maximal perceptual quality $M(q)$ under $Q$. As a consequence, the watermarking algorithm can be declared robust since a large degradation needs to be introduced in the image to remove the mark. On the other hand, the embedded watermark is not robust if $M(q)$ assumes values higher than the threshold $Q$ (i.e., $R(q)$ is less than 1). Such robustness assessment is summarized in Table 3.1.

In this work, we attempt to maximize the function WPSNR, obtaining $M(q)$. Since we consider combinations of attacks, a suitable optimization technique is needed in order to avoid brute force computation.

### 3.2.2 Genetic Algorithm

GA can be used to achieve an optimal or near-optimal solution in multidimensional non-linear problems, such as the one to be handled in this context. GA are robust, stochastic search methods modeled on the principles of natural selection and evolution [52]. GA differ from conventional optimization techniques in that:

- they operate on a group (population) of trial solutions (individuals) in parallel: a positive number (fitness) is assigned to each individual representing a measure of goodness;
- they normally operate on a coding of the function parameters (chromosome) rather than on the parameter themselves;
- they use stochastic operators (selection, crossover, and mutation) to explore the solution domain.

Initially a set of individuals is encoded with chromosome-like bit strings to form an initial population. The cardinality of the set of individuals is called population size [52]. At each iteration, called generation, the genetic operators of crossover and mutation are applied to selected chromosomes in order to generate new solutions belonging to the search space. The optimization process terminates when a desired termination criterion is satisfied, for

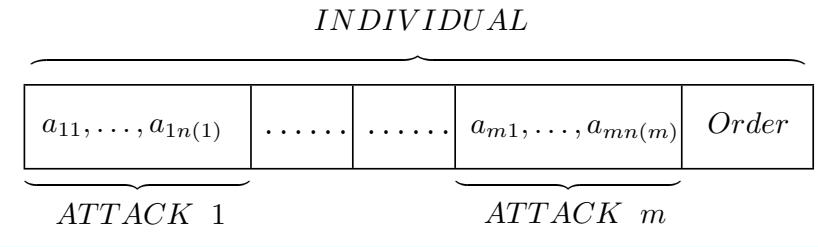



Figure 3.1: Individual (chromosome) definition by Genetic Algorithms. It encodes a possible parametrization of the pre-selected attacks, together with the order they must be applied.

example, the maximum number of generations is reached, or the fitness value is below a fixed threshold.

GA have been widely employed to solve nonlinear optimization problems dealing with a large solution space. Although they are not guaranteed to find out the global optimum, they are less likely to get locked into a local optimum compared to traditional optimization techniques. Moreover, GA allow dealing with a larger searching space than conventional techniques; as a consequence, they are more likely to find suitable solutions to highly nonlinear and constrained problems. The efficiency and the computational complexity of GA are heavily dependent on tuning parameters and can be calculated in terms of the number $\nu_{op}$ of elementary operations required by the algorithms, as follows [52]:

$$\nu_{op} = (P_C + P_M)PK_{max}, \tag{3.4}$$

where $P_C$ and $P_M$ are the crossover and mutation probability, respectively, $P$ is the population size, and $K_{max}$ is the number of iterations.

In this specific application, an individual represents one possible pattern of parameters to be assigned to the preselected attacks, plus the order in which they must be applied (see Figure 3.1). Each attack can be parameterized by $n$ values, according to its specification. The image processing operators must be chosen before running the tool, according to the considered application scenario and are applied to the marked image in order to remove the embedded watermark. The evolution of the population leads to a fine tuning of the parameterization of these attacks, such that they succeed to un-mark the image while introducing a minimal degradation. Moreover, the GA support the optimization of the order of the attacks reaching an optimal or near-optimal solution. In particular, we have experimentally shown the influence of the applied order, since attacks parameterized in the same way but applied with different order may produce a loss of quality of even more than 1 dB.

### 3.2.3 Tool description

In the proposed tool, GA are applied in the detection procedure of the watermarking scheme. An image previously watermarked by the algorithm to be tested and with per-

ceived quality $q$ is attacked with different combinations of selected image processing operators, in order to remove the embedded mark. The aim is to find a near-optimal combination of attacks to apply in order to remove the watermark, while granting a perceptual quality of the resulting image as high as possible. The algorithm robustness is then measured via Equation 3.3. The optimization process is performed by GA and WPSNR is the fitness value to be maximized. We remark that the choice of this fitness function has been done to measure perceptual quality of unmarked images, but the user may adopt any other quality metric. In the following, we briefly depict the operations of the process performed by GA and reported in Figure 3.2:

1. Randomly generate combinations of parameters to be applied to processing operators and convert them into chromosomes. This way, an initial population is created.
   The population size is typically set to 10 times the number of variables the algorithm has to deal with (length of the chromosome). Therefore, it depends on the number of values needed to parameterize each attack we want to perform in the robustness evaluation process. In Section 3.3, we report the experimental analysis, where GA deal with four variables; thus the population size is set to 40. In this work, where the population size is not particularly large, we can reach a good efficiency of GA; in fact, we have small population over a large search space with a consequent fast convergence to optimal or near-optimal solution. Experimental results reported in Section 3.3 show in detail the efficiency of GA in terms of computational time and costs. Note that as the number of attacks to be performed increases, the population dimension and correspondingly the execution time increase as well.

2. Apply each generated attack to the input image and evaluate the WPSNR of each chromosome in the current population which removes the watermark, i.e. which generates an unmarked image, and then create a new population by repeating the following steps:

   - pick as parents the chromosomes with the higher WPSNR, according to the selection rule;
   - form new children (new patterns of attacks) by applying to parents the stochastic operator of crossover with probability $P_C$;
   - mutate the position in the chromosome with probability $P_M$.

   In this work, widely used parameters for genetic operators have been selected (see Section 3.3). Among all individuals of the current population which allow removing the watermark, the one that provides an image with the higher WPSNR will survive to the next generation. We set to zero the fitness value of those chromosomes which do not succeed in removing the mark.

   If in this step no solutions for the problem are found, i.e., none of the individuals of the population succeeds in removing the watermark, another population is reinitialized and the process is repeated until a termination criterion is met (number

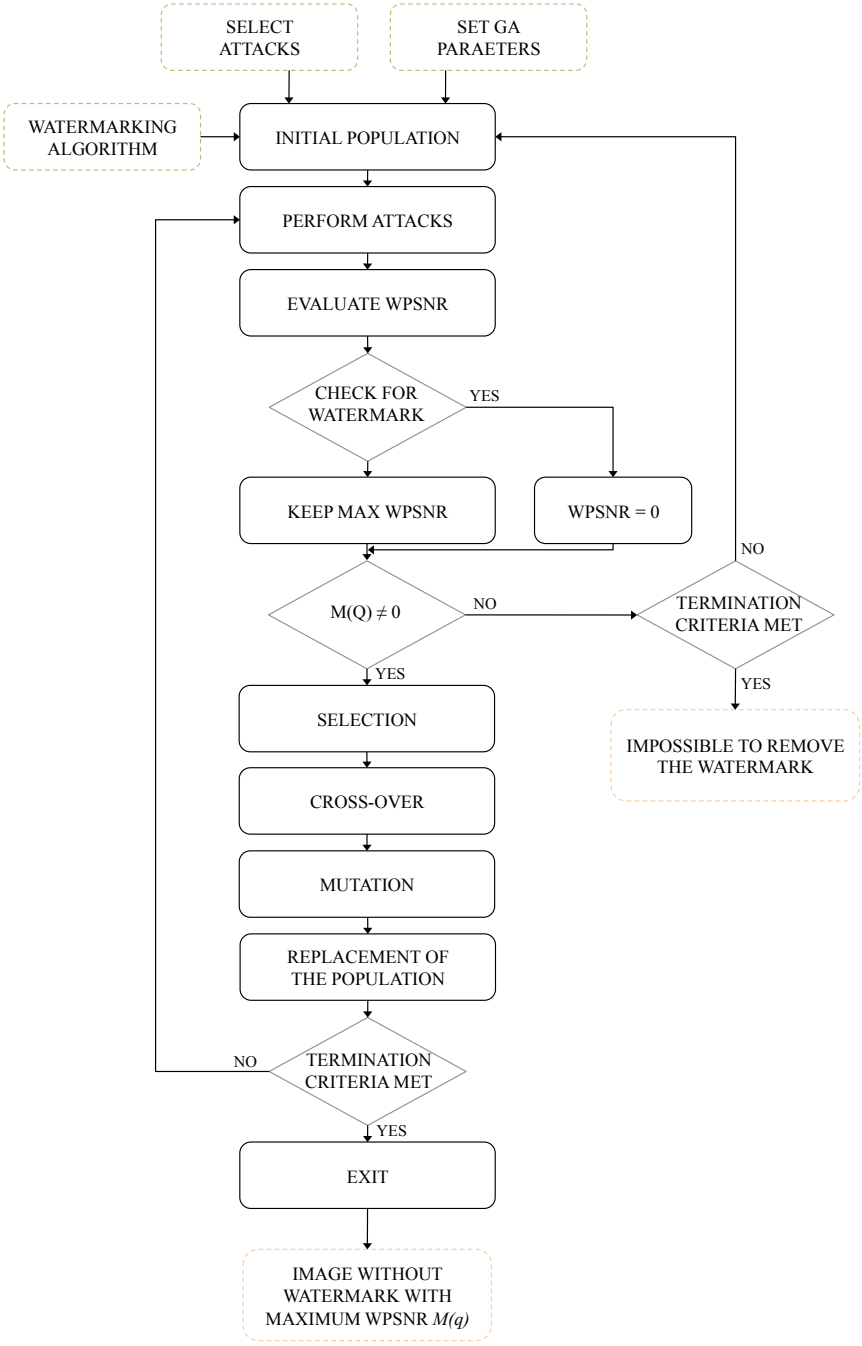



Figure 3.2: Block diagram of the proposed benchmark, involving all the genetic algorithm steps taken to attempt removing the watermark.

of generation exceeded). Consequently, the result of the test is that the analyzed watermarking technique is robust to the selected attacks.

3. A new iteration with the just generated population is processed. This new population provides new attacks parameters, their corresponding fitness values are evaluated, and at every generation the individual with the highest fitness value is kept.

4. The process ends when a given number of generation is exceeded (termination criteria). At that point a near-optimal combination of attacks removing the watermark from the image has been discovered. This way, the lacks of the tested algorithm with respect to the selected attacks are stressed out. At the end of the process, GA return the maximized fitness value, i.e., the maximized WPSNR $M(q)$. According to Equation 3.3, it is possible to calculate the robustness index $R(q)$ and assess the global robustness performances of the watermarking technique.
   In particular, given the quality threshold $Q$, $M(q) \leqslant Q$ means that it is hard to remove the watermark while keeping a high perceptual quality, hence, the watermarking technique is declared to be robust. On the other hand, if $M(q) > Q$, our robustness measure indicates a serious weakness of the scheme corresponding to high quality of the unmarked image.

## 3.3 Results

### 3.3.1 Setup

In this section, we set up the robustness analysis of two perceptual-based watermarking algorithms, the former presented by Barni et al. in [6] and the latter proposed by Li and Cox in [93]. The main difference between them lies in the watermark recovery process, allowing watermark detection in the first case and watermark decoding in the second one.

In order to assess the robustness of these algorithms, we take into account several grayscale common images of size $512 \times 512$. The parameters for the embedding procedure are carefully selected so that the resulting watermarked images present the same WPSNR. They are then processed by the proposed GA-based tool which requires the selection of attacks, as shown in Figure 3.2.

In this work, we take into consideration a combination of some (2 or 3) of the following attacks, each of them parametrized by a single variable:

A. JPEG2000 compression, parameterized by the compression ratio $CR$ ranging from 8 (no compression) down to 0.01 as a float number;

B. JPEG compression, parameterized by the quality factor $QF$ ranging from 100 (no compression) down to 20 as a float number;

| | |
|---|---|
| Population size | ≈10 times number of variables |
| Creation function | Uniform |
| Fitness scaling | Proportional |
| Parents selection | Roulette wheel |
| Crossover function | Single point |
| Crossover probability | 0.8 |
| Elite count | 1 |
| Mutation function | Uniform |
| Mutation rate | 0.1 |
| Stopping criteria | 1000 iterations |

Table 3.2: Genetic Algorithms parametrization used in the experimental session.

C. Additive white Gaussian noise (AWGN), parameterized by the noise power $NP$ expressed in decibels, ranging from 0 to 40, and, for [93] parameterized in terms of standard deviation $\delta$, ranging from 0.1 to 2;

D. Resize, parameterized by the resize factor $RES$ ranging from 1 down to 0.1;

E. Amplitude Scaling, parameterized by the scaling factor $SF$, ranging from 0.1 up to 3.

The choice of the attacks will depend on the application for which the investigated algorithm is intended for. We selected largely used processing operations whose combination represents a realistic scenario. Notice that the choice of both operator and parameter ranges is fully arbitrary and application driven although it affects the computational cost according to Equation 3.4.

Moreover, GA look also for the order the attacks are applied in, since there is no theoretical reason why the attacks should be commutative and indeed we experimentally notice the difference in the resulting image quality when the selected attacks are combined with different orders.

In this work, GA parameters have been tuned according to standard settings [52], as reported in Table 3.2. Genetic algorithms allow a stable convergence of the fitness function. Figure 3.3 illustrates the fitness trend in the case of algorithm [6] and the Baboon image with $q = 59$ dB (similar trends for all other simulations are not reported here). All simulations are carried on an Intel Core 2 Quad CPU at 2.4 GHz, with 2-GB Memory RAM.

GA tool available at : http://www.mathworks.com/access/helpdesk/help/toolbox/gads/

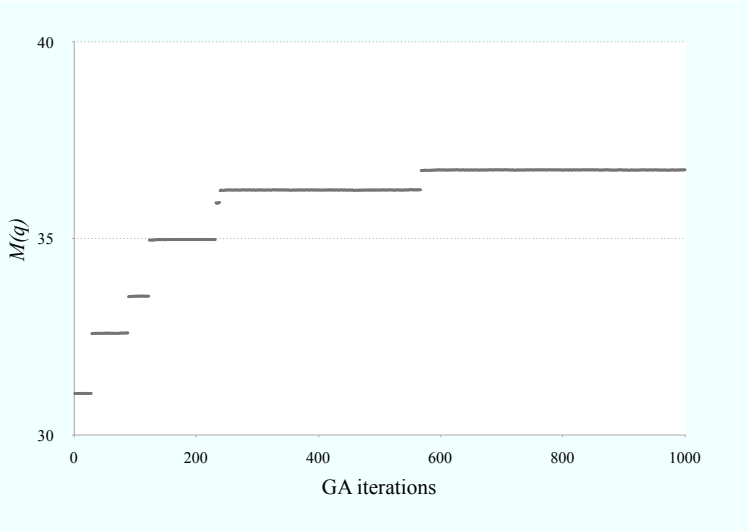



Figure 3.3: Genetic algorithms allow a stable convergence of the fitness function. Shown is the plot of the fitness function *M(q)* ( on the horizontal axis) versus the number of GA iterations (on the vertical axis).

### 3.3.2 Analysis of Barni's algorithm

This watermarking method works on wavelet domain and exploits perceptual masking in order to embed the mark improving invisibility and robustness [6]. The main advantage of this method with respect to existing algorithms operating in the discrete wavelet transform (DWT) domain is that masking is accomplished pixel by pixel by taking into account the texture and the luminance content of all image sub-bands. The mark $w$ (a pseudorandom sequence) is adaptively inserted into the DWT coefficients of the three largest detail sub-bands, as follows:

$$\widetilde{I}_{\theta}^{w} = I^{\theta}(i,j) + \alpha w_{\theta}(i,j) x_{\theta}(i,j), \tag{3.5}$$

where $I_{\theta}^{w}$ are the sub-band coefficients with $\theta \in \{0,1,2\}$, $\alpha$ is the global parameter for strength, $w_{\theta}(i,j)$ is a weighting function considering the local sensitivity of the image to noise, and $x_{\theta}(i,j)$ is the mark to be embedded. To detect the presence of the watermark, the correlation between the extracted DWT coefficients and the watermark is computed by

$$\rho = \frac{1}{3MN} \sum_{\theta=0}^{2} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_{\theta}^{w}(i,j) x_{\theta}(i,j), \tag{3.6}$$

where $2M \times 2N$ is the dimension of the host image, and compared to a threshold $T_{\rho}$, chosen dependently on the admitted false alarm probability [6].

In order to evaluate the robustness of this method and compute the value $R(q)$ defined in Equation 3.3, we apply the following steps:

| Quality parameters | | | | | Attack parameters | | | |
|---|---|---|---|---|---|---|---|---|
| q | $\text{PSNR}_m$ | $\alpha$ | $M(q)$ | $\text{PSNR}_u$ | $CR$ | $NP$ | $RES$ | $Order$ |
| Lena Image | | | | | | | | |
| 47 | 24.6 | 2.11 | 37.8 | 23.3 | 0.48 | 0 | 0.25 | C-A-D |
| 50 | 27.4 | 1.51 | 37.6 | 28.2 | 1.63 | 0 | 0.25 | D-C-A |
| 53 | 30.4 | 1.07 | 37.8 | 26.9 | 0.17 | 1 | 0.3 | C-D-A |
| 56 | 33.4 | 0.75 | 38.4 | 28.2 | 2.66 | 1 | 0.25 | C-A-D |
| 59 | 36.4 | 0.53 | 39.0 | 29.2 | 2.74 | 0 | 0.28 | C-A-D |
| 62 | 39.6 | 0.37 | 40.0 | 30.1 | 2.01 | 0 | 0.3 | A-C-D |
| Baboon Image | | | | | | | | |
| 47 | 24.3 | 1.57 | 32.5 | 19.4 | 4.3 | 1 | 0.25 | C-A-D |
| 50 | 27.3 | 1.11 | 32.9 | 20.2 | 4.02 | 0 | 0.26 | C-A-D |
| 53 | 30.4 | 0.78 | 34.4 | 21.0 | 4.14 | 0 | 0.3 | A-C-D |
| 56 | 33.4 | 0.55 | 34.8 | 21.4 | 3.93 | 0 | 0.31 | C-A-D |
| 59 | 36.4 | 0.39 | 36.5 | 21.9 | 3.86 | 0 | 0.35 | D-A-C |
| 62 | 39.3 | 0.27 | 37.8 | 22.3 | 3.75 | 0 | 0.38 | A-C-D |

Table 3.3: Robustness results of Barni's algorithm under the combination of JPEG2000 compression ($CR$), addition of WGN ($NP$) and resize attack ($RES$), for different initial quality of the watermarked image $q$. For each experiment, quality parameters, all measured in Db, are shown. Starting from the first column: the quality of the watermarked image $q$ and $\text{PSNR}_m$, measured with WPSNR and PSNR respectively, the strength $\alpha$ of the embedded watermark and the quality of the un-marked image $M(q)$ and $\text{PSNR}_u$, measured with WPSNR and PSNR respectively. The parameterizations of the selected attacks are reported, together with the order they must be applied.

I. tune the embedding strength $\alpha$ in Equation 3.5 so that $q = \text{WPSNR}(I^w)$;

II. select the detection threshold $T_\rho$ in such a way that the false positive probability is less than a fixed value $P_{fa}$;

III. run the GA in order to determine $M(q)$ and set $R(q) = \frac{Q}{M(q)}$ as in Equation 3.3.

In our simulations, the detection threshold is adaptively changed depending on the parameter $\alpha$ and imposing a probability of false alarm $P_{fa} \leqslant 10^{-6}$(refer to [6]).

In Table 3.3, experimental results for the Lena and Baboon images are reported. Both are processed with the combination of three image processing operations A, C, and D described in Section 3.3.1. For completeness sake, we report also the PSNR values for both the marked images PSNR ($\text{PSNR}_m$)and the unmarked ones $\text{PSNR}_u$. The elapsed

(a) (b) (c) (d)

Figure 3.4: Marked and Unmarked example images for Barni's algorithm: in panel (a) Lena image watermarked with $q = 47$ dB, while in panel (b) its unmarked version with $M(q) = 37.8$ dB. Baboon image watermarked with $q = 47$ dB is shown in panel (c) and in panel (d) this image unwatermarked with $M(q) = 32.53$ dB.

time for obtaining such results is almost $55s$ per iteration. We underline the weakness of the algorithm with respect to the resize operation, which plays the main role in the watermark removal process. In particular, in the case of Lena, the resulting unmarked image presents a high value of WPSNR; this highlights a robustness limitation of the algorithm under test for this image.

On the other hand, this is no longer true for Baboon: indeed, it is worth noticing that the behavior of the algorithm is image-dependent. In Figure 3.4, examples of the output images (referring to Table 3.3) are shown. The difference is due to the intrinsic nature of the watermarking algorithm. Being a perceptual method, its behavior varies depending on the texture of the content it is dealing with.

To have a whole evaluation of the robustness of the method we are analyzing referring to the single images, we calculate the robustness index $R(q)$ according to Equation 3.3. Averaging over different watermarks we get the plots reported in Figure 3.5 , where the quality threshold is set to 40 and 35 dB. In the first case, the method results to be very robust, since $R(q) \geqslant 1 \ \ \forall q$. In the second case, instead, the dependence of the behavior on the image content is evident. For very highly textured images, such as Baboon, the robustness of the method can be preserved using an embedding strength $\alpha \geqslant 0.5$. On the other hand, for the Lena image, the algorithm turns out to be not robust.

The plots in Figure 3.5 highlight the importance of the choice of the quality threshold $Q$, which strictly depends on the application scenario and greatly influences the robustness assessment. We stress that the user of the proposed tool can properly choose the signal processing operations to give as input to the framework and the results presented here are just an example of application.

As a proof of concept, we carried out simulations also on another image processing operator that is more dangerous and subtle, such as geometric manipulation called shear-

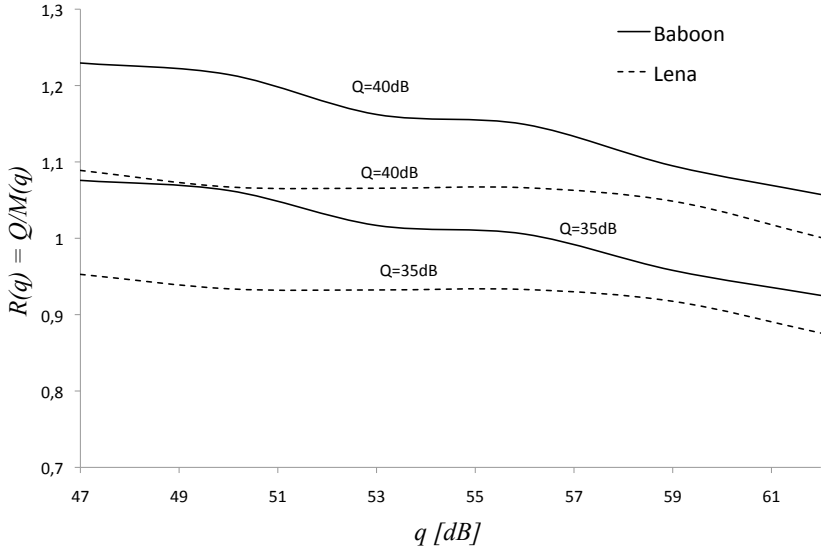



Figure 3.5: Performance plots for Barni's algorithm, when setting the threshold to $Q = 40$ db and $Q = 35$ dB, under the combination of JPEG2000 compression, addition of WGN and resize attack. On the vertical axis is the robustness metric, which, when greater than 1, indicates a good robustness of the algorithm.

ing. This has been done to verify the robustness of the algorithm also from a security point of view. Together with JPEG2000 compression (A) we apply the following attacks:

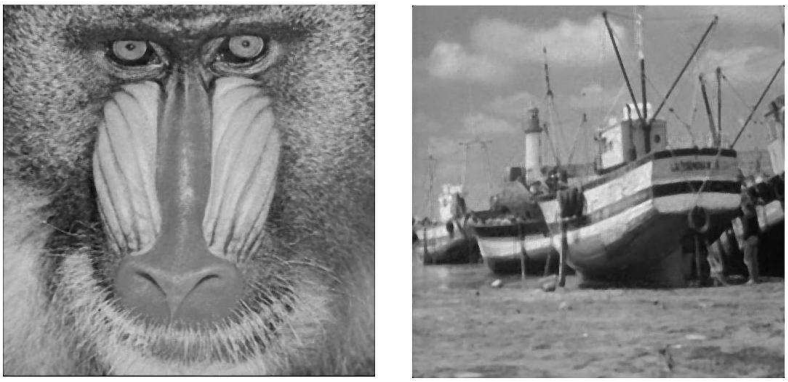
F. Median Filter, parameterized by its window size ($WS$);

F. Shearing, parameterized by X and Y, representing respectively the geometric distortion on the $X$ and $Y$ axis;

In Table 3.4 results are reported for Baboon and Boat images. We underline that in the case of Baboon image all the attacks play a relevant role in the removal process, while in case of Boat image the shearing attack does not influence the process. This could be justified by analyzing the content of the images, which differs in texture. The PSNR values remain pretty low, due to the nature of the embedding algorithm. We point out that the combination of more attacks produces a gain of quality of the un-marked image compared to the exploitation of one single image processing operator, as reported in Table 3.4. In both cases, the median filter can be considered the most effective attack, since it introduces a lower degradation compared to the others, while shearing is devastating. Figure 3.6 shows the good perceptual quality of the attacked Baboon and Boat images.

| Quality parameters | | Attack parameters | | | | |
|---|---|---|---|---|---|---|
| $M(q)$ | $\text{PSNR}_u$ | $CR$ | $WS$ | $X$ | $Y$ | $Order$ |
| Baboon Image | | | | | | |
| 34.66 | 19.83 | 2.64 | 4 | 0.1 | 0 | E-F-A |
| 29.33 | 18.79 | 0.05 | – | – | – | – |
| 34.36 | 19.83 | – | 4 | – | – | – |
| 18.68 | 11.81 | – | – | 1 | 14 | – |
| Boat Image | | | | | | |
| 37.02 | 23.34 | 1.78 | 5 | 0 | 0 | F-E-A |
| 31.71 | 21.77 | 0.05 | – | – | – | – |
| 36.75 | 23.34 | – | 4 | – | – | – |
| 20.12 | 13.65 | – | – | 8 | 1 | – |

Table 3.4: Security analysis of Barni's algorithm under the combination of JPEG2000 compression ($CR$), median filtering ($CR$WS) and shearing ($X, Y$) for different initial quality of the watermarked image $q$. For each experiment, the quality of the un-marked image *M(q)* and $\text{PSNR}_u$, measured with WPSNR and PSNR respectively, are reported in dB. The parameterizations of the selected attacks are reported, together with the order they must be applied. A gain of quality of the un-marked image is achieved exploiting the combination of these attacks, rather than using one single image processing operator.



(a) (b)

Figure 3.6: Successfully attacked and un-marked Baboon (a) and Boat (b) images referred to Table 3.4.

### 3.3.3 Analysis of Li's algorithm

The algorithm presented in [93] by Li and Cox is an important enhancement of traditional quantization index modulation (QIM) methods overcoming the sensitivity to volumetric changes of QIM schemes by adaptively selecting the quantization step size according to a modified version of Watsons model [171]. The authors first describe the need of dither modulation (DM) to address the problem of poor fidelity in some areas of the cover object in traditional QIM schemes due to the fixed quantization step size. DM, first proposed by Chen and Wornell as an extension of the original QIM method [18], introduces a pseudo-random dither signal reducing in such a way the perceptual artifacts caused by quantization. The embedding function embeds the message bit $m_n$ by

$$y_n(x_n, m_n) = Q(x_n + d(n, m_n), \Delta) - d(n, m_n), \tag{3.7}$$

where

$$d(n,1) = \begin{cases} d(n,0) + \Delta/2 & \text{if d(n,0)} < 0 \\ d(n,0) + \Delta/2 & \text{if d(n,0)} > 0 \end{cases} \tag{3.8}$$

and $d(n,0)$ is a pseudorandom signal with a uniform distribution over $[-\Delta/2, \Delta/2]$ and $L$ is the number of samples.

To improve fidelity, Watsons perceptual model is adopted to calculate a slack, that is the maximal distortion allowed for each discrete cosine transform (DCT) coefficient. A slack is then employed to adaptively adjust the quantization step size used in the quantization process of the DCT coefficients.
Operating in the DCT domain, slacks calculated from Watsons model are multiplied by a global constant $G$ in order to get the final quantization step size for each DCT coefficient (notice that $G$ must be known in the decoding phase). Moreover, $G$ is tuned to empirically control the quality of the watermarked image. QIM schemes are generally weak with respect to volumetric scaling. To address this problem, Watsons model has been modified so that the quantization step size is scaled linearly with respect to scaling amplitude of the volumetric attack. This way decoding can be correctly performed. Rational dither modulation (RDM) is then introduced: in particular, the authors propose to calculate the quantization step size for the current block using the slacks of previously watermarked blocks. The final perceptually adaptive RDM method is referred to as rational dither modulation modified Watson model (RDM-MW).

In the detection phase, two signals, namely $S_r(n,0)$ and $S_r(n,1)$, are calculated as follows:

$$S_r(n,0) = Q(r_n + d(n,0), \Delta) - d(n,0) \tag{3.9}$$
$$S_r(n,1) = Q(r_n + d(n,1), \Delta) - d(n,1) \tag{3.10}$$

where $r_n$ is the received signal. The recovered bit is the closest in the Euclidean metric

| q | G | $PSNR_m$ | WD | DWR | q | G | $PSNR_m$ | WD | DWR |
|---|---|---|---|---|---|---|---|---|---|
| Lena Image | | | | | Baboon Image | | | | |
| 47 | 1.149 | 23.28 | 63 | 8.5 | 47 | 0.886 | 25.08 | 45.9 | 9.51 |
| 53 | 0.574 | 29.27 | 32.02 | 14.24 | 53 | 0.443 | 31.09 | 23.01 | 15.52 |
| 59 | 0.284 | 35.36 | 16 | 20.33 | 59 | 0.218 | 37.22 | 11.46 | 21.65 |
| 65 | 0.136 | 41.17 | 8.15 | 26.67 | 65 | 0.104 | 43.56 | 5.81 | 27.99 |
| 71 | 0.053 | 49.47 | 4.48 | 34.45 | 71 | 0.041 | 51.06 | 3.29 | 35.49 |

Table 3.5: Embedding values for Li's algorithm, for both Lena and Baboon image. Starting from the first column, we report the quality of the un-marked image $q$, measured with WPSNR , the embedding parameter $G$, the PSNR f the watermarked image $PSNR_m$, the Watson Distance ($WD$) and the Document to Watermark Ratio ($DWR$).

to the received signal $r$:

$$\hat{m}_n = \underbrace{\arg\min}_{l \in (0,1)} (r_n - S_r(n,l))^2 \tag{3.11}$$

Since one message bit is spread into a sequence of $N$ samples, the code rate is $1/N$ and the detected message bit is determined by accumulating the two Euclidean distances for $N$ samples, as follows:

$$\hat{m}_n = \underbrace{\arg\min}_{l \in (0,1)} \sum_{h=(n-1)N+1}^{nN} (r_h - S_r(h,l))^2 \tag{3.12}$$

with $n = 1, 2, \ldots, L/N$. This watermarking method implies a decoding process, which is evaluated in terms of bit-error rate (BER).
In this analysis context, we embed a message of length 8129 using a 1/31 rate repetition code, following the reference paper [93], and the BER threshold is fixed to 0.2.
In order to compute the value $R(q)$ defined in Equation 3.3, we proceed as follows:

I. tune the global constant $G$ such that $q$ =WPSNR($I_w$);

II. fix a BER threshold $T_{BER}$ discriminating between watermarked and un-watermarked images;

III. run the GA in order to determine $M(q)$ and set $R(q) = \frac{Q}{M(q)}$ as in Equation 3.3.

Concerning point I, we selected $G$ in order to have values of $q$ and corresponding $PSNR_m$ as reported in Table 3.5. The document-to-watermark ratio and the Watson

(a) (b) (c)

Figure 3.7: Lena image, marked with Li's algorithm: in panel (a) with $q = 47$ dB, in panel (b) with $q = 53$ dB and in panel (c) with $q = 65$ dB.

distance of the marked image are also reported, following the setting choices in [93]. Figure 3.7 shows three images of Lena, watermarked in order to reach a quality of $q = 47, 53, 65$ respectively, following embedding parameters of Table 3.5.

First we test the algorithm under the combination of attacks suggested in the reference paper: JPEG compression (B), AWGN (C), and amplitude scaling (E). Results for the Baboon and Lena images are reported in Table 3.6. The elapsed time for obtaining such results is almost $100s$ per iteration. As underlined in the description of the algorithm, we can notice a weakness with respect to JPEG compression. This attack is able to remove the mark while introducing a minimal degradation in the resulting image. This is not surprising considering QIM-based algorithms. Notwithstanding, these tests allow us to demonstrate the effectiveness of the proposed tool.

Figure 3.8 plots the robustness index and underlines once again the weakness of the algorithm, since $R(q) < 1 \ \forall q$. Notice the similar behavior of the algorithm for different images.

A further experimental analysis is carried out in order to analyze the algorithm under the effect of JPEG2000 compression (A) instead of classical JPEG. This attack has been chosen because it is expected to become the new standard for image compression; it is, therefore, interesting to examine the robustness of watermarking algorithms with respect to this attack. The elapsed time for obtaining such results is about $120s$ per iteration.

In Table 3.7, the obtained results for the Baboon and Lena images are reported. It is clear that as the quality of the marked image increases, it becomes easier and easier to remove the mark, introducing little degradation in the image. It is worth stressing the crucial role of JPEG2000 compression and AWGN in the watermark removing process, while the amplitude scaling does not enter into play: indeed the algorithm has been designed to be resistant to this last kind of attack.

| Quality parameters | | | Attack parameters | | | |
|---|---|---|---|---|---|---|
| q | $M(q)$ | $\text{PSNR}_u$ | $QF$ | $\delta$ | $SF$ | $Order$ |
| Lena Image | | | | | | |
| 47 | 49.4 | 24.5 | 54 | 1 | 0.98 | C-B-E |
| 59 | 54.0 | 36.3 | 89 | 0.9 | 0.98 | E-C-B |
| 71 | 54.4 | 45.3 | 98 | 1 | 0.98 | E-C-B |
| Baboon Image | | | | | | |
| 47 | 51.4 | 25.8 | 64 | 1.2 | 0.98 | E-B-C |
| 59 | 55.2 | 37.6 | 91 | 1.3 | 0.98 | B-E-C |
| 71 | 55.4 | 47 | 98 | 0.1 | 0.99 | B-C-E |

Table 3.6: Robustness results of Li's algorithm under the combination of JPEG compression ($QF$), addition of WGN ( $\delta$) and amplitude scaling attack ($SF$). Starting from the first column: the quality of the watermarked image $q$ and the quality of the un-marked image $M(q)$ and $\text{PSNR}_u$, measured with WPSNR and PSNR respectively. The parameterizations of the selected attacks are reported, together with the order they must be applied.

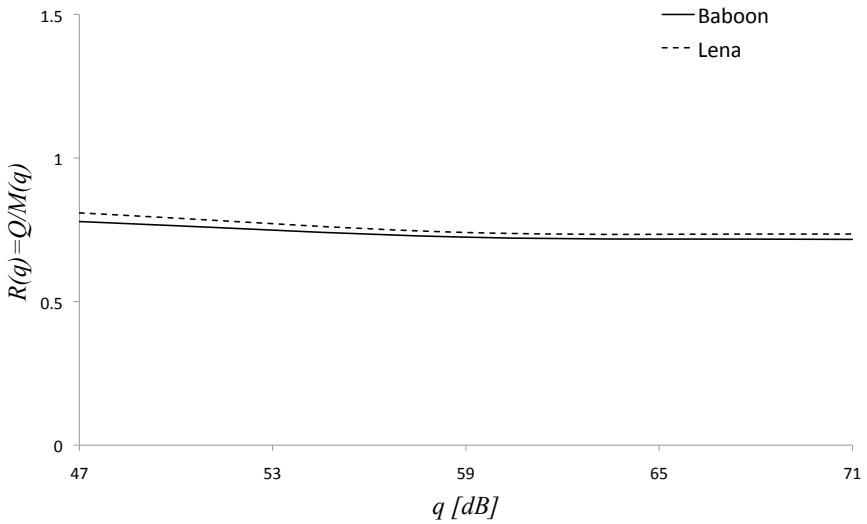



Figure 3.8: Performance plots for Li's algorithm, when setting a quality threshold $Q = 40$ db, under the combination of JPEG compression, addition of WGN and amplitude scaling attack. The robustness metric, $R(q)$, reported on the vertical axis, indicates a weakness of the algorithm to the selected attacks, since $R(q) \leq 1$.

| Quality parameters | | | Attack parameters | | | |
|---|---|---|---|---|---|---|
| q | $M(q)$ | $\text{PSNR}_u$ | $CR$ | $\delta$ | $SF$ | $Order$ |
| Lena Image | | | | | | |
| 47 | 37.6 | 22.9 | 0.17 | 1.2 | 0.97 | E-C-A |
| 53 | 42.1 | 28.3 | 0.31 | 2.0 | 0.97 | A-C-E |
| 59 | 45.6 | 33.1 | 0.48 | 2.5 | 0.97 | C-A-E |
| 65 | 49.8 | 38.1 | 0.85 | 0.8 | 0.98 | A-C-E |
| 71 | 53.2 | 43.4 | 3.87 | 2.2 | 0.98 | C-A-E |
| Baboon Image | | | | | | |
| 47 | 35.8 | 22.3 | 0.55 | 1.8 | 0.97 | C-A-E |
| 53 | 40.9 | 27.2 | 1.12 | 3.1 | 0.97 | A-E-C |
| 59 | 45.7 | 33 | 1.8 | 0.8 | 0.98 | C-E-A |
| 65 | 49.2 | 37.6 | 3.53 | 4.0 | 0.97 | A-E-C |
| 71 | 55.1 | 44.9 | 6.01 | 1.8 | 0.98 | A-C-E |

Table 3.7: Robustness results of Li's algorithm under the combination of JPEG2000 compression ($CR$), addition of WGN ($NP$) and amplitude scaling attack ($SF$). Starting from the first column: the quality of the watermarked image $q$ and the quality of the un-marked image *M(q)* and $\text{PSNR}_u$, measured with WPSNR and PSNR respectively. The parameterizations of the selected attacks are reported, together with the order they must be applied.

We would like to underline, once again, the benefits of the search of the more suitable order the attacks should be applied. Referring to Table 3.7, we can notice the difference in the resulting bit error rate and in the maximum $M(q)$ when different orders are applied (see Table 3.8) to Baboon image watermarked with $q = 47dB$.

By computing the robustness index $R(q)$ of Equation 3.3, we get the plot in Figure 3.9. By setting the quality threshold to 40 dB, we can assess the robustness of the under-test-algorithm for values of the perceptual quality of the input image $q$ lower than 53 dB ($R(q) > 1$ for $q < 53$). It means that the mark has to be embedded with a quite high strength to be robust in this case. Once again, notice the uniform behavior of the algorithm for different images. This highlights a big advantage of this algorithm: it seems to be independent of the image content, thus allowing a wider application. In order to verify this statement, we have repeated the last experiment over ten different standard images (Baboon, Lena, Boat, Cameraman, Peppers, Barbara, Goldhill, Clown, Airplane, Walkbridge). Results have been plot in Figure 3.10, where both mean and variance are reported for different values of $q$ and $Q = 40$dB. This provides clear evidence of the image independency of the algorithm, since the variance of WPSNR is at most 5%.

| $M(q)$ | BER | $Order$ |
|---|---|---|
| 35.78 | 0.21 | A-C-E |
| 35.78 | 0.208 | A-E-C |
| **35.83** | **0.222** | **C-A-E** |
| 35.61 | 0.191 | C-E-A |
| 35.62 | 0.191 | E-C-A |
| 35.63 | 0.193 | E-A-C |

Table 3.8: Performance comparison for Li's algorithm, when the same set of attacks is applied with different orders. results are reported in terms of BER and quality of the attacked image, for Baboon image marked with $q = 47$ dB.

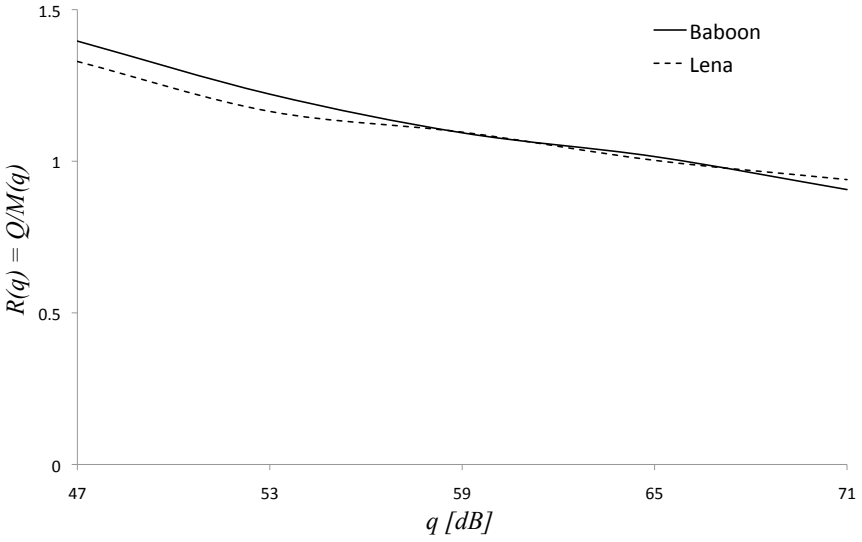



Figure 3.9: Performance plots for Li's algorithm, when setting a quality threshold $Q = 40$ db, under the combination of JPEG compression, addition of WGN and amplitude scaling attack. The robustness metric, $R(q)$, reported on the vertical axis, indicates a robustness of the algorithm to the selected attacks, since $R(q) \geq 1$.

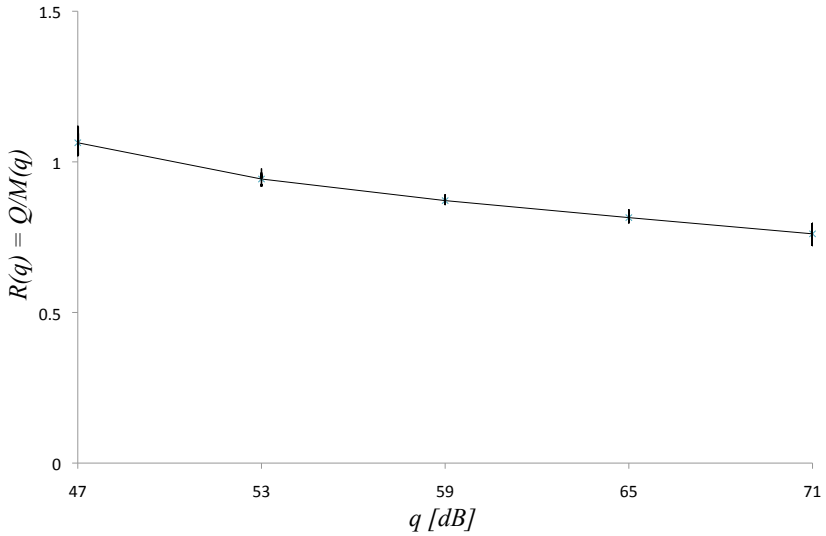
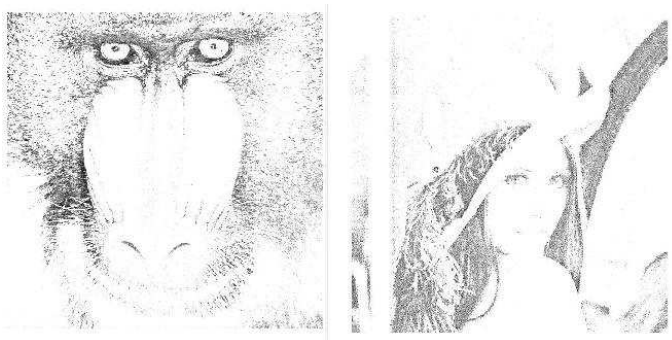



Figure 3.10: Evidence of image independency. The mean and variance of the quality of the un-marked images are reported for different values of $q$ and setting $Q = 40$dB. This provides clear evidence of the image independency of the algorithm, since the variance of WPSNR is at most 5%.

Finally we analyze the algorithm avoiding compression. We apply the combination of the two attacks: AWGN (C) and volumetric scaling (E). GA still looks for the order the two operators need to be performed. In Table 3.9, results for the Baboon and Lena images are reported. The elapsed time for obtaining such results is almost $85s$ per iteration. We stress the fact that the quality of the unmarked images is in this case substantially decreased, compared to previous experiments. This is mainly explained by the choice of the selected image processing operators: in fact, the analyzed algorithm is designed to resist against the chosen attacks. The significant degradation introduced to remove the mark is shown in Figure 3.11. In Figure 3.12, the robustness index $R(q)$ is plotted for different values of the quality threshold $Q$. In particular, it is shown that robustness is preserved even in the case of a low quality threshold ($Q = 30$ dB) and the combination of both attacks, as expected from [93].

| Quality parameters | | | Attack parameters | | |
|---|---|---|---|---|---|
| q | $M(q)$ | $\text{PSNR}_u$ | $\delta$ | $SF$ | $Order$ |
| Lena Image | | | | | |
| 47 | 12.7 | 6.6 | 3 | 2.97 | C-E |
| 53 | 13.2 | 7.1 | 3.4 | 2.54 | E-C |
| 59 | 14.6 | 8.6 | 4 | 0.26 | E-C |
| 65 | 20.2 | 13.9 | 4 | 0.59 | C-E |
| 71 | 52.6 | 42.7 | 2.6 | 0.98 | C-E |
| Baboon Image | | | | | |
| 47 | 13.4 | 7.2 | 0.1 | 2.97 | E-C |
| 53 | 13.9 | 7.8 | 2.1 | 2.54 | C-E |
| 59 | 14.6 | 8.5 | 0.6 | 0.26 | C-E |
| 65 | 20.2 | 15.5 | 4 | 0.59 | C-E |
| 71 | 55.1 | 45.1 | 1.8 | 0.98 | C-E |

Table 3.9: Robustness results of Li's algorithm under the combination of addition of WGN ($\delta$) and amplitude scaling attack ($SF$). Starting from the first column: the quality of the watermarked image $q$ and the quality of the un-marked image $M(q)$ and $\text{PSNR}_u$, measured with WPSNR and PSNR respectively. The parameterizations of the selected attacks are reported, together with the order they must be applied.



(a) (b)

Figure 3.11: Un-watermarked images for Li's algorithm under the attack of addition of WGN and amplitude scaling attack : in panel (a) Baboon, marked with $q = 47$dB, and unmarked with $M(q) = 13.4dB$. In panel (b) Lena, marked with $q = 47$dB, and unmarked with $M(q) = 12.7dB$.

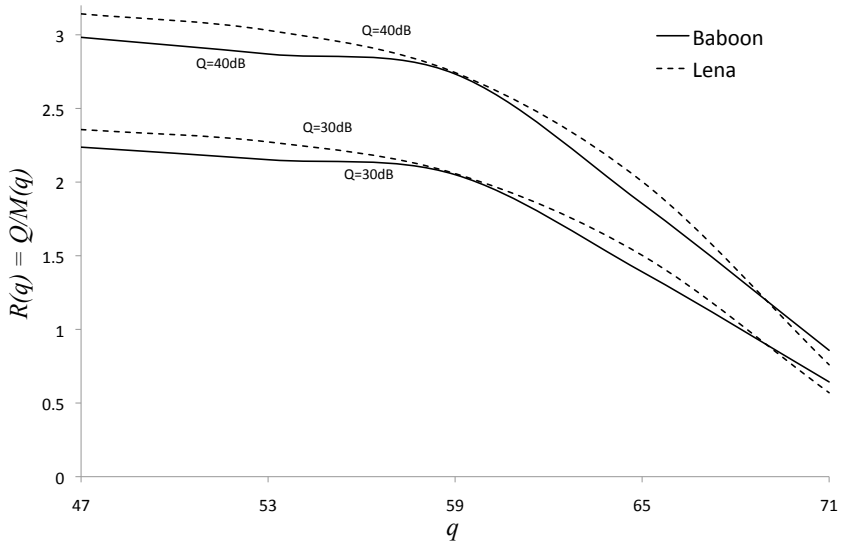



Figure 3.12: Performance plots for Li's algorithm, when setting a quality threshold $Q = 40$db and $Q = 30$dB under the combination of addition of WGN and amplitude scaling attack. The robustness metric, $R(q)$, reported on the vertical axis, indicates a robustness of the algorithm to the selected attacks, since $R(q) \geq 1$.

## 3.4 Discussion

We have presented an innovative benchmarking tool to evaluate the robustness of any digital watermarking technique considering the quality of the unmarked images in terms of perceived quality. Therefore, a new metric based on WPSNR is introduced. The goal is to remove the watermark from a content while maximizing perceptual quality. So, given a set of attacks, we look for a parameterization able to remove the watermark, optimizing the WPSNR of the unmarked image. This nonlinear optimization problem is supported by GA. The effectiveness of the present tool has been demonstrated by extensive simulations pointing out the weaknesses of two well-known methods. We also point out that with the proposed tool, it is possible to fairly compare two different watermarking algorithms performing the same kind of watermark recovery (namely, either both detection or both decoding).

# Chapter 4

# Reversible Data Hiding Based On Adaptive Prediction

*In this chapter we present a new near lossless reversible watermarking algorithm using adaptive prediction for embedding. The prediction is based on directional first-order differences of pixel intensities within a suitably selected neighborhood. The proposed scheme results to be computationally efficient and allows achieving high embedding capacity while preserving a high image quality. Extensive experimental results demonstrate the effectiveness of the proposed approach.*

## Acknowledgment

I would like to deeply thank prof. Karen Egiazarian, Tampere University of Technology (Finland) and prof. Marco Carli, University of RomaTre (Italy), for the inspiring collaboration and valuable comments.

Parts of this Chapter appear in:

- V. Conotter, G. Boato, M. Carli and K. Egiazarian "Near Lossless Reversible Data Hiding Based On Adaptive Prediction", *IEEE International Conference on Image Processing 2010*, Hong Kong, September 2010.
- V. Conotter, G. Boato, M. Carli and K. Egiazarian "High Capacity Reversible Data Hiding based on Histogram Shifting and Non-local Means" (invited paper), *IEEE International Conference on Linear and Non-Linear Approximation 2009*, Tuusula (Finland), August 2009.

## 4.1 Introduction

As outlined in Chapter 2, generally, in digital watermarking, it is desirable to embed as much information as possible (capacity), while guaranteeing the ability to preserve the hidden data after processing (robustness), and the fidelity of the watermarked data to the original one (imperceptibility). However, the relative importance of these factors strictly depends on the application. Some applications, such as medical, forensics, artwork, or military, require data integrity preservation, thus preferring imperceptibility over robustness.

Reversible watermarking [25] relies on the ability to perfectly recover the original content after the extraction of the embedded data. More in detail, this class of techniques allows a lossless or near-lossless restore of the host, while guaranteeing high quality of the watermarked image. Their effectiveness is measured by considering imperceptibility, payload capacity, and computational complexity. Robustness is barely achieved due to the challenging integrity requirement. Many reversible watermarking algorithms have been proposed in the literature over the past years. The main difference among them lies in the particular technique adopted for achieving reversibility: data compression [48][16], difference expansion [152][63], histogram shifting [115], and pixel prediction [58][39]. Prediction based methods have been originally introduced for compression since they guarantee reduced computational complexity with respect to the transform-domain based methods, by reducing coding, inter pixel, or psycho-visual redundancy via prediction and coding of the difference between the original and predicted pixel values [54].

In this work we proposed to extend prediction-based schemes, by designing a high-capacity reversible watermarking algorithm based on adaptive pixel prediction. It computes the to-be-predicted pixel by linearly combining its neighboring values. Different weights are suitably assigned to all values and computed based on directional first order differences between the predicted pixel and those in the surrounding area to adaptively exploit their similarity. Prediction errors are used to embed the watermark and a histogram-shift based approach is employed to control under/overflows issues. The algorithm is designed to achieve high embedding capacity while preserving high perceptual quality.

## 4.2 The proposed approach

We introduce a method that adaptively applies prediction in both embedding and detection phases and allows a lossless recover of the image in the detection process. In order to compute prediction errors used to embed the mark we present two methods: $(i)$ by linearly combining the to-be predicted pixel and its neighborhood, $(ii)$ by taking into account also some well known local predictors. Directional first order differences between the predicted pixel and those in the surrounding area are exploited to compute weights to be assigned to the values adaptively exploiting their similarity. We employ under/overflows control

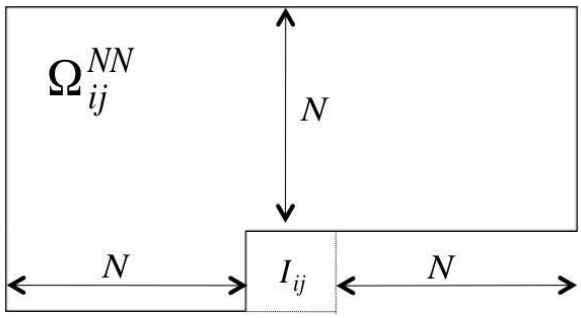



Figure 4.1: Neighborhood considered in the prediction.

based on a histogram shifting method to prevent values to exceed the range [0, 255]. Thus, the proposed algorithm consists of four main stages: prediction, embedding, under/overflows control and detection.

### 4.2.1 Prediction

Let $I$ be the host image with values $I_{ij}$ where $i = 1, \ldots, n$ and $j = 1, \ldots, m$ and $\hat{I}_{ij}$ the to-be-predicted pixel at position $(i, j)$. In order to compute $\hat{I}_{ij}$, we exploit an adaptive prediction. In particular, we consider pixel values within a region $\Omega_{ij}^{NN}$ around it (see Fig. 4.1) by defining:

$$V_{ij} = [I_{i-N,j-N:j+N}, \ldots, I_{i-1,j-N:j+N}, I_{i,j-N:j-1}]\,. \tag{4.1}$$

Therefore, the total number of elements used for prediction (i.e., the length of $V_{ij}$) is:

$$L_V = 2N(N+1). \tag{4.2}$$

It is worth noticing that $\Omega_{ij}^{NN}$ does not include pixels coming after (in a scan-raster order) the target pixel. In fact, to exactly retrieve the original data during the extraction phase, not all the neighbors surrounding the to-be-predicted pixel can be used for estimating its value since not available.

Once we have defined $V_{ij}$, we evaluate $\hat{I}_{ij}$ as a normalized linear combination of $V_{ij}$ elements:

$$\hat{I}_{ij} \;=\; \frac{\sum_{k=1}^{L_V} \alpha_{ij}(k) V_{ij}(k)}{\sum_{k=1}^{L_V} \alpha_{ij}(k)} \tag{4.3}$$

where $\alpha_{ij}(k)$ are weights defined by introducing another region $\Omega_{ij}^{TS}$ surrounding $I_{ij}$. Such neighboring pixels are exploited to evaluate the similarity in different directions and, thus, adapting the weights $\alpha_{ij}(k)$ accordingly. First, directional differences are computed as Euclidean distance between the pixel value $I_{ij}$ and all $V_{ij}$ elements:

$$g_{ij}(k) = |I_{ij} - V_{ij}(k)| \tag{4.4}$$

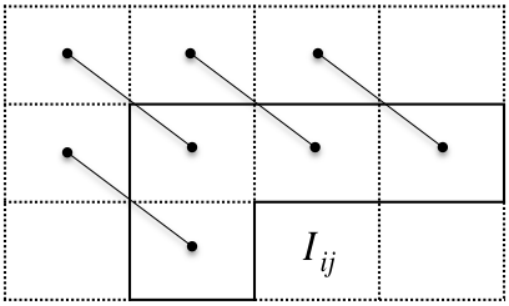



Figure 4.2: Pixel pairs involved in directional distances $g_{ts}(k)$ evaluation when $T = S = 1$ and $k = 2N^2 - 1$.

where $k = 1, \ldots, L_V$. Then such distances are integrated over $\Omega_{ij}^{TS}$ taking into consideration different pairs of values depending on $k$:

$$G_{ij}(k) = \sum_{I_{ts} \in \Omega_{ij}^{TS}} g_{ts}(k). \tag{4.5}$$

For sake of clarity, we report an example in Fig. 4.2 with $T = S = 1$ and $k = 2N^2 - 1$. Notice that the evaluation of $G_{ij}(k)$ in Equation (4.5) depends on $T$ and $S$ but also on $N$. Such values can be suitably tuned to achieve maximum embedding capacity of the algorithm (see Sec. 4.3). Finally, given $G_{ij}(k)$ and two constants $c$ and $h$ we define $\alpha_{ij}(k)$ as:

$$\alpha_{ij}(k) \quad = \quad c^{-\frac{G_{ij}(k)}{h}}. \tag{4.6}$$

Once the prediction process has been performed, the prediction error $E_{ij}$ can be computed as the difference of the original and predicted value:

$$E_{ij} = I_{ij} - \hat{I}_{ij}. \tag{4.7}$$

### 4.2.2 Improved Prediction

To further improve the prediction, besides the neighborhood we take into account also two well know local predictors: Median Edge Detector (MED) [58] and Gradient Adjusted Prediction (GAP) [39].
Median Edge Detector exploit three neighbors of the to-be predicted pixel, namely $I_{i,j-1}$, $I_{i-1,j-1}$ and $I_{i-1,j}$ for prediction, as shown in Fig. 4.3 straight line. MED uses raster scan order and applies edge rule to evaluate the predicted pixel, accord- ing to the following criteria:

$$\hat{I}_{ij} = \begin{cases} min(I_{i-1,j}, I_{i,j-1}) & \text{if } I_{i-1,j-1} \geqslant max(I_{i-1,j}, I_{i,j-1}) \\ max(I_{i-1,j}, I_{i,j-1}) & \text{if } I_{i-1,j-1} \leqslant min(I_{i-1,j}, I_{i,j-1}) \\ I_{i-1,j} + I_{i,j-1} - Ii - 1, j - 1 & \text{otherwise} \end{cases} \tag{4.8}$$

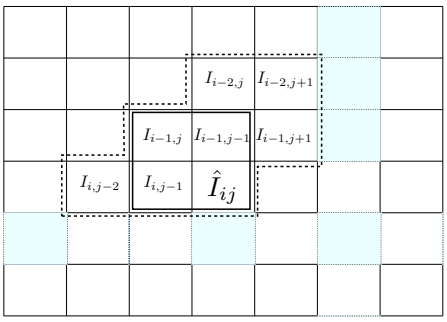



Figure 4.3: Neighborhood considered for prediction by MED (straight line) and GAP (dashed line).

Gradient Adjusted Prediction (GAP) considers seven neighbors of the to-be-predicted pixel (see Fig. Fig. 4.3 dashed line ) and compute its prediction $\hat{I}_{ij}$ as follows:

$$
\begin{aligned}
d_h &= |I_{i-1,j} - I_{i-2,j}| + |I_{i,j-1} - I_{i-1,j-1}| + |I_{i,j-1} - I_{i+1,j-1}| \\
d_v &= |I_{i-1,j} - I_{i-1,j-1}| + |I_{i,j-1} - I_{i,j-2}| + |I_{i+1,j-1} - I_{i+1,j-2}| \\
D &= d_v - d_h \\
\tilde{I}_{ij} &= (I_{i-1,j} + I_{i,j-1})/2 + (I_{i+1,j-1} + I_{i-1,j-1})/4 \\
\hat{I}_{ij} &= \begin{cases}
I_{i,j-1}) & \text{if } D < -80 \\
(\tilde{I}_{ij} + I_{i,j-1})/2 & \text{if } -80 \leqslant D < -32 \\
(3\tilde{I}_{ij} + I_{i,j-1})/4 & \text{if } -32 \leqslant D < -8 \\
\tilde{I}_{ij} & \text{if } -8 \leqslant D < 8 \\
(3\tilde{I}_{ij} + I_{i-1,j})/4 & \text{if } 8 \leqslant D < 32 \\
(\tilde{I}_{ij} + I_{i-1,j})/2 & \text{if } 32 \leqslant D < 80 \\
I_{i-1,j}) & \text{if } D \geqslant 80
\end{cases}
\end{aligned}
$$

As can be noticed, the prediction schemes based on MED and GAP exploit only local information obtained from a small neighborhood of the to-be-predicted pixel.

We add this two predictors to our vector $V_{ij}$ defined in Equation 4.1 built for prediction. Thus, we define a new vector $V'_{ij}$:

$$V'_{ij} = [V_{ij}, MED, GAP]\,. \tag{4.9}$$

where $MED$ and $GAP$ represent the predicted values in Equation 4.8 for MED and in Equation 4.9 for GAP, respectively.

The number of elements used for prediction is in this case:

$$L'_V = 2N^2 + 2N + 2. \tag{4.10}$$

For example, in case $N = 1$, we get $L'_V = 6$ and $V'_{ij} = [I_{i-1,j-1}, I_{i-1,j}, I_{i-1,j+1}, I_{i,j-1}, MED, GAP]$.

According to Equation (4.3) the vector $V'_{ij}$ is then used for prediction following the procedure described in the Sec. 4.2.1.

### 4.2.3 Embedding

Here we present our embedding technique which exploits prediction errors. Each $E_{ij}$, computed as in Equation (4.7), is controlled in the range $[-Q, Q)$, with $Q$ being an integer greater than zero, and is modified according to:

$$\hat{E}_{ij} = \begin{cases} E_{ij} + b_z \text{sign}(E_{ij})Q & \text{if } -Q \leq E_{ij} < Q \\ E_{ij} + b_z Q & \text{if } E_{ij} = 0 \\ E_{ij} + \text{sign}(E_{ij})Q & \text{otherwise} \end{cases} \tag{4.11}$$

where $b$ is the stream of binary entries $b_z$ to be embedded. Usually $b$ represents the watermark $w$ to be inserted, but in some cases we may embed besides $w$ an overhead information to deal with under/overflow problems (see Sec. 4.2.4).

The overall embedding scheme is:

$$I^w_{i,j} = \hat{I}_{i,j} + \hat{E}_{i,j} \tag{4.12}$$

where we denote the watermarked image as $I^w$.

### 4.2.4 Under/overflow control

During the embedding procedure some pixels may exceed the allowed intensity range [0, 255] leading to the so-called under/overflow problem. To prevent this, here we exploit a histogram shifting based technique, which allows the histogram of the host content to be narrowed. Thus, extreme values are modified and can be safely used for embedding. In particular, given an image with intensity values in the range [0, 255], the histogram shifting based process modifies the grayscale boundary pixels to be in the range [$Q$, 255-$Q$], where $Q$ is the parameter used in Equation (4.11).

Since the overall watermarking algorithm is designed to be reversible, the shift imposed to prevent under/overflow must be traceable. To this end, an overhead information $OI$ is embedded together with the watermark as follows:

$$T_{bit} = [OI, w] \tag{4.13}$$

where $T_{bit}$ is the total embedded bitstream ($b = T_{bit}$ in Equation (4.11)). The overhead information is defined as in Fig. 4.4. $L$ represents the size of the overhead, is represented

| L | Q | Left-side scan sequence | Right-side scan sequence |
|---|---|---|---|

Figure 4.4: Overhead information ($OI$).

| Overflow control | | | Underflow control | | |
|---|---|---|---|---|---|
| Original vale | New value | Left-side scan sequence | Original value | New value | Left-side scan sequence |
| 255 | 253 | 10 | 0 | 2 | 10 |
| 254 | 253 | 01 | 1 | 2 | 01 |
| 253 | 253 | 00 | 2 | 2 | 00 |

Table 4.1: Example of overflow and underflow control when $Q = 2$.

with a fixed number of bits, and is included in the $OI$ since required in the detection phase. The right and left side scan sequences keep trace of the modified pixels for under/overflow controls, respectively. In the case of overflow control, the host image is scanned in raster order and whenever a pixel value is greater than 255-$Q$ it is changed to 255-$Q$ and the shift amount is memorized in binary form in the left-side scan sequence. In the case the pixel value equals 255-$Q$ a zero is appended. A similar process is applied to control underflow by defining the right side scan sequence. In Table 4.1 an example of such procedures is presented, for $Q = 2$.

### 4.2.5 Detection

In order to have a reversible scheme, we need to fully recover both the watermark and the original image. To this aim the detection procedure has to be symmetric to the embedding. Given the watermarked image $I^w$, we reconstruct predicted values $\hat{I}^w_{ij}$ following the same procedure described in Sec. 4.2.1. The prediction error can thus be computed as:

$$E^w_{ij} = I^w_{ij} - \hat{I}^w_{ij}. \tag{4.14}$$

Thus, we can reconstruct the bitstream $\bar{b}$ as follows:

$$\begin{aligned}
&\bar{b}_z = 0 \text{ and } \bar{E}^w_{ij} = E^w_{ij} && \text{if } -Q \leq E^w_{ij} < Q \\
&\bar{b}_z = 1 \text{ and } \bar{E}^w_{ij} = E^w_{ij} - \text{sign}(E^w_{ij})Q && \\
&\qquad \text{if } -2Q \leq E^w_{ij} < -Q \text{ or } Q \leq E^w_{ij} < 2Q && \\
&\bar{E}^w_{ij} = E^w_{ij} - \text{sign}(E^w_{ij})Q && otherwise
\end{aligned}$$

where $\bar{E}^w_{ij}$ is the suitably modified prediction error used to recover $I^r$ by removing the watermark from $I^w$:

$$I^r_{ij} = \hat{I}^w_{ij} + \bar{E}^w_{ij}. \tag{4.15}$$

| | Prediction | | | Improved Prediction | | |
|---|---|---|---|---|---|---|
| | $T_{bit}$ | $OI$ | WPSNR | $T_{bit}$ | $OI$ | WPSNR |
| Lena | 60464 | 0 | 57.92 | 61688 | 0 | 5 8.44 |
| Barbara | 50822 | 0 | 58.12 | 51783 | 0 | 58.67 |
| Boat | 62341 | 0 | 56.78 | 62083 | 0 | 56.54 |
| Goldhill | 42553 | 0 | 57.43 | 43534 | 0 | 57.92 |
| Baboon | 19615 | 76 | 56.78 | 19688 | 76 | 56.6 |
| Peppers | 51451 | 251 | 57.48 | 50595 | 251 | 57.75 |
| Zelda | 60978 | 41 | 55.18 | 61832 | 41 | 55.51 |

Table 4.2: Results about capacity and perceived quality for both types of prediction with $T = N = S = 1$.

Since the implemented scheme is reversible we have $\bar{b} = w$ and $I^r = I$, if no under/overflow control is needed. Otherwise, we first extract $L$ (known length) and then identify $OI$ in order to invert the pixel shifts described in Sec. 4.2.4 and to extract the watermark $w$ from $\bar{b} = T_{bit}$ (see Equation (4.13)).

## 4.3 Results

We describe here a set of experimental tests carried out on grayscale images of size $[512 \times 512]$ pixels taken from the UWaterloo database [158] to verify the efficiency and effectiveness of the proposed method. First we consider the case when $N = T = S = 1$ which leads to minimum computational cost. We set $Q = 1$ and run simulations for both types of predictions (see Sec. 4.2.1 and 4.2.2). Results for 7 images are shown in Table 4.2, where we report the embedded payload $T_{bit}$, the overhead information $OI$, the Weighted PSNR (WPSNR), a modified version of the PSNR which takes into account the behavior of the human visual system [158]. Notice that PSNR values strictly depend on the binary sequence chosen as watermark and on average are $48.5 \pm 0.5$ dB. It can be noticed that in both cases we achieve high capacity while guaranteeing high quality of the watermarked image. Results are slightly better employing the improved prediction.

In order to further improve performances, it is possible to optimize the parameterization of the algorithm, obviously increasing the computational complexity. We exploit a brute force search for tuning parameters $N$, $T$ and $S$ and verify a significant increase in capacity without affecting quality. The parameterization for the algorithm is suitably tuned for each image since it depends on the image content. Also in this case we get PSNR values of $48.5 \pm 0.5$ dB for all the images.

[158] available at : http://links.uwaterloo.ca/Repository.html

| | Prediction | | | | | | Improved Prediction | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $T$ | $S$ | $T_{bit}$ | $OI$ | WPSNR | $N$ | $T$ | $S$ | $T_{bit}$ | $OI$ | WPSNR |
| Lena | 2 | 1 | 3 | 61682 | 0 | 57.59 | 2 | 1 | 4 | 63492 | 0 | 58.11 |
| Barbara | 3 | 2 | 3 | 54366 | 0 | 57.24 | 3 | 2 | 3 | 56028 | 0 | 57.77 |
| Boat | 3 | 2 | 5 | 64826 | 0 | 56.25 | 3 | 2 | 4 | 65992 | 0 | 56.3 |
| Goldhill | 1 | 1 | 3 | 42995 | 0 | 57.53 | 2 | 1 | 4 | 44275 | 0 | 57.7 |
| Baboon | 1 | 2 | 1 | 19639 | 76 | 56.69 | 1 | 3 | 3 | 20172 | 76 | 56.77 |
| Peppers | 3 | 1 | 2 | 51931 | 251 | 56.7 | 2 | 2 | 1 | 52294 | 251 | 57.22 |
| Zelda | 1 | 1 | 3 | 62349 | 41 | 55.1 | 2 | 1 | 5 | 63077 | 41 | 55.25 |

Table 4.3: Results for improved prediction about capacity and perceived quality with optimized parameters.

| | $\mu_+$ | $\mu_\times$ | $\mu_C$ | $\mu_S$ | Capacity Gain |
|---|---|---|---|---|---|
| [58] | 2 | 0 | 4 | 0 | 16.4% - 54.5% |
| [39] | 18 | 8 | 6 | 0 | 1.8% - 12.7% |
| Proposed | 26 | 4 | 0 | 8 | – |

Table 4.4: Performances in terms of prediction computational complexity and capacity gain percentage for the proposed method with $N = T = S = 1$ with respect to [58] and [39].

Finally, we analyze the performances of the proposed algorithm in comparison with two well known reversible watermarking techniques, [58] and [39], employing MED and GAP prediction respectively. Results for computational costs (number of additions $\mu_+$, multiplications $\mu_\times$, comparisons $\mu_C$ and shift operations $\mu_S$) are reported in Table 4.4, where the proposed adaptive prediction with $N = T = S = 1$ is compared with MED and GAP prediction. In Table 4.4 we also show the range of capacity gain percentage computed for the 7 images, comparing the proposed method with [58] and [39], respectively.

To fully compare the proposed method with MED and GAP, we compared the capacity achievements over three different kinds of embedding: *(a)* right shifted embedding where interval [0,1) is used for embedding, i.e., $E_{ij} \in [0, 1)$, *(b)* $Q = 1$ and *(c)* $Q = 2$ [39] . In Table 4.5 results, in terms of capacity and PSNR, for MED, GAP and our improved method are reported, testing all the three embedding methodologies. Of course, a higher capacity can be achieved as the interval used for embedding increases, therefore slightly loosing in quality. Note that in all cases and for all images we gain in capacity by exploiting more sophisticated predictors, i.e., GAP outperforms MED and the proposed technique outperforms GAP. In Fig. 4.5 we plot comparison results of Table 4.5 to visually enlighten the outperforming of our algorithm over MED and GAP.

| | | Embedding (a) | | | Embedding (b) | | | Embedding (c) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | [58] | [39] | Proposed | [58] | [39] | Proposed | [58] | [39] | Proposed |
| Lena | Payload | 26293 | 28971 | 31714 | 46670 | 57949 | 63492 | 75640 | 108540 | 116026 |
| | PSNR | 51.52 | 52.1 | 51.82 | 48.55 | 49.2 | 48.74 | 42.8 | 44.1 | 43.26 |
| Barbara | Payload | 20745 | 23816 | 27885 | 38098 | 47363 | 56028 | 62520 | 89296 | 104511 |
| | PSNR | 51.5 | 51.9 | 51.83 | 48.47 | 49 | 48.71 | 42.68 | 43.7 | 43.17 |
| Boat | Payload | 25296 | 28941 | 33349 | 45530 | 56850 | 65992 | 74659 | 105900 | 118960 |
| | PSNR | 51.55 | 52.1 | 51.1 | 48.54 | 49.2 | 48.8 | 42.8 | 44 | 43.32 |
| Goldhill | Payload | 18977 | 20837 | 22080 | 36036 | 41466 | 44275 | 62278 | 79110 | 83564 |
| | PSNR | 51.54 | 52 | 51.7 | 48.45 | 48.9 | 48.57 | 42.67 | 43.5 | 42.92 |
| Baboon | Payload | 8882 | 9629 | 10038 | 16858 | 19277 | 20172 | 32712 | 37678 | 39561 |
| | PSNR | 51.33 | 51.5 | 51.37 | 48.29 | 48.5 | 48.32 | 42.4 | 42.7 | 42.47 |
| Peppers | Payload | 18429 | 22841 | 25808 | 33294 | 45669 | 52294 | 55810 | 86373 | 97659 |
| | PSNR | 51.47 | 51.9 | 51.74 | 48.43 | 49 | 48.64 | 42.63 | 43.6 | 43.08 |
| Zelda | Payload | 23909 | 28906 | 31432 | 44438 | 57908 | 63077 | 74508 | 110798 | 119711 |
| | PSNR | 51.56 | 52.2 | 51.7 | 48.53 | 49.2 | 48.74 | 42.8 | 44.2 | 43.3 |

Table 4.5: Results comparison for three different embedding methodologies.

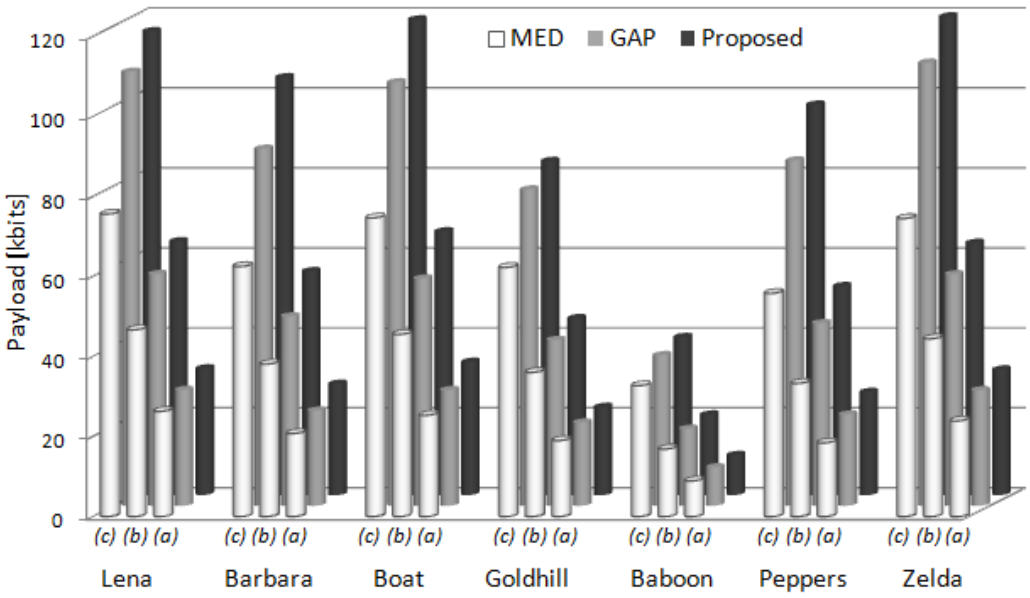



Figure 4.5: Performances comparison in terms of capacity among MED (white), GAP (gray) and proposed predictor (black) for the three types of embedding *(a)*, *(b)* and *(c)*.

## 4.4 Discussion

We have presented a new reversible watermarking technique, where embedding is based on adaptive prediction errors. Neighboring pixels of the to-be-predicted value are combined and suitably weighted depending on their directional distances, assigning to similar pixels a greater role in the computation. The algorithm achieves high performances in terms of capacity, complexity and quality of the watermarked image.
Future work would deal with the investigation of different prediction schemes and the optimization of the similarity matching algorithm, in order to improve the performances of the algorithm while keeping the computational complexity low.

# Chapter 5

# Passive Digital Forensics

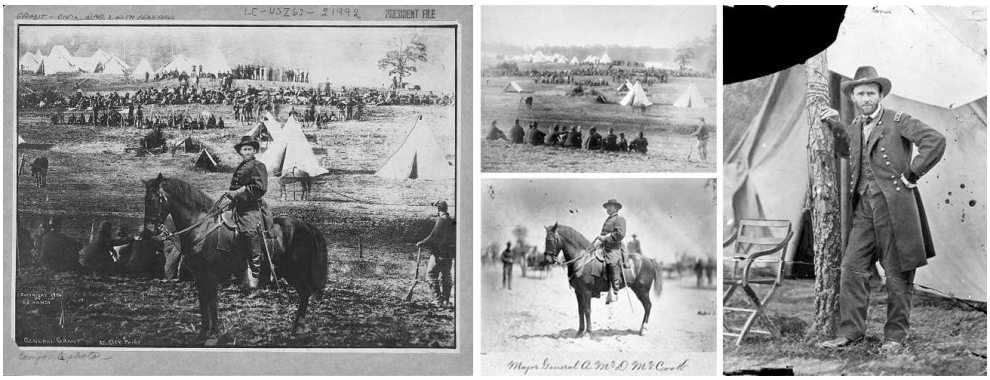
*In this chapter we review existing tools for digital passive forensics, mostly focusing on forgery detection. In contrast to statistical techniques, geometric forensic techniques are largely insensitive to resolution, post-processing, compression and re-compression. As such, we mainly carried on our research on geometric-based methods, analyzing the projection geometry of the image formation. Our contributions both on image and video forgery detection are finally introduced.*

Thanks to the wide spread of the Internet as a massive communication mean coupled with the diffusion of low-cost and high-resolution cameras, we are nowadays overwhelmed in our daily-life with digital multimedia contents, such as images and videos. They are easily spread out via web-based content-sharing tools (social networks, Youtube, Picasa, Flickr) and widely employed in several fields such as news, sport, information reporting, scientific publication, politics campaigns and forensic courts.

From a traditional point of view, we assume a photograph or a video to be a trusty and close representation of reality. But in today's digital age, this assumption is undermined by the simplicity of malleability and manipulation of digital multimedia content, due to the wide spread of low-cost and high-performance computers, high-resolution digital cameras and sophisticated photo-editing and computer graphics software tools (e.g. Photoshop). Even a non-expert user can easily manipulate and alter multimedia contents, change the information they represent, without leaving any obvious traces of the occurred tampering. As a consequence, we can no longer take the fidelity and authenticity of images and videos for granted, especially when we consider scenarios such as forensic and criminal investigation, surveillance systems, medical imaging and journalism. Moreover, altered multimedia data may influence people opinions and even alter their attitudes in response to the represented event [136, 111].

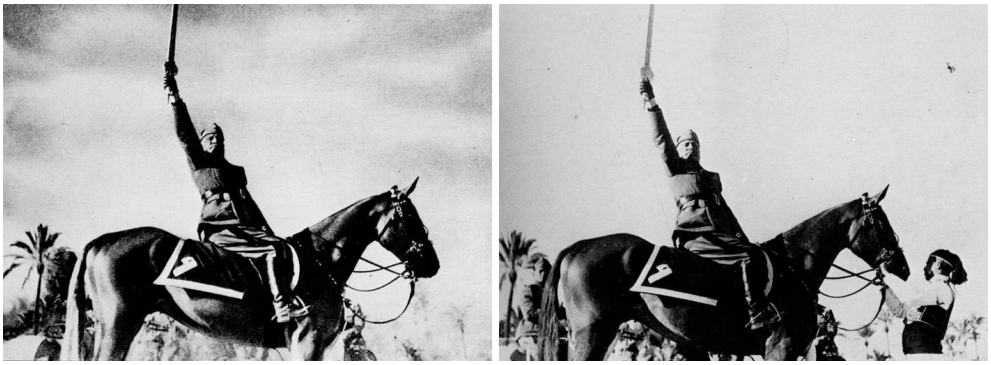
Multimedia content fakery has a long history, probably starting right after Joseph Nicephore Nipce produced the first known photograph in 1825.

All the following reported forged photographs are courtesy of the Dartmouth Image Science Group [42]



(a) (b)

Figure 5.1: Circa 1864: the print (a) portraying General Ulysses S. Grant resulted to be a composite of three other pictures (b), taken in different moments and settings.



(a) (b)

Figure 5.2: Circa 1942: the portrait (a) of Mussolini is the result of altering the original picture by deleting the horse handler (b).

Some of the most notorious examples of early photographic tampering involved military generals and politicians. Figure 5.1, taken circa in 1864, portrays General Ulysses S. Grant in front of his troops at City Point, Virginia, during the American Civil War. Researchers at the Library of Congress revealed that this print is a composite of three separate pictures: the head is taken from another portrait of Grant, the horse and body are those of Major General Alexander M. McCook and the background is taken from a photograph of Confederate prisoners captured at the battle of Fisher's Hill, VA.

Another example comes from the Second World War, when Mussolini altered one of his portraits by removing the horse handler, pretending to look more heroic and valorous. (see Figure 5.2).


(a) (b)

Figure 5.3: 1989: the cover of a TV magazine picturing the American TV host Oprah Winfrey (a) was created by splicing her head and Ann-Margret's body (b).

At that time, a high degree of technical expertise and specialized equipment was required to tamper a photograph. Nowadays, modern softwares have made the forgery of photographs and, more recently, of videos, easier to be done and much harder to be uncovered than ever before.

Starting from the 1980's we witness a significant increase in image tampering, especially in magazine covers in order to appear more appealing and attractive to customers. In Figure 5.3 the American television host Oprah Winfrey is pictured on the cover of the magazine *TV Guide*. It was uncovered that this was created by splicing her head and the body of Ann-Margret, without the permission of any of the two. The forgery was discovered by the fashion designer of Ann-Margret, who recognized the dress.

In the 90's magazines started to alter their covers also to alter people attitudes in response to the represented event. A popular example is shown in Figure 5.4, where *Time* magazine published a photograph of OJ Simpson, right after he was arrested for murder. His image on the cover appeared to be be darker and menacing compared to the original photograph, published by the magazine *Newsweek*. A couple of years later, the *Los Angeles Times* published a digital composite of a British soldier in Basra, gesturing to Iraqi civilians urging them to seek cover, shortly after the U.S. led invasion of Iraq (see Figure 5.5). When the composite was uncovered by the editors, the photographer for the newspaper was fired. He said he spliced the images in order to "improve" the image.


(a) (b)

Figure 5.4: 1994: the image of murderer OJ Simspon which appeared on the cover of the magazine *Time* (a) was altered in brightness and color to make the subject look more menacing. The original was published by the magazine *Newsweek* (b).

(a) (b)

Figure 5.5: 2003: the cover of the *Los Angeles Times* (a) picturing a soldier during the Iraq war in front of civilians has been revealed to be a composite of two different pictures (b).

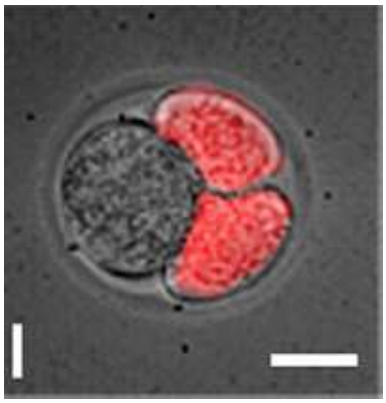


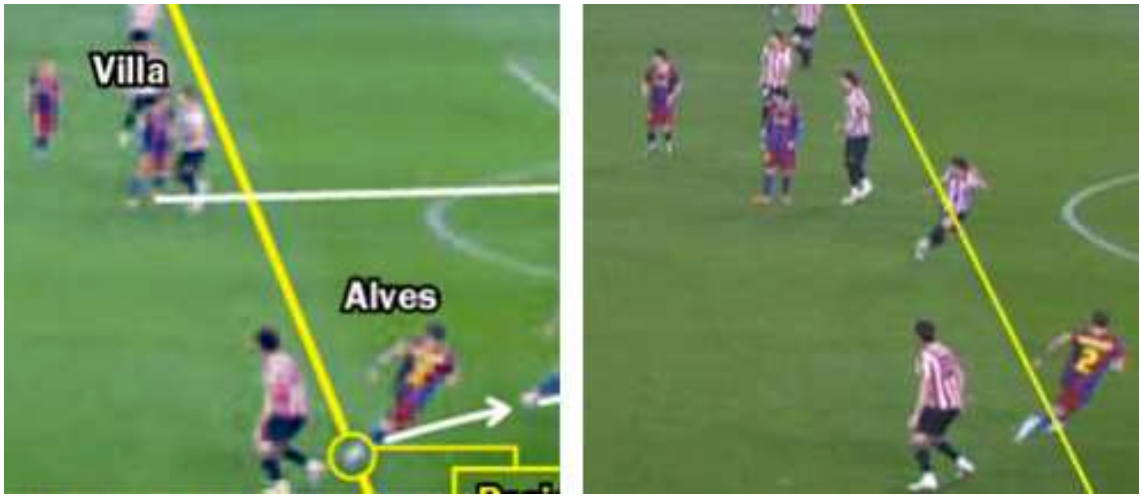
Figure 5.6: July 2007: the investigating university committee revealed image results in a scientific publication to be doctored and forced the authors to retract the paper.

Also the scientific community has been experiencing the ease of image manipulation in journal publications. The impact of digital photography processing software in science was addressed in an article in the journal *Nature* published in April 2005. It reported that, in 1990, 2.5% of contentions examined by the U.S. Office of Research Integrity, which monitors scientific misconduct, involved suspect scientific images. By 2001, the trend was impressively increasing reaching nearly 26%.
One well-known episode happened in 2007, when Professor R. Michael Roberts and co-authors from Missouri University published the paper "Cdx2 Gene Expression and Trophectoderm Lineage Specification in Mouse Embryos" on the journal *Science* pretending to break through the conventional wisdom demonstrating that the first two cells of mouse embryos possess markers that indicate from a very early stage whether they will grow into a fetus or placenta. But the images in the paper (Figure 5.6) resulted to be doctored and the authors had to retract the paper when the investigating university committee found that Roberts and his post-doc deliberately altered images of the embryos. The latter abruptly resigned his position and moved with no further information, while professor Roberts behavior was judged by the committee as wrong, but *"since he addressed that in the letter he sent to Science, we had no reason to suspect anything other than that he had been tricked"*.

Recently, we are literally overwhelmed by retouched images, especially in magazines and advertisements. A well known case concerns the cosmetic industry Olay's advertisement shown in Figure 5.7. The ad read: "Olay is my secret to brighter-looking eyes.. reduces the look of wrinkles and dark circles for brighter, younger-looking eyes". It turned out that the model Twiggy was heavily retouched in order to look younger, but this post-production retouching was judged misleading for costumers by the Advertising Standards Authority (ASA). It sentenced that the ad gave consumers a *"misleading impression of the effect the product could achieve"* and was therefore banned in the United Kingdom.


(a) (b)

Figure 5.7: December 2009: The model in the magazine ad for an Olay beauty product commercial (a) has been heavily retouched (b). The Advertising Standards Authority (ASA) banned it because judged to be misleading.




(a) (b)

Figure 5.8: February 2011: a Spanish sport magazine pretended to give evidence of an offside violation in the match between Athletic Bilbao and Barcelona with the picture (a), but a defender was digitally removed from the original (b), thus no violation occurred.

Figure 5.9: September 2007: for political reasons, Mikhail Delyagin has been removed from the video sequence of a popular TV show, but the forgery was not accurate and the leg and hand remained visible in some frames.

A very recent example of image fakery was registered in February 2011, when a Spanish sport newspaper published an altered photo of the match between Athletic Bilbao and Barcelona, pretending to demonstrate an offside violation (see Figure 5.8). However, the original frame shows that a defender had been digitally deleted from the photo and thus no violation occurred. The newspaper did apologize for this, claiming it was due to an info-graphics error.

Even if it is clearly more difficult and time consuming to manipulate and forge a video than tampering an image, the wide spread of video-sharing websites and the availability of video processing softwares make this task easier nowadays. An example of malicious tampering with video is shown in Figure 5.9. After Mikhail Delyagin criticized the Russian Prime Minister Vladimir Putin, he was removed from a video sequence of the TV shows *"The people want to know"*. But the leg and hand remained visible to the right of the man holding the microphone.

As shown by all the presented examples, the growth in image and video tampering is having a significant impact in our daily life and in our society. The reliability of digital content cannot be taken for granted and has been seriously questioned. Does the content truly represent a shot of reality? Where does this image come from? What is its history? To answer such types of questions, passive digital forensics has recently attracted the attention of the scientific community, as the increasing number of publications in this field confirms. It is said to be passive, or blind, since it can operate where no prior information about the content is available or no integrity protection mechanisms (e.g. digital watermarking) have been previously applied.

Passive digital forensics is tightly connected with many different science fields, such as computer science, signal processing and criminal justice. As commonly accepted, we can identify the following research areas of passive digital forensics [140]:

- **Source identification** : the aim is to identify the device which captured the content, by exploiting traces left by the different steps taken during the image acquisition process. The basic idea comes from classical forensic science, where bullet analysis is carried on based on the distinct markings introduced onto it when fired. Such markings are distinctive and unique for every specific gun and can be therefore used to link the fired bullet with the weapon which shot it. Similarly, when shooting an image, a specific and unique fingerprint is introduced into the content, depending on the device which took it. In particular, a noise-like pattern is overlaid onto the image when captured, which may take the form of image artifacts, distortions or statistical properties of the data. Such noise is invisible to the human eye, but it can be analyzed to successfully contribute to identification.

  A plethora of remarkable works has been presented in the literature for source identification, exploiting distortions introduced by lens distortion (aberration) [23] or demosaicing artifacts [11, 15], while others rely on the sensor imperfections, such as defective pixels [50], fixed pattern noise [35] and photo-response non-uniformity noise (PRNU) [47, 20] . Similar concepts may be adapted to videos. For example, dark current pixels can be exploited to link a video to the device that generated it [85]. The photo-response non-uniformity can be used as a unique fingerprint to identify the source camcoder [19, 109]. In [61] authors propose to analyze the correlation of noise residual on a block level, computing it or each pair of temporally adjacent blocks. Low-resolution issues related to the pattern noise are analyzed in [155], successfully linking youtube videos to their source.

- **Discrimination between synthetic and real images**: the purpose is to differentiate between real and computer generated images, given that the increasing photorealism of imagery created by sophisticated 3D graphic tools make this task challenging at only visual inspection [113]. This goal is mainly achieved via machine learning algorithms suitably modified to classify natural and artificial images.

  A first approach has been made by Lyu in [100], where higher order statistics of wavelet transform coefficients are used to train a classifier and are shown to be effective in discriminating natural and computer generated images. After this, several techniques have been developed, among which, physics-based [114] and features-based methods [138]. Following the core idea of source identification techniques, methods have been developed based on the assumption that the generative process for computer graphics is substantially different from the process real images are undertaken during acquisition. Therefore, techniques based on demosaicing and chromatic aberration have been proposed [34], as long as pattern noise based approaches [29].

- **Forgery detection** : the goal is to authenticate digital contents, including images or videos, based on the assumption that forgeries may not leave any clue of their occurrence but they alter the underlying statistics of the content [44, 103].

  Several techniques have been proposed in the literature and we will review them later in this chapter.

It is worth mentioning an emerging research areas in digital forensics, called *Computer forensics*, whose principal goal is to recovery data from corrupted or physically damaged devices for legal purposes [117]. It requires both hardware and software approaches, since once the data has been recovered from the hardware it is necessary to re-organize it via a software mean. Although included in the forensics area of research, computer forensics substantially differ in the approach compared to the above mentioned classes: in [12] similarities and differences are pointed out.

For more detailed literature on the forensic areas of research we address the reader to [135, 31, 156, 140, 132] and the references therein. For a complete and constantly updated on-line bibliography in the field, please refer to [41, 91].

## 5.1 Forgery detection

Following we review the state-of-the-art of forgery detection, following the classification presented in [44]. In particular, forensic tools designed for image authentication and forgery detection can be grouped into five categories: *(i)* Pixel-based, *(ii)* Format-based, *(iii)* Camera-based, *(iv)* Physically-based, *(v)* Geometric-based. . We are undoubtedly but involuntarily omitting some representative contributions in the field of forgery detection, but our aim is to give the reader a global overview of the research areas in order to understand the motivations which lead us to develop our contributions, enlightening their innovative aspects.

### 5.1.1 Pixel-based

The basic assumption of this class of techniques is that any form of manipulation, if applied properly, is not visually detectable but alters specific statistics at a pixel level. For example, in [7] statistical regularities in natural images that remain independent of the image content are exploited to authenticate images.
Depending on the occurred forgery, pixel-level correlations can be analyzed.
Cloning (or "copy and paste") is probably one of the easiest form of forgeries, usually performed in order to conceal an object in the scene by covering it with other parts of the image itself [9]. Although visually challenging to be disclosed, such kind of tampering can be detected by looking for statistically similar parts within the content. However, it results computationally unaffordable to perform a brute search all over the image. To reduce computational complexity and speed up the process, efficient algorithms have been proposed, based on DCT transform [146], DWT transform [76], PCA analysis [124], Fourier-Mellin transform [10] and features matching [118]. Similar concepts are applied

to video duplication detection [165].
Resampling is a process required when resizing, rotating or stretching an image, operations that are likely to happen when creating a fake image. It introduces some specific correlations in neighboring pixels, that can be analyzed as evidence of manipulation. All produced work in such a direction show that interpolated signals and their derivatives contain specific detectable periodic properties [102, 78, 94].
When creating a composite, two or more images are spliced together. Such operation has been demonstrated to alter higher order Fourier statistics, alteration that can be used as evidence of tampering [112]. Recently, in [162] the co-occurrence matrix of thresholded edge image of image chroma is analyzed.
Removing a moving object by video inpainting generally introduces ghost shadow artifacts, which can be accurately detected and their inconsistencies across the sequence can be taken as evidence of tampering [178].

### 5.1.2 Format-based

This class of techniques aims at disclosing statistical correlations introduced by compression schemes. The JPEG compression is well known to be a lossy scheme, i.e. some information is lost during the process. In particular, a quantization step taken on the DCT coefficient is mainly responsible for such loss. The full quantization is based on a table of 192 values, associated with each frequency on a $8 \times 8$ block-basis and may vary depending on the quality setting. In [40] it has been demonstrated that such table can be estimated and extracted from the content. Inconsistencies of it all over the image, or video, can be taken as evidence of tampering [82, 21]. Moreover, it has been shown that JPEG compressing the data twice, some specific artifacts are introduced in an image [99, 128], as well as in a video [168]. Lately, also JPEG2000 traces have been studied as proof of alteration [127] and also for double compression [179]. Block artifacts introduced by a JPEG compression at the border of neighboring pixels are studied in [92] for forensics purposes supposing that manipulations are likely to alter such artifacts. Assuming that, when creating a composite, it is unlikely to match the same level of quantization of the two spliced parts, [43] analyzes the quantization coefficients to prove tampering.
Although that traces of JPEG quantization are not necessarily proof of malicious tampering, the class of technique above described represents an helpful mean to gather information about the history of the content.

### 5.1.3 Camera-based

Similarly to source identification task, camera-based techniques for tampering detection are based on the analysis of traces left by the various stages of the imaging process. These artifacts are inherent to camera manufacturing processes and inconsistencies can be taken as evidence of tampering.
Chromatic aberration is an artifact due to a spatial shift in the location where the light

with different wavelength hits the sensor. In [68] local aberration is evaluated and inconsistencies of it with respect to the global image are taken as evidence of tampering.
Most of digital cameras are equipped with spherical lens which cause a radial distortions on images. When two image are spliced together, it is unlikely to match such distortion among spliced parts and such ireegularities may be taken as evidence of tampering [22].
Moreover, generally digital cameras are provided with a single sensor and color are captured with a color filter array (CFA). For each pixel, only one color sample is recorded and the missing ones are obtained by interpolation. As a consequence, specific correlations are introduced, which are unlikely to survive when tampering occurs [125].
Based on the linear relationship between the amount of light measured by the sensor and the corresponding pixel values (camera response), in [62] a forensic tool for detecting inconsistencies in such mapping is presented.
Correlations introduced by the camera or software algorithms for de-interlaced video can be exploited for video authentication, by analyzing their inconsistencies across the sequence [164].

### 5.1.4 Physics-based

A recent research area in compositing detection considers inconsistencies in lighting. As a matter of fact, even with the more sophisticated editing tools, it is difficult to match the lighting effects on each part of the composite. Therefore, differences in lighting across the entire image can be taken as evidence of forgery.
Pioneering contributions in this field have been proposed by Johnson et al. in [67], where 2-D surfaces normals are estimated to analyze light directions of different object in the scene, whose consistency can be used as proof of compositing. Such idea has been further expanded to 3-D models [72], exploiting light reflection in human eyes.
However, these work assume a simplified lighting mode, but a scene lighting can be complex. Recently, it has been demonstrated that the 3D lighting environment can be approximated with a low-dimensional model [71, 75] .

### 5.1.5 Geometric-based

Typically images may undergo a variety of post-processing and re-compression, which may impair the effectiveness of traditional techniques for forgery detection. In contrast to statistical techniques, geometric-based forensic techniques have been proposed, which exploit measurements of objects in the world and their position relative to the camera analyzing the projection geometry. Their major advantage over techniques based on low-level image statistics is that the modeling and estimation of geometry is less sensitive to resolution and compression that can easily confound the statistical analysis of images and videos.
Several techniques have been proposed in the literature, each exploiting the principles of image formation and projective geometry.

In [70] the authors show how photo compositing can influence the position of the projection of the camera center into the image plane. The so-called principal point is one of the internal parameters of the $3 \times 3$ planar projective transformation $\mathbf{H}$ (called homography) which relates world points $\mathbf{x}$ and image coordinates $\tilde{\mathbf{x}}$ ( $\tilde{\mathbf{x}} = \mathbf{H}\mathbf{x}$ ). In particular, the homography can be decomposed in camera's intrinsic (focal length, principal point and skew parameter) and extrinsic (camera motion) parameters. Under certain conditions, the homography can be factored and the parameters estimated, including the principal point [55]. The algorithm in [70] is based on the assumption that generally the principal point is located near the center of the image and translations in the image plane correspond to an equivalent shift of the principal point across the image. Exploiting the known geometry of a pair of eyes, inconsistencies in the principal points across persons in an image are used as evidence of tampering.
Similarly, Zhang et al. in [180] describe a technique for detecting image composites by enforcing two geometrical constraints on the homography. The approach can detect fake regions efficiently on a pair of images at the same scene, but requires two images correlated with $\mathbf{H}$ (planar homography) or $\mathbf{F}$ (fundamental matrix) constraints.
When making composites, also the matching of shadow is a challenging task. The imaged shadow can be modeled by a planar homography. By imposing geometric constraints on it from a single image, it is possible to detect digital forgeries [181].

In a forensic setting it is very important to be able to retrieve objects measurements when only a picture of them is available. In [160] the authors show that the camera matrix and some available scene constraints can be used to retrieve geometrical constraints of the scene (e.g. height of an object with respect to the reference plane).
Also rectification of planar surfaces under perspective projection can be used to the same aim. In [69] the authors review several techniques for image rectification of planar surfaces. The knowledge of polygons of known shape, two or more vanishing points, and two or more coplanar circles allow the image to world transformation of the planar surface to be computed. The removal of such projective distortion can be applied straightforwardly and metric measurements can me made on the planar surface. Moreover, in the case of spliced images, it is unlikely that all the parts obey the same rule of perspective projection. By rectifying the entire image, inconsistencies can be revealed and the forgery detected.

A geometric-based approach has been proposed also for video authentication [167]. The technique is based on the fact that for re-projected videos the skew parameter, one of the intrinsec parameters of the homography, may be non-zero, in contrast with the expected parameter of an authentic video, which should be zero.

## 5.2 Innovative Contributions

As we have seen, statistical techniques for tampering detection may suffer from resolution, post-processing, compression, and re-compression. Geometric-based techniques offer a valid support being less sensitive to such issues.

Based on this consideration, we developed a couple of innovative forensic techniques which exploit the geometric principles of perspective projection:

- **Geometric-based Image Forensics for detecting manipulation of text**:
  This forensic technique aims at detecting if text on signs or billboards has been forged. Adding or changing text in an image is relatively an easy task, even in a visually compelling way, so that the modification are hard to be detected at simple visual inspection. This technique explicitly models the projection from the sign in the world to the image to determine if this projection satisfies the expected planar perspective mapping. It is indeed unlikely that inserted text precisely satisfy such rule, therefore deviations from the model can be taken as evidence of tampering.
  To the best of our knowledge, this is the first forensic approach to authenticate text in images.
  Chapter 6 presents a detailed description of such forensic framework.

- **Geometric-based Video Forensics for ballistic motion authentication**
  Videos purportedly showing sensational basketball and soccer shots and death-defying acrobatics have become popular on video sharing websites. Some are real, but many are fake. We proposed a geometric-based forensic technique for authenticating such videos determining if a projectile (e.g., a thrown ball) in a video is authentic. This technique explicitly models the 3-D ballistic motion and 2-D projection of the trajectory of a projectile. Deviations from this model are used as evidence of fakery.
  The analysis of projectiles in video has not previously been considered in the forensic community. It has, however, been addressed in the robotics and computer vision communities. Most notably, in the development of ball-catching robots [116, 134, 131] and in the analysis of sporting events (*e.g.*, soccer, football, basketball) [77, 130, 4, 97]. Unlike the approach we developed, these techniques either require a stereo camera pair or require additional information about the scene geometry.
  A more closely related approach considers the estimation of a projectile's trajectory from a single camera [133]. That approach, however, requires a calibrated camera and is only applicable when the camera is stationary. In contrast, the approach described here does not require a calibrated camera and is applicable to a stationary or moving camera (with unknown camera motion).
  Our technique makes minimal assumptions about the nature of the projectile trajectory and camera. The computational requirements are simple, intuitive, and computationally efficient. Experimental tests on videos of our own creation and on videos obtained from video sharing websites verify the efficacy of this analysis.
  Chapter 7 presents a detailed description of the proposed video authentication tool.

# Chapter 6

# Detecting Photo Manipulation on Signs and Billboards

*The manipulation of text on a sign or billboard is relatively easy to do in a way that is perceptually convincing. When text is on a planar surface and imaged under perspective projection, the text undergoes a specific distortion. When text is manipulated, it is unlikely to precisely satisfy this geometric mapping. In this Chapter we describe a technique for detecting if text in an image obeys the expected perspective projection, deviations from which are used as evidence of tampering. Tests on forged images confirm the efficiency of the proposed approach.*

## Acknowledgment

I would like to thank prof. Hany Farid for hosting this research within the Image Science Group at Dartmouth College (USA).

Part of this Chapter appears in:

- V. Conotter, G. Boato, H. Farid "Detecting Photo Manipulation on Signs and Billboards", *IEEE International Conference on Image Processing 2010*, Hong Kong, September 2010.


Figure 6.1: Examples of signs and billboards.

## 6.1 Introduction

In our daily life we are overwhelmed by signs and billboards. They are everywhere and some of them may appear funny and weird. In Figure 6.1 some examples of odd signs are shown: one may hardly believe they are real, unless she saw them in person or took the picture herself. Thus their authenticity is hard to be established at first glance, thus requiring forensics tools able to determine if a text in a given image is authentic of it has been inserted and manipulated.

Throughout photograph history [42], we have many examples of text manipulation in photographs. In 2004 a controversial photo of a U.S. Marine posing with two Iraqi children while purportedly holding an inappropriate sign was widely circulated on the Internet. The Marine claimed that the image was manipulated, and that the sign originally read "Welcome Marines". The photo created a significant enough controversy that a military inquiry was launched. The investigation, however, was inconclusive and the authenticity of the image was never determined.

In April 2005 another well known doctored photograph appeared on the web, in which text was altered in order to lead a different meaning (see Figure 6.2 (b) ). It shows the British politicians Ed Matts, conservative candidate for Dorset South, and Ann Widdecombe, conservative candidate for Maidstone and the Weald, holding a pair of signs that together read "controlled immigration – not chaos and inhumanity". This picture appeared as part of Matts' election literature. The original photograph (see Figure 6.2 (a)), however, shows the two politicians during a campaign to support asylum for a Malawian


(a) (b)

Figure 6.2: A well known example of manipulated photograph, where text has been forged in order to lead a different meaning. Shown in panel (a) the original image and in panel (b) a doctored version created by manually distorting text onto a planar surface.

family in order to allow them to stay in Britain. Widdecombe declared she was "happy to be associated with either message".

The adding or changing of text in an image is relatively easy to do in a way that is perceptually convincing, thanks to the availability of powerful personal computers and sophisticated photo-editing software. Figure 6.3 shows an example, of our creation, of forged text, showing the steps usually, but somehow, simplified, taken to create a text forgery. Given the original image intended to be forged (Figure 6.3 (a)), we first delete the undesired text and we then replace it with another text string, roughly matching color, font and size (Figure 6.3 (b)). In order to render the forgery visually compelling, we need to apply to the text string on the planar plane a distortion that visually simulates the regular perspective projection the image has been undergone (Figure 6.3 (c)). Such operation can be made using popular image editing softwares. In particular, we used image transformations available in Photoshop in order to create the presented forgery. Although this image looks visually compelling, the introduced distortion does not satisfy a proper perspective projection, but pretends to visually have the same impact. When inserting text into an image, it is likely that the precise rules of perspective projection will be violated, and these violations will not be perceptually obvious [45].

Following, we describe a new forensic technique for determining if typed text on a sign or billboard obeys the rules of perspective projection. This method explicitly identifies the projection of text on a planar surface and detects deviations from this model. We consider the case when the font style of the text in question is known and when it is unknown. In the context of geometric-based forensic techniques [70, 167, 180], this is a first approach dealing with detecting manipulated text.


(a) (b) (c)

Figure 6.3: The process performed to manipulate text in a photograph is shown. The text that needs to be forged in the original image (panel (a)) is first deleted and then replaced with another text string, roughly matching color, font and size (panel (b)). The final doctored photo (panel (c)) is created by manually distorting the text onto the planar surface, pretending to visually simulate a perspective projection, but indeed not following its rule.

## 6.2 The proposed approach

In this section we describe the forensics technique we developed, by firstly reminding some basic concept about homography and how it can be evaluated.

With regard to notation, scalars are denoted with lower-case italic letters $x$, vectors are denoted with lower-case bold-face letters $\mathbf{x}$, and matrices are denoted with upper-case bold-face letters $H$. Image coordinates are denoted as $\tilde{\mathbf{x}}$, and world and coordinates are denoted as $\mathbf{x}$. Superscripts are used to denote the components of a vector $\mathbf{x} = \begin{bmatrix} v^1 & v^2 & v^3 \end{bmatrix}$ or the rows of a matrix: $\mathbf{h}^k$ is the $k^{th}$ row of matrix $H$.

### 6.2.1 Planar Homography

The process of image formation involves the projection of three-dimensional points (the world) to two-dimensional points (the image). In general, this process can be described by the central projection of points in space onto a plane. Specifically, this model is in accord with a simple pinhole camera model, in which a ray of light from a 3D world point $\mathbf{x}$ passes through the lens of a camera (centre of projection $\mathbf{c}$) and produce an image on the point $\tilde{\mathbf{x}}$ by intersecting a specific plane in space, called image plane. The perspective mapping between points in 3-D world coordinates to 2-D image coordinates can be expressed by the projective imaging equation $\tilde{\mathbf{x}} = \mathbf{P}\mathbf{x}$, where the $3 \times 4$ matrix $\mathbf{P}$ embodies the projective

transform, the vector $\mathbf{x}$ is a 3-D world point in homogeneous coordinates, and the vector $\tilde{\mathbf{x}}$ is a 2-D image point also in homogeneous coordinates.

We consider a special case of this geometric transform where all of the world points $\mathbf{x}$ lie on a single plane and $\mathbf{P}$ reduces to a $3 \times 3$ planar projective transform $\mathbf{H}$, known as a homography:

$$\tilde{\mathbf{x}} = \mathbf{H}\mathbf{x}, \tag{6.1}$$

where the world $\mathbf{x}$ and image points $\tilde{\mathbf{x}}$ are represented by homogeneous coordinates.

Reformulating Eq. (6.1) as a cross product yields:

$$\begin{pmatrix} \tilde{x}^1 \\ \tilde{x}^2 \\ \tilde{x}^3 \end{pmatrix} \times \left[ \begin{pmatrix} h^1 & h^2 & h^3 \\ h^4 & h^5 & h^6 \\ h^7 & h^8 & h^9 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \right] = \mathbf{0}. \tag{6.2}$$

The homography $\mathbf{H}$ is defined up to a scale factor. Specifically, a multiplication by an arbitrary non-zero scale factor does not alter the relation of Equation 6.2. Therefore, we can say that the matrix $\mathbf{H}$ is homogenous, since as in the homogenous representation of a point only the ratios of its elements are important. Given that we have eight ratios amongst the nine elements of $\mathbf{H}$, we can say that the homography has 8 degrees of freedom.

We briefly review Direct Linear Transformation algorithm for the estimation of the planar homography $\mathbf{H}$, Eq. (6.1), from known world and image coordinates [55].

Evaluation of the cross product, Equation 6.2 yields:

$$\begin{pmatrix} \tilde{x}^2(h^7x^1 + h^8x^2 + h^9x^3) - \tilde{x}^3(h^4x^1 + h^5x^2 + h^6x^3) \\ \tilde{x}^3(h^1x^1 + h^2x^2 + h^3x^3) - \tilde{x}^1(h^7x^1 + h^8x^2 + h^9x^3) \\ \tilde{x}^1(h^4x^1 + h^5x^2 + h^6x^3) - \tilde{x}^2(h^1x^1 + h^2x^2 + h^3x^3) \end{pmatrix} = \mathbf{0}. \tag{6.3}$$

This constraint is linear in the unknown elements of the homography $h^i$. Re-ordering the terms yields the following system of linear equations:

$$\begin{pmatrix} 0 & \tilde{x}^3x^1 & -\tilde{x}^2x^1 \\ 0 & \tilde{x}^3x^2 & -\tilde{x}^2x^2 \\ 0 & \tilde{x}^3x^3 & -\tilde{x}^2x^3 \\ -\tilde{x}^3x^1 & 0 & \tilde{x}^1x^1 \\ -\tilde{x}^3x^2 & 0 & \tilde{x}^1x^2 \\ -\tilde{x}^3x^3 & 0 & \tilde{x}^1x^3 \\ \tilde{x}^2x^1 & -\tilde{x}^1x^1 & 0 \\ \tilde{x}^2x^2 & -\tilde{x}^1x^2 & 0 \\ \tilde{x}^2x^3 & -\tilde{x}^1x^3 & 0 \end{pmatrix}^T \begin{pmatrix} h^1 \\ h^2 \\ h^3 \\ h^4 \\ h^5 \\ h^6 \\ h^7 \\ h^8 \\ h^9 \end{pmatrix} = \mathbf{0}$$

$$\mathbf{A}\mathbf{h} = \mathbf{0}. \tag{6.4}$$

where $\mathbf{A}$ is a $3 \times 9$ matrix and $\mathbf{h}$ is a 9-vector containing the entries of the matrix $\mathbf{H}$.

A matched set of points $\tilde{\mathbf{x}}$ and $\mathbf{x}$ appear to provide three constraints on the eight unknown elements of $\mathbf{h}$ (the homography is defined only up to an unknown scale factor,

reducing the unknowns from nine to eight). The rows of the matrix $\mathbf{A}$, however, are not linearly independent. As such, this system provides two constraints in eight unknowns. In order to solve for $\mathbf{h}$, we require four or more points with known image, $\tilde{\mathbf{x}}$, and (planar) world, $\mathbf{x}$, coordinates. From four or more points , standard least-squares techniques can be applied, as described in [55]. Of course we seek for a solution that is non-zero, since the obvious solution $\mathbf{h} = \mathbf{0}$ is not of our interest. Since data are usually noisy, we attempt to find an approximate solution, namely a vector $\mathbf{h}$ which minimizes a cost function. To avoid the null solution, we can impose a constraint on the norm, such as $\mathbf{h}\| = 1$. Since the homography $\mathbf{H}$ is defined up to a scale factor, the value imposed norm is not important. We can minimize the the norm $\|\mathbf{Ah}\| = 0$ under the constraint of $\|\mathbf{h}\| = 1$. The solution of such system is the minimal eigenvalue eigenvector of $\mathbf{A}^T\mathbf{A}$.

With the Direct Linear Transformation algorithm it is possible to compute an homography, starting from correspondences between image $\tilde{\mathbf{x}}$ and world $\mathbf{x}$ coordinates. For images, many interest points, such as corners, blobs, edges or points, can be identified and subsequently matched according to correlation matching algorithms.

Scale-invariant feature transform (SIFT) [98] is one of the most popular method to extract features in images. The algorithm was initially proposed by Lowe in 1999 and is now patented in the United States by the University of British Columbia. It extracts highly distinctive image keypoint descriptors, which have the characteristic to be invariant to certain amounts of image scale, rotation, affine distortion, noise, and illumination differences. The image is firstly convoluted with Gaussian filters at different scales and the difference of adjacent blurred images generates the Difference-of-Gaussian images. SIFT keypoints correspond to local maxima or minima (extrema) of the difference-of-Gaussians images in the scale-space. The position of all the interest points is accurately determined by interpolating neighboring data. Low contrast keypoints and responses along edges are removed, in order to make keypoints more robust for matching and recognition. An orientation is assigned to each keypoint based on gradient orientation histogram. All the properties of a keypoint are measured relative to this orientation, providing invariance to rotation. In particular, the full feature descriptor is computed as a set of orientation histograms on a $4 \times 4$ pixel neighborhoods, for a total of 128 elements. Keypoints are then matched between two images using a variant of nearest neighbor matching on the feature vectors. This association accounts for a geometric transformation between the images by matching keypoints up to a planar homography.

The RANdom SAmple Consensus (RANSAC) algorithm [46] is subsequently used to minimize the effect of mis-matched keypoints (outliers). RANSAC iteratively selects a random subset of the input data. Assuming these data to be inliers, a parametric model is fitted to them (usually an homography). All the rest of the data is then tested against the reconstructed model.: those points which are in agreement with the model will be considered inliers, while those which do not fit the model will be classified as outliers. The

The 3-D world and 2-D image coordinates should be translated so that their centroid is at the origin, and scaled isotropically so that the average distance to the origin is $\sqrt{2}$. This normalization improves stability of the homography estimation in the presence of noise [55].

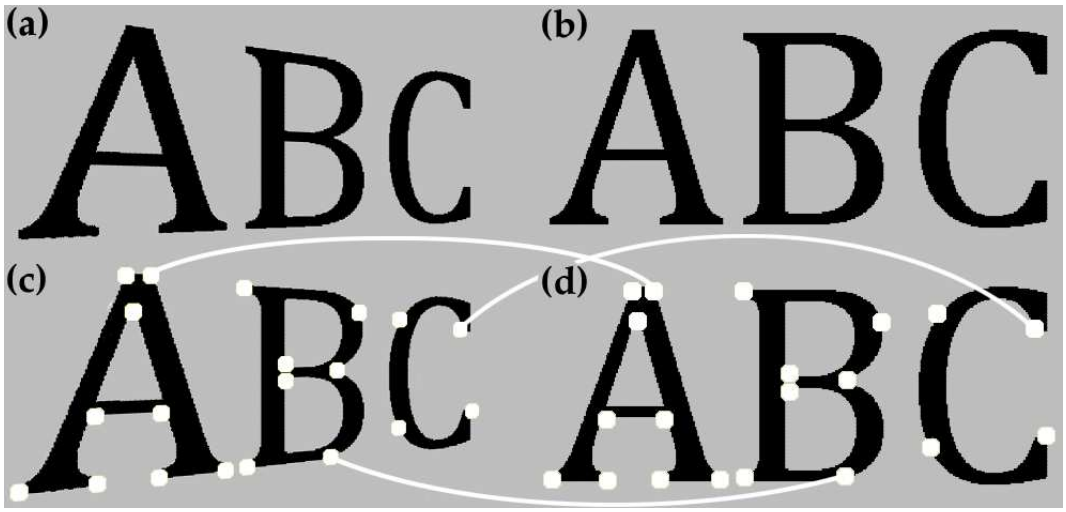



Figure 6.4: Shown in panels (a) and (c) is a text string in image coordinates, and shown in panels (b) and (d) is the same string in world coordinates. The dots in panels (c) and (d) correspond to a subset of the extracted coordinates used to estimate the image to world homography.

more inliers, the better will be the estimated model. Since it was initially estimated from a random set of inliers, the model is re-estimated with all the hypothetical inliers. Finally an error is calculated by estimating the total deviation of the inliers to the model.This process is repeated iteratively until a maximum number of iteration is reached. The model which will return the smallest error will be taken, successfully identifying the outliers.

In our case, the required world coordinates are determined by re-creating the text in question with no distortion. We next consider the case when the font style is known and when the font style is unknown. Since the homography is only estimated up to an unknown scale factor, the font size is arbitrary.

### 6.2.2 Known Font

Shown in Figure 6.4(a) is the text string "ABC" after distortion by a planar homography, as in Eq. (6.1). Assuming that the font style is known, this string in its world coordinate system can easily be determined, as depicted in Figure 6.4(b). From this pair of images, we automatically extract the image and world coordinates required for the planar homography estimation, as described next.

Exploiting their sufficient invariance to affine distortion, we employ the SIFT operator [98], and RANSAC algorithm, to identify the coordinates of distinctive image keypoint positions. Recall that the matching algorithm accounts for a geometric transformation between the images by matching keypoints up to a planar homography.

Shown in Figure 6.4(c) and (d), for example, are a subset of the extracted keypoints (dots) for the images shown in panels (a) and (b).

### 6.2.3 Unknown Font

When the font of the text in question cannot be easily determined by visual inspection, we adopt the following technique for automatically identifying the font style. We begin by constructing the text in question in undistorted world coordinates with all available font styles. Then, the SIFT operator is applied to each of these images, as described in Section 6.2.2. The font style that returns the largest number of matched keypoints is taken to be the correct font.

### 6.2.4 Photo Composite

Given an image of text that has undergone planar perspective projection (i.e., a homography), we have described how to determine the required image and world coordinates, and how to estimate the world to image homography. Except for degenerate cases, it is always possible to calculate a homography regardless of the authenticity of the underlying text. We will show, however, that when the text is inconsistent with a perspective planar projection, the estimated homography yields a large reconstruction error.

Specifically, the inverse homography is applied to the keypoints in image coordinates yielding rectified world coordinates:

$$\mathbf{x}_r \;=\; \mathbf{H}^{-1}\tilde{\mathbf{x}}. \tag{6.5}$$

It is unlikely in an inauthentic image to have the image coordinates precisely satisfy the proper planar perspective distortion. In this case, the rectified image $\mathbf{I}_r$ is unlikely to match the world image $\mathbf{I}_w$. On the other hand, in the case of an authentic image, the rectified image should be a good approximation of the world image. As such, we use the root mean square (RMS) error between the world and rectified image as a measure of authenticity:

$$e \;=\; \frac{1}{\sqrt{n_x n_y}}\|\mathbf{I}_w - \mathbf{I}_r\|, \tag{6.6}$$

where $n_x$ and $n_y$ are the image dimensions and $\|\cdot\|$ denotes vector 2-norm. Note that this error is computed on the underlying intensity image, as opposed to the extracted keypoint coordinates.

Because a homography captures a broad range of distortions, we have found it more effective to estimate the homography from several small subsets of the matched keypoints and then compute the RMS error for each estimated homography. The average RMS error is used as a measure of authenticity:

$$E \;=\; \frac{1}{N}\sum_{i=1}^{N} e_i, \tag{6.7}$$

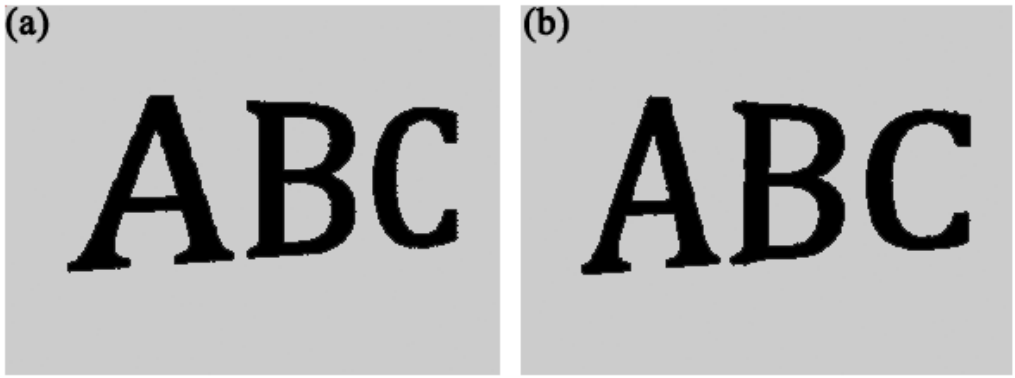



Figure 6.5: Shown are (a) authentically projected text and (b) matched inauthentic text generated by isotropically scaling and affine transforming text to best match panel (a).

where $N$ is the total number of subsets and $e_i$ is the RMS error, Eq. (6.6), for the $i^{th}$ subset. Specifically, an error $E$ above a specified threshold is taken to be evidence of tampering.

## 6.3 Results

We describe a set of simulations to verify the efficacy of the proposed technique. A set of authentic images were first created by generating images consisting of a text string with six letters in one of 350 font styles. A planar homography, Eq. (6.1), was then applied by considering physically plausible intrinsic and extrinsic camera parameters. Specifically, with a fixed focal length ($f$=10), we let the principal point vary in $[-0.5, 0.5]$, we allowed a rotation up to 20 degrees. A matched inauthentic image was generated that approximated the appearance of the authentic image, while not precisely satisfying a planar homography, Figure 6.5. Specifically, the image in world coordinates was subjected to a non-linear anisotropic scaling followed by a six parameter affine transformation constructed to optimally match (in the least-square sense) the authentic image. The result was a perceptually convincing transformation.

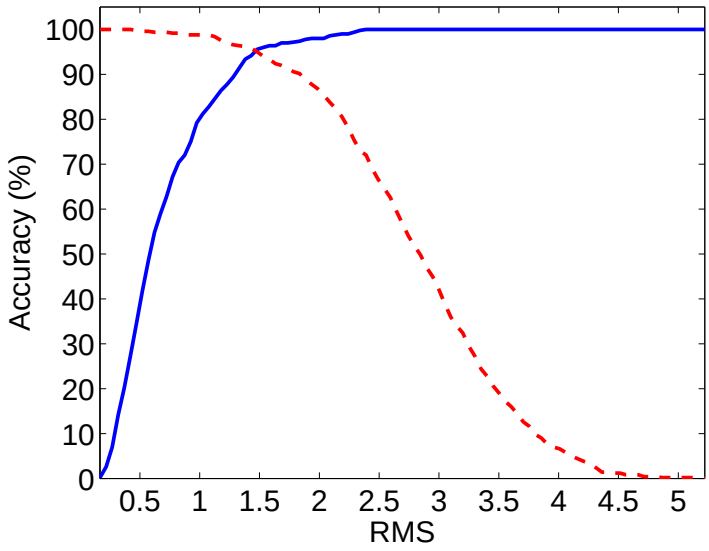
We generated 500 such authentic and inauthentic images. Each image was $1200 \times 900$ pixels in size, and rendered as a 1-bit binary image. For each image, we assumed a known font style, automatically extracted the image and world coordinates, estimated the world to image homography, and computed the reconstruction error with $N = 100$, Eq. (6.7).

Shown in Figure 6.6(a) is the resulting ROC curve where the horizontal axis corresponds to the RMS error, and the vertical axis to the classification accuracy. The solid curve corresponds to the authentic images, and the dashed curve corresponds to the inauthentic images. The intersection of these curves corresponds to an overall accuracy of

95%. The detection accuracy and false alarm rate (incorrectly classifying an authentic image as inauthentic) can be controlled by adjusting the RMS threshold. For example, a false alarm rate of 2% yields a detection accuracy of 88% and a false alarm rate of 1% yields a detection accuracy of 82%.

Shown in Figure 6.6(b) is the ROC curve for the case when the font style is unknown. In this case, the intersection of the curves corresponds to an overall accuracy of 92%. Note that the overall accuracy is similar, with a slight degradation due to some errors in the font identification stage.

(a)




(b)

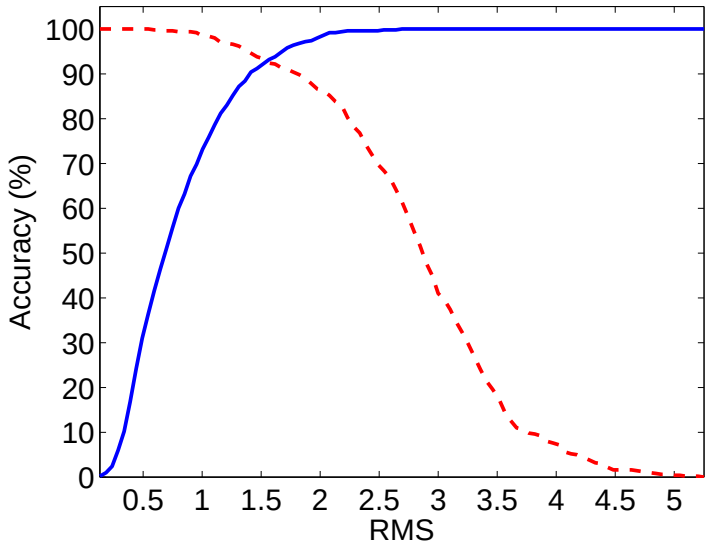
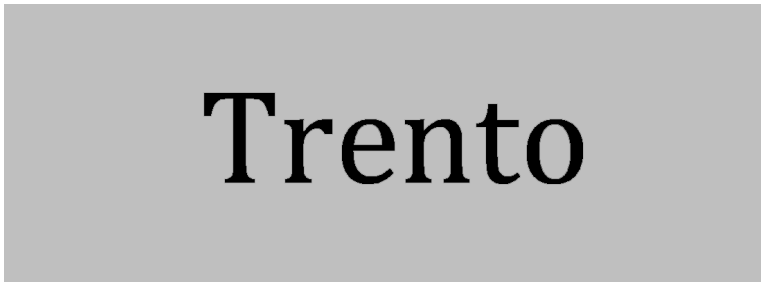



Figure 6.6: Shown are ROC curves for classification with (a) known font style and (b) unknown font style. The solid curve corresponds to the authentic images, and the dashed curve corresponds to the inauthentic images.


(a) Original photographs


(b) Binary image

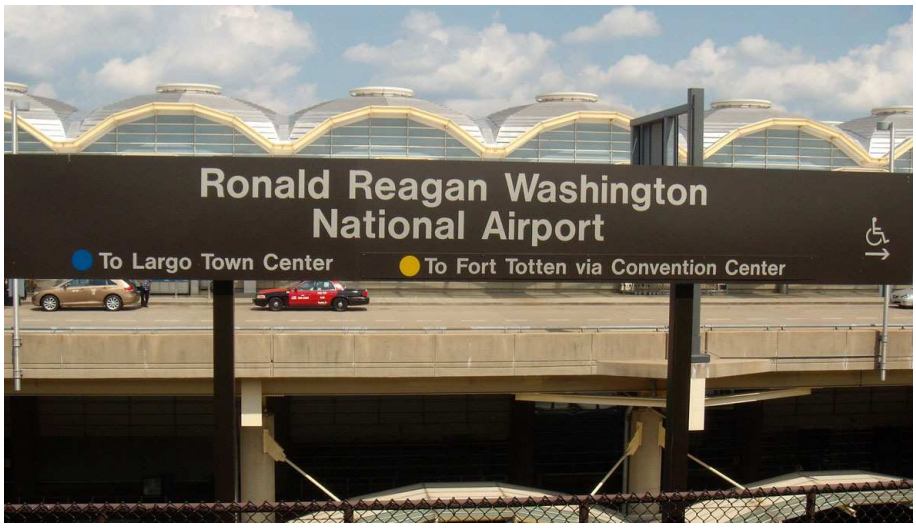



(c) Matched and rectified image.

Figure 6.7: Example of real tested image containing text. In panel (a) the original photograph is shown. In order to identify the font via SIFT matching, the image is binarized (panel (b)). The font style that returned the largest number of matched keypoints was taken to be the correct font (panel (c)).

Following we analyzed real images. As a proof of concept, we tested at first images of our creation containing simple text strings. Since we generated the text we can have a ground truth also for the font identification step. Figure 6.7 shows an example of this analysis. A photograph with a printed text reading "Trento" was taken to be analyzed. The image has been firstly made binary not to consider the background during SIFT computation (panel (b)). The font style that returned the largest number of matched keypoints was taken to be the correct one. Specifically, the original text was produced with "Cambria" style and "Cambria Math" font style was identified, among the large database. Given the correct font style, we could compute the homography between the images and rectify the original text. The reconstruction error, Eq. (6.7), for the text "Trento" was $E = 0.75$. Therefore such image could be correctly classified with a threshold of 1.5 ( i.e., an overall accuracy of 92%, the intersection of the ROC curves in Figure 6.6(b)).

(a)




(b)

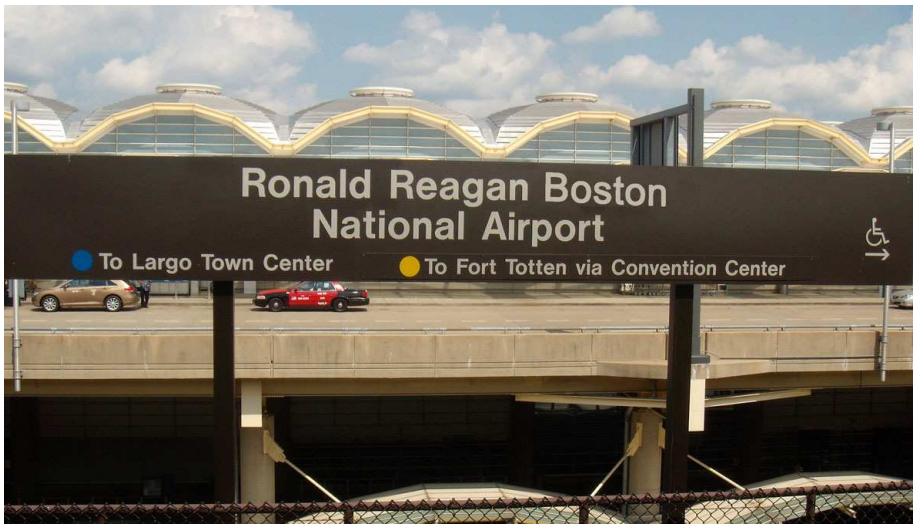



Figure 6.8: Shown in panel (a) is an authentic image, while in panel (b) a forged version is depicted, where the text "Washington" has been replaced with the string "Boston".

Finally, we analyzed real images, where text has been forged. Shown in Figure 6.8 is an example our creation. In panel (a) the authentic image is represented, and shown in panel (b) is the corresponding visually plausible fake. The reconstruction error, Eq. (6.7), for the manually extracted strings "Washington" and "Boston" were $E = 0.94$ and $E = 1.9$, respectively. These images are correctly classified with a threshold of 1.5 ( i.e., an overall accuracy of 92%, the intersection of the ROC curves in Figure 6.6(b)). A second example of tested image is shown in Figure 6.9. In the original image (panel (a)) the text "Amore" has been replaced with the string "Hong Kong", creating a visually compelling forgery (panel (b)). For the these images, the evaluated reconstruction error for the strings

(a)

(b)


Figure 6.9: Shown in panel (a) is an authentic image, while in panel (b) a forged version is depicted, where the text "Amore" has been replaced with the string "Hong Kong".

"Amore" and "Hong Kong" were $E = 1.23$ and $E = 1.94$, respectively. As before, by imposing a threshold of 1.5, the photographs are correctly classified.


Figure 6.10: Examples of tested images. In the red circle, the text string we extracted and analyzed.

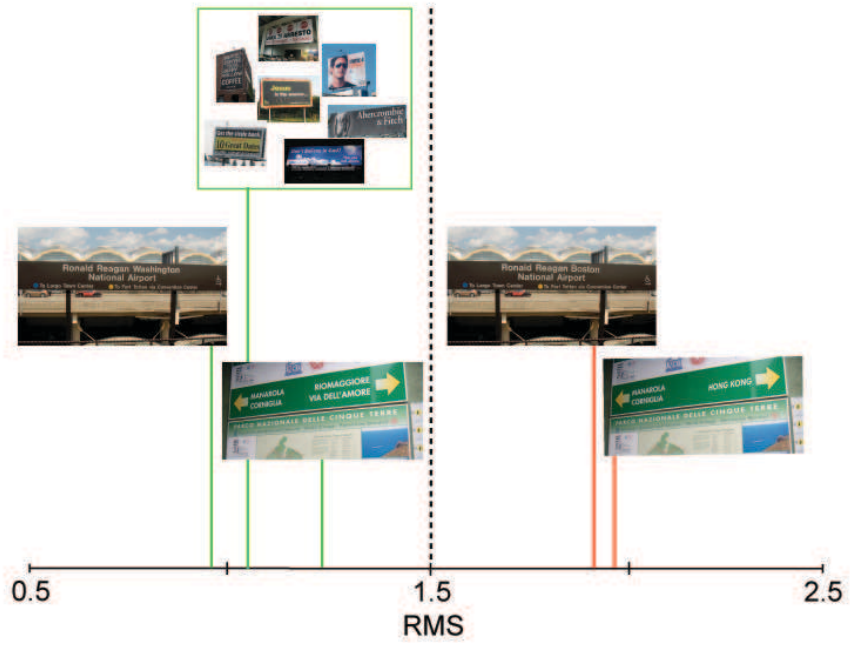



Figure 6.11: Shown are the tested images, plotted based on their reconstruction errorr. When setting an RMS threshold of 1.5, based on the simulations, all the images are correctly classified.

To further validate this method, we tested a total of ten authentic images, obtaining reconstruction errors in the range of 0.93 to 1.35, with a median of 1.05. In Figure 6.10 examples of the tested images are shown, where the red circle indicates the analyzed text. Also for these images, by setting the threshold to 1.5, a correct classification could be achieved, as can be seen in Figure 6.11.

## 6.4 Discussion

We have presented a new forensic technique for authenticating text in photographs. Because it is relatively easy to digitally insert text into a photo in a visually compelling manner, it can be difficult to manually determine if text is authentic. Our forensic technique explicitly estimates the perspective projection of text onto a planar surface. We have shown that inauthentic text often violates the rules of perspective projection and can therefore be detected.

This approach is semi-automatic, requiring only a user to manually select the text in question. In the case when the text font style in question is unknown, this approach requires a sufficiently large database of font styles from which the required world coordinates are extracted.

A determined forger could circumvent this technique by applying the correct homography to the inserted text [33]. This would, of course, require the forger to estimate the correct homography from the image, which is outside of the expertise of the average Photoshop user.

# Chapter 7

# Exposing Digital Forgeries in Ballistic Motion

*In this chapter we describe a geometric technique to detect physically implausible trajectories of objects in video sequences. This technique explicitly models the three-dimensional ballistic motion of objects in free-flight and the two-dimensional projection of the trajectory into the image plane of a static or moving camera. Deviations from this model provide evidence of manipulation. The technique assumes that the object's trajectory is substantially influenced only by gravity, that the image of the object's center of mass can be determined from the images, and requires that any camera motion can be estimated from background elements. The computational requirements of the algorithm are modest, and any detected inconsistencies can be illustrated in an intuitive, geometric fashion. We demonstrate the efficacy of this analysis on videos of our own creation and on videos obtained from video-sharing websites.*

## Acknowledgment

I would like to thank prof. Hany Farid for hosting this research within the Image Science Group at Dartmouth College (USA) and prof. James O'Brien, University of Berkeley (USA), for the inspiring collaboration and valuable comments.
Morover, a special thank goes to Marty Banks for his helpful comments and for inspiring this work by forwarding the video "megawoosh" (*http://www.youtube.com/watch?v=E2D6MJzva7E*) and to Eric Kee for his basketball skills and helpful comments.

Part of this Chapter appears in:

- V. Conotter, J. O'Brien and H. Farid, "Exposing Digital Forgeries in Ballistic Motion". Submitted to *IEEE Transaction on Information Forensics & Security*, 2011.

## 7.1 Introduction

Increasingly sophisticated video editing and special effects software has made it possible to create forged video sequences that appear to contain realistic dynamic motion. For example, video sharing websites are littered with titles like "Seriously Amazing Best Beer Pong Shots," "Dude Perfect Amazing Basketball Shots," and "Epic Pool Jump." These videos appear to show spectacular basketball shots, gravity-defying acrobatics, and the bone-crushing results of daredevil leaps and jumps. Some of these videos are real, but many are fake.

We describe a forensic technique that is tailored to determine if video of a purportedly ballistic motion, such as a ball being thrown or a person jumping through the air, is consistent with the geometry and physics of a free-falling projectile.
Posted videos are often of low-quality and typically have undergone a variety of post-processing and re-compression. As such, statistical techniques that focus on double-compression artifacts [163, 21, 169], interlaced and de-interlaced correlations [164], or sensor noise patterns [110, 60, 59] are unlikely to apply. In addition, forensic techniques based on detecting frame or region duplication [166, 8] will only apply when some form of duplication was necessary to create the fake.

In contrast to these statistical techniques, we propose a geometric forensic technique that is largely insensitive to resolution, post-processing, compression, and re-compression. We begin by describing a plausible, albeit somewhat simplified, physical model for the expected trajectory of a projectile motion, and a basic imaging model for a static or moving camera. We then describe a technique to determine if the image of the trajectory of a projectile motion is consistent with this physical model. Fig. 7.1 visually explains such idea. An authentic (top) and fake video (bottom) are shown as a time-lapse composite. The small red dots specify the tracked position of the ball, and the larger yellow dots and solid lines denote the best fit model of a ballistic trajectory. Inconsistencies between the two trajectories provides evidence of tampering.

Because a projectile's motion is influenced by gravity, the vertical component of acceleration is roughly constant (assuming no air resistance or other external forces). In principle a human observer binocularly viewing a projectile could use this information to estimate the absolute size and distance to the projectile [172]. In practice, however, most observers do not consciously take advantage of this information [57]. These studies suggest that if observers are not able to take advantage of such constraints in a three-dimensional setting, then it is most likely that they will not be able to accurately reason about the two-dimensional projection of a projectile. It follows that relying on the visual system to authenticate a projectile's motion becomes even more unlikely when the camera moves along an arbitrary and unknown path.

The technique here proposed makes minimal assumptions about the nature of the trajectory and camera, and requires only limited manual input. The core computational requirements of the algorithm are modest, and the analysis can be presented as a simple, intuitive geometric construction. We demonstrate the efficacy of this analysis on videos of our own creation and several obtained from video-sharing websites.

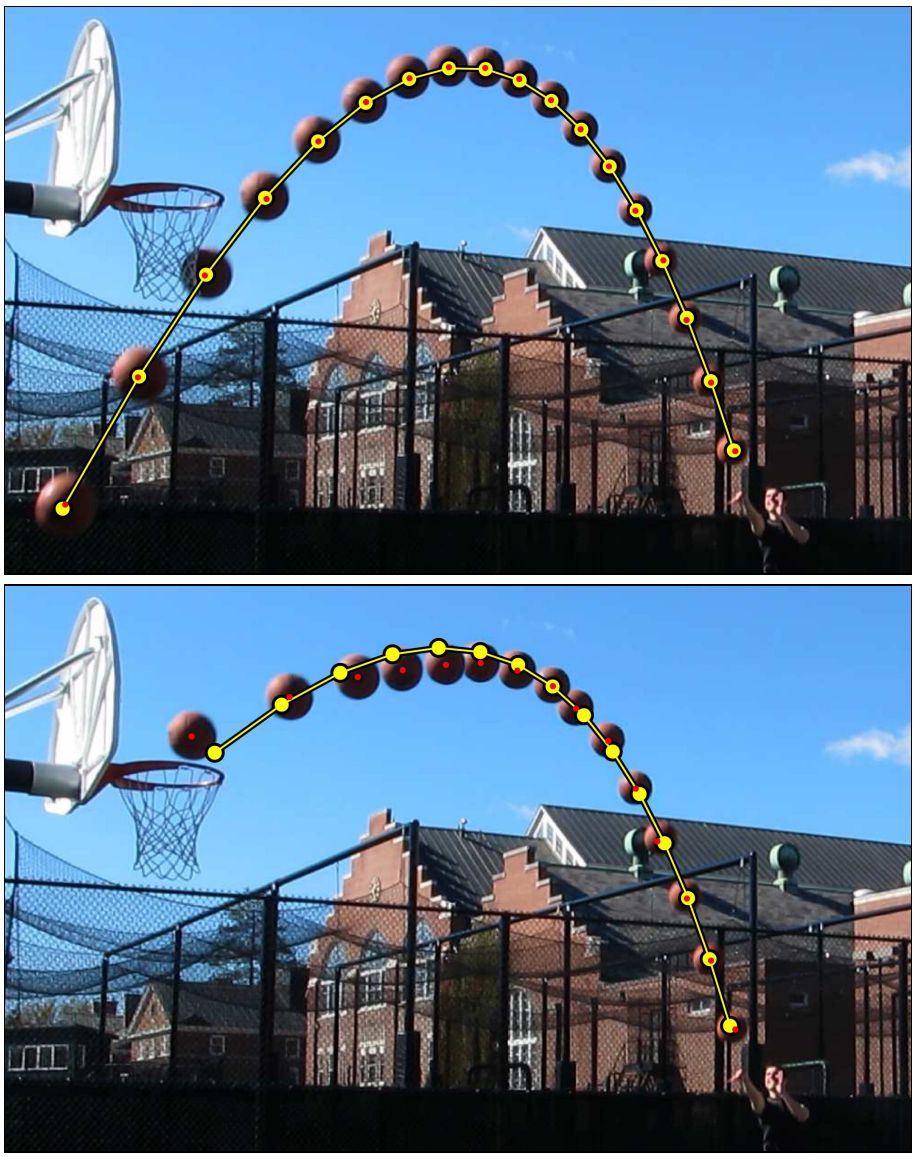


Figure 7.1: An authentic (top) and fake video (bottom) each shown as a time-lapse composite. The small red dots specify the tracked position of the ball, and the larger yellow dots and solid lines denote the best fit model of a ballistic trajectory. A discrepancy between the model and the ball's actual positions in the lower image provides proof that the depicted motion is fake.

## 7.2 The proposed approach

Our approach to exposing inauthentic motion relies on four explicit assumptions:

- The motion of the object in three-dimensional space is ballistic (*i.e.,* only gravitational acceleration) so that the trajectory of the object's center of mass will describe a simple parabola.
- The moving object is sufficiently well-known so that the location of its center of mass can be estimated for each frame.
- The scene is imaged under linear perspective projection.
- Any movement or changes in the camera parameters during the sequence can be computed from the scene background so that all frames can be placed into a common coordinate system.

With regard to notation, scalars are denoted with lower-case italic letters $s$, vectors are denoted with lower-case bold-face letters $\mathbf{v}$, and matrices are denoted with upper-case bold-face letters $\mathbf{M}$. Image coordinates are denoted as $\tilde{\mathbf{p}}$, and world and camera coordinates are denoted as $\mathbf{p}$. Superscripts are used to denote the components of a vector $\mathbf{v} = \begin{bmatrix} v^x & v^y & v^z \end{bmatrix}$ or the rows of a matrix: $\mathbf{m}^k$ is the $k^{th}$ row of matrix $\mathbf{M}$. Subscripts are most commonly used to denote position as a function of time, $\mathbf{p}_t$.

### 7.2.1 Trajectory Geometry

The center of mass of a projectile, in the absence of air resistance or any other external forces, follows a ballistic trajectory that can be described in three dimensions with a time-parametrized parabolic curve:

$$\mathbf{p}_t \;=\; \mathbf{p}_0 + \mathbf{v}t + \tfrac{1}{2}\mathbf{a}t^2 \tag{7.1}$$

where $t$ denotes time, $\mathbf{p}_0$ is the initial position, $\mathbf{v}$ is the initial velocity, and $\mathbf{a}$ is the acceleration (due to gravity). In the absence of any other external forces, the path of the projectile is planar. Assuming linear perspective projection under a pinhole camera model, the image of a projectile's trajectory, in homogeneous coordinates, is:

$$\tilde{\mathbf{q}}_t \;=\; \mathbf{H}\mathbf{p}_t, \tag{7.2}$$

where $\mathbf{H}$ is a $3 \times 3$ matrix embodying the planar perspective projection (*i.e.*, a homography).

Consider the special case where the optical axis of the camera is orthogonal to the plane of motion, the focal length is unit length, and the principal point is the image center. In this case, the z-component of the velocity is zero, and the x- and z-components

of the acceleration are zero. The trajectory, Equation (7.1), then simplifies to:

$$\begin{bmatrix} p_t^x \\ p_t^y \\ p_t^z \end{bmatrix} = \begin{bmatrix} p_0^x \\ p_0^y \\ p_0^z \end{bmatrix} + \begin{bmatrix} v^x \\ v^y \\ 0 \end{bmatrix} t + \frac{1}{2} \begin{bmatrix} 0 \\ -g \\ 0 \end{bmatrix} t^2, \tag{7.3}$$

where $g$ is gravity and the world coordinate system is defined such that the z-axis is the optical axis and the positive y-axis points upward. In addition, the world to image transformation is simply $\tilde{\mathbf{q}}_t = \mathbf{p}_t$ (*i.e.*, $\mathbf{H} = \mathbf{I}$). In non-homogeneous coordinates, this yields:

$$\tilde{q}_t^x = \frac{p_0^x + v^x t}{p_0^z} \tag{7.4}$$

$$\tilde{q}_t^y = \frac{p_0^y + v^y t - \frac{1}{2} g t^2}{p_0^z}. \tag{7.5}$$

Note that in this special case the projectile's path maps to a parabola in image coordinates, which can be seen more clearly by rewriting the above equations as:

$$\begin{bmatrix} \tilde{q}_t^x \\ \tilde{q}_t^y \end{bmatrix} = \begin{bmatrix} \frac{p_0^x}{p_0^z} \\ \frac{p_0^y}{p_0^z} \end{bmatrix} + \begin{bmatrix} \frac{v^x}{p_0^z} \\ \frac{v^y}{p_0^z} \end{bmatrix} t + \frac{1}{2} \begin{bmatrix} 0 \\ \frac{-g}{p_0^z} \end{bmatrix} t^2. \tag{7.6}$$

Under an arbitrary homography, however, the image of a projectile will not necessarily be a parabola. Specifically:

$$\tilde{\mathbf{q}}_t = \mathbf{H}\mathbf{p}_t = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \mathbf{p}_t, \tag{7.7}$$

In non-homogeneous image coordinates, the projectile takes the form:

$$\tilde{q}_t^x = \frac{\mathbf{h}^1 \cdot \mathbf{p}_t}{\mathbf{h}^3 \cdot \mathbf{p}_t} = \frac{\mathbf{h}^1 \cdot (\mathbf{p}_0 + \mathbf{v}t + \frac{1}{2}\mathbf{a}t^2)}{\mathbf{h}^3 \cdot (\mathbf{p}_0 + \mathbf{v}t + \frac{1}{2}\mathbf{a}t^2)} \tag{7.8}$$

$$\tilde{q}_t^y = \frac{\mathbf{h}^2 \cdot \mathbf{p}_t}{\mathbf{h}^3 \cdot \mathbf{p}_t} = \frac{\mathbf{h}^2 \cdot (\mathbf{p}_0 + \mathbf{v}t + \frac{1}{2}\mathbf{a}t^2)}{\mathbf{h}^3 \cdot (\mathbf{p}_0 + \mathbf{v}t + \frac{1}{2}\mathbf{a}t^2)}, \tag{7.9}$$

where $\cdot$ is inner product and $\mathbf{h}^i$ is the $i^{th}$ row of the homography $\mathbf{H}$. Note that these image coordinates follow a rational parabola described by a ratio of second-order polynomials, and cannot be expressed as a single second-order polynomial as in Equation (7.6).

Given a composite image from a still camera showing the path of an object, one could test the authenticity of the motion by verifying that it can be fit by a rational parabola using Equations (7.8) and (7.9). However that approach does not explicitly reveal the direction of gravity, provide a trajectory in world space, or generalize to a moving camera.

### 7.2.2 Projectile Estimation: Static Camera

In the previous section, the projectile was specified in a world coordinate system, Equation (7.1), and its projection was specified in an image coordinate system, Equation (7.2). In the following sections, we will specify all coordinates with respect to a common three-dimensional coordinate system in which the origin is the camera center, the image plane is $f$ units from the origin (the focal length), and the optical axis is orthogonal to the image plane. With this notation, a projectile is specified as $\mathbf{p}_t$ in Equation (7.1), its projection is specified as a three-vector $\mathbf{q}_t$ in which the z-component $q_t^z = f$, and the camera center, $\mathbf{c}$, is the origin of the three-dimensional coordinate system. (See Fig. 7.2(a).)

For each moment in time $t = 1 \ldots n$, define a line from the camera center, $\mathbf{c}$, through the image of the projectile's center of mass, $\mathbf{q}_t$, as:

$$\mathbf{l}_t = \mathbf{c} + s_t(\mathbf{q}_t - \mathbf{c}), \tag{7.10}$$

where $s_t$ is a parametric variable for the line. If the projectile $\mathbf{p}_t$ follows a parabolic trajectory then there exists a value of the parametric variable at each moment in time that satisfies:

$$\mathbf{p}_t = \mathbf{l}_t \tag{7.11}$$

$$\mathbf{p}_0 + \mathbf{v}t + \tfrac{1}{2}\mathbf{a}t^2 = \mathbf{c} + s_t(\mathbf{q}_t - \mathbf{c}) \tag{7.12}$$

for some values of $s_t$. Expanding in terms of the individual components yields:

$$p_0^x + v^x t + \tfrac{1}{2}a^x t^2 = c^x + s_t(q_t^x - c^x) \tag{7.13}$$

$$p_0^y + v^y t + \tfrac{1}{2}a^y t^2 = c^y + s_t(q_t^y - c^y) \tag{7.14}$$

$$p_0^z + v^z t + \tfrac{1}{2}a^z t^2 = c^z + s_t(q_t^z - c^z) \tag{7.15}$$

This system of equations is linear in terms of both the nine unknowns that specify the object's trajectory through space ($\mathbf{p}_0$, $\mathbf{v}$, and $\mathbf{a}$) and the $n$ unknowns that specify the object's parametric location along the line from the camera ($s_t$). With a sufficient number of frames ($n \geq 5$) where the projectile's position can be observed, it is possible to solve for these unknowns by performing a least-squares fit to the data. However, in the absence of noise, the linear system will always be rank deficient so that the least-squares solution is not unique. This deficiency corresponds to a scale ambiguity in the solution: any solution could be scaled about the origin and still satisfy the constraints, Fig. 7.2(a).

Before describing a robust solution that copes with both this ambiguity and noise, we will first discuss the situation for a moving camera. Our solution treats both still and moving cameras using the same framework.

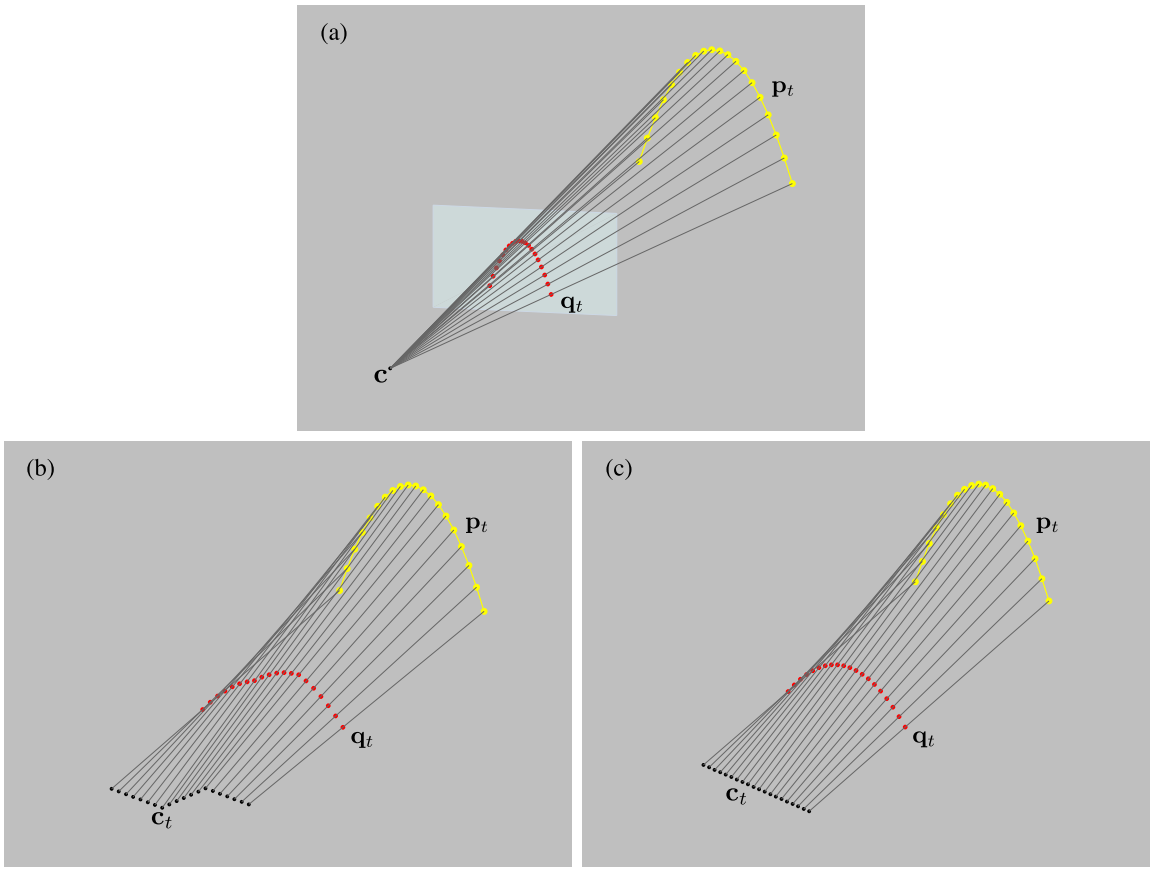



Figure 7.2: For a stationary camera (a), a ballistic trajectory will appear as a rational parabola in the image plane (red dots). Lines from the center of projection $\mathbf{c}$ (black dot) through these points $\mathbf{q}_t$ on the image plane will intersect the object's true parabolic trajectory through space $\mathbf{p}_t$ (yellow dots). A scale ambiguity appears because any slice through the generalized cone formed by the lines would also produce a parabolic arc if that slice is parallel to the object's true parabolic trajectory. With a moving camera (b), the lines from the moving center of projection $\mathbf{c}_t$ (black points) through observed locations for the object do not form a cone, but they must still intersect the object's parabolic trajectory through space. In general, other slices through the lines will not form parabolae unless the camera's path is also a quadratic curve as shown in (c) in which case any isoparametric slice through the lines will also produce a parabolic trajectory. The scale ambiguity for a stationary camera is special case of the parametric ambiguity for a camera on a quadratic path. For clarity, the image planes for each camera location in (b) and (c) are not shown.

### 7.2.3 Projectile Estimation: Moving Camera

In the previous section we described the estimation of a projectile that was being imaged by a static camera. In that case, the projection of the projectile was a parabola subjected to a planar homography. However, if the camera is not stationary, then the projection of the projectile can take an arbitrary path in the image plane that results from combining image coordinates from different local coordinate systems. If the camera motion can be estimated, then these image coordinates can be described in a single common coordinate system. The resulting estimation of projectile motion then becomes equivalent to the static camera case with the difference being that the location of the camera center varies over time. We will arbitrarily use the coordinate system of the initial frame in a sequence as our common coordinate system.

Consider a camera undergoing a rigid-body transformation (rotation and translation) with fixed intrinsic parameters (focal length, principal point, etc.). The image of a projectile's trajectory can be expressed with respect to the common coordinate system as $\mathbf{R}_t\mathbf{q}_t + \mathbf{t}_t$, where $\mathbf{q}_t$ is the image of the projectile in the camera's local coordinate system, and $\mathbf{R}_t$ and $\mathbf{t}_t$ are respectively the $3 \times 3$ rotation matrix and $3 \times 1$ translation vector that define the transformation between the local and common coordinate systems. The camera center is expressed with respect to the common coordinate system as $\mathbf{c}_t + \mathbf{t}_t$, where $\mathbf{c}_t$ is the location of the camera center in the local coordinate system.

Given the rotation and translation that relates a camera at each time to the common coordinate system, all coordinates can be expressed with respect to a single common coordinate system. As in the static case, for each frame a line is defined from the camera center, $\mathbf{c}_t$ (now a function of time), through the center of mass of the projection of the projectile into the image plane, $\mathbf{q}_t$, as:

$$\mathbf{l}_t = (\mathbf{c}_t + \mathbf{t}_t) + s_t(\mathbf{R}_t\mathbf{q}_t - \mathbf{c}_t) \tag{7.16}$$

where $s_t$ is the parametric variable. Note the similarity to Equation (7.10), with the exception that $\mathbf{q}_t$ and $\mathbf{c}_t$ undergo a coordinate transformation. If the projectile follows a parabolic trajectory, $\mathbf{p}_t$, then there exists a value of the parametric variable $s_t$, at each moment in time that satisfies:

$$\mathbf{p}_t = (\mathbf{c}_t + \mathbf{t}_t) + s_t(\mathbf{R}_t\mathbf{q}_t - \mathbf{c}_t) \tag{7.17}$$

or

$$p_0^x + v^x t + \tfrac{1}{2}a^x t^2 = (c_t^x + t_t^x) + s_t(\mathbf{r}_t^1 \cdot \mathbf{q}_t - \mathbf{c}_t) \tag{7.18}$$

$$p_0^y + v^y t + \tfrac{1}{2}a^y t^2 = (c_t^y + t_t^y) + s_t(\mathbf{r}_t^2 \cdot \mathbf{q}_t - \mathbf{c}_t) \tag{7.19}$$

$$p_0^z + v^z t + \tfrac{1}{2}a^z t^2 = (c_t^z + t_t^z) + s_t(\mathbf{r}_t^3 \cdot \mathbf{q}_t - \mathbf{c}_t) \tag{7.20}$$

where $\cdot$ is inner product and $\mathbf{r}_t^i$ is the $i^{th}$ row of the rotation matrix $\mathbf{R}_t$. Note that, as in the previous section, this system of equations is linear in the unknown projectile parameters ($\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$), and the parametric variables ($s_t$).

It is possible to contend with a varying focal length over time, but for notational simplicity we assume that the focal length is fixed.

As before, with a sufficient number of frames ($n \geq 5$) where the projectile's position can be observed, it becomes possible to solve for these unknowns by performing a least-squares fit to the data. However, unlike the situation with a still camera there is no longer a necessary scale ambiguity. Because the camera now moves arbitrarily, scaling the solution would also require scaling the camera path as well, but the scale of that path will have been fixed when the camera's motion was estimated.

Nevertheless, specific combinations of camera and projectile paths may still fail to produce a fully determined system, Fig. 7.2(b)-(c). In particular, if the motion of the camera center is also described by a parabola then any isoparametric interpolation between the projectile's true trajectory and the camera's trajectory will also be a parabola. These extra solutions correspond to uniform scaling of the vector of $s_t$ values that would produce the true trajectory. This ambiguity is the general case of the scale ambiguity that appears for a static camera, and the static camera can be viewed as just a special case where the camera motion follows the trivial parabola.

### 7.2.4 Camera Calibration

Before continuing, we briefly describe how to estimate the required camera transformations $\mathbf{R}_t$ and $\mathbf{t}_t$ needed to relate each video frame to a common coordinate system. Modern-day camera calibration techniques consist of three basic steps: (1) image features (SIFT, corners, edges) are automatically detected in each video frame; (2) the features are matched across video frames (cross-correlation, nearest neighbor, etc..); and (3) the matched features are used to estimate the camera parameters: extrinsic ($\mathbf{R}_t$, $\mathbf{t}_t$) and intrinsic (focal length, principal point). A commonly used and effective optimization technique for this last step is bundle adjustment [55]. This techniques optimizes the estimation of the projection parameters by minimizing the reprojection error, that is. the geometric image distance between the homogenous detected and reprojected image points for every view in which the 3D point appears.

Among the many available software tools for camera calibration, we employ *Voodoo Camera Tracker* [151]. Within this software, we employ the SIFT operator [98] to extract image features and the RANSAC algorithm for matching features (see Section 6.2.1). Please refer to Section 6.2.1 for a brief description of SIFT keypoints and RANSAC algorithm.

### 7.2.5 Size Constraints

Although the perspective projection of a parabola makes an ideal model for representing the image of ballistic trajectory, the previously mentioned scale ambiguity can lead to bad behavior when data are noisy or if the underlying motion is not actually ballistic. In these cases, the solution may be skewed dramatically in the otherwise unconstrained part of the solution space. This skew typically manifests as trajectories in planes that are nearly parallel to the view lines with unreasonably high velocities and accelerations.

To prevent these degenerate solutions, an optional constraint can be imposed based on the variation in size of a projectile over time. With the assumption that the actual projectile is of constant size, then its projected size in the image is inversely proportional to the distance between the object and camera center as measured orthogonal to the image plane. Accordingly, the additional constraints require that the trajectory's distance to the image plane vary based on measurements of the object's size in the input images.

Consider a spherical projectile with diameter $d$ at position $(x_1, y_1, z_1)$ relative to the camera center with the z-axis being perpendicular to the image plane. The projected size of this projectile will be $\tilde{d}_1 = fd/z_1$, where $f$ is the camera focal length. As the projectile moves to another position $(x_2, y_2, z_2)$, the projected size is $\tilde{d}_2 = fd/z_2$. The ratio of these projections is $\tilde{d}_2/\tilde{d}_1 = z_1/z_2$. Note that this ratio does not depend on the focal length $f$ or diameter $d$.

For the static camera case, this basic constraint takes the form:

$$\frac{l_1^z - c^z}{l_k^z - c^z} = \frac{\tilde{d}_k}{\tilde{d}_1}, \tag{7.21}$$

where $c^z$ is the z-component of the camera center $\mathbf{c}$, $\tilde{d}_1$ and $\tilde{d}_k$ are the measured sizes of the projectile in the image at times $t = 1$ and $t = k$, and $l_1^z$ and $l_k^z$ are the z-components of $\mathbf{l}_1$ and $\mathbf{l}_k$ as defined in Equation (7.10). This constraint expands to:

$$s_1 \left[\frac{q_1^z - c^z}{q_k^z - c^z}\right] - s_k \left[\frac{\tilde{d}_k}{\tilde{d}_1}\right] = 0. \tag{7.22}$$

Note that this constraint is linear in the unknown parametric variables $s_1$ and $s_k$. These linear constraints for all $k = 2 \ldots n$ can be included when solving for the trajectory parameters. For the moving camera case, the constraint in Equation (7.22) takes the form:

$$s_1 \left[\frac{\mathbf{r}_1^3 \cdot \mathbf{q}_1 - \mathbf{c}_1}{\mathbf{r}_k^3 \cdot \mathbf{q}_k - \mathbf{c}_k}\right] - s_k \left[\frac{\tilde{d}_k}{\tilde{d}_1}\right] = 0, \tag{7.23}$$

where $\mathbf{r}_k^3$ is the third row of the rotation matrix $\mathbf{R}_k$.

### 7.2.6 Trajectory Estimation

The constraints from the proceeding sections can be assembled into a linear system of equations. For both the static and moving camera cases, this system will have a similar structure. This system can be over-constrained, either because the data cannot be fit by a parabola or because of noise caused by small measurement errors. This system can also be under-constrained due to a scale ambiguity, as previously described in Sections 7.2.2 and 7.2.3. The following least-squares solution contends with all of these cases.

For the static camera case, the unknown parameters, $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$, and $s_t$ ($t = 1 \ldots n$), are gathered into a length $n + 9$ vector:

$$\mathbf{u} = [p_0^x \; v^x \; a^x \; p_0^y \; v^y \; a^y \; p_0^z \; v^z \; a^z \; s_1 \cdots s_n]^\mathsf{T} \quad . \tag{7.24}$$

The constraints, Equations (7.13)-(7.15), are assembled into a linear system $\mathbf{Mu} = \mathbf{b}$ where:

$$\mathbf{M} = \begin{bmatrix}
1 & t_1 & t_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & c^x - q_1^x & 0 & \ldots & 0 \\
0 & 0 & 0 & 1 & t_1 & t_1^2 & 0 & 0 & 0 & c^y - q_1^y & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & t_1 & t_1^2 & c^z - q_1^z & 0 & \ldots & 0 \\
1 & t_2 & t_2^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c^x - q_2^x & \ldots & 0 \\
0 & 0 & 0 & 1 & t_2 & t_2^2 & 0 & 0 & 0 & 0 & c^y - q_2^y & \ldots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & t_2 & t_2^2 & 0 & c^z - q_2^z & \ldots & 0 \\
\vdots & & & & & & & \ddots & & & & & \vdots \\
1 & t_n & t_n^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & c^x - q_n^x \\
0 & 0 & 0 & 1 & t_n & t_n^2 & 0 & 0 & 0 & 0 & 0 & \ldots & c^y - q_n^y \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & t_n & t_n^2 & 0 & 0 & \ldots & c^z - q_n^z
\end{bmatrix}$$

and

$$\mathbf{b} = [c^x \; c^y \; c^z \; c^x \; c^y \; c^z \cdots c^x \; c^y \; c^z]^\mathsf{T} .$$

For the moving camera case, Equations (7.18)-(7.20), the $\mathbf{c} - \mathbf{q}_t$ in the rows of $\mathbf{M}$ are replaced with $\mathbf{R}_t\mathbf{c}_t - \mathbf{q}_t$), and $\mathbf{b}$ is replaced with:

$$\mathbf{b} = [c_1^x + t_1^x \;\; c_1^y + t_1^y \;\; c_1^z + t_1^z \ldots c_n^x + t_n^x \;\; c_n^y + t_n^y \;\; c_n^z + t_n^z]^\mathsf{T} .$$

If the optional size constraints from Section 7.2.5 are used, then $\mathbf{M}$ is extended by appending an additional $n - 1$ rows with a corresponding number of zeros appended to $\mathbf{b}$. For a static camera these rows have the form:

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{q_1^z - c^z}{q_2^z - c^z} & -\frac{\tilde{d}_2}{\tilde{d}_1} & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{q_1^z - c^z}{q_3^z - c^z} & 0 & -\frac{\tilde{d}_3}{\tilde{d}_1} & \ldots & 0 \\
\vdots & & & & & & & \ddots & & & & & & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{q_1^z - c^z}{q_n^z - c^z} & 0 & 0 & \ldots & -\frac{\tilde{d}_n}{\tilde{d}_1}
\end{bmatrix} .$$

For the moving camera case, the $\mathbf{q}_t - \mathbf{c}$ are correspondingly replaced with $\mathbf{R}_t\mathbf{q}_t - \mathbf{c}_t$).

The least-squares solution to the system of constraints is given by $\mathbf{u} = \mathbf{M}^+\mathbf{b}$, where $\mathbf{M}^+$ denotes pseudo-inverse computed as $\mathbf{M}^+ = \mathbf{VS}^{-1}\mathbf{U}^\mathsf{T}$, where $\mathbf{M} = \mathbf{USV}$ is the singular-value decomposition of $\mathbf{M}$.

Due to the possible scale ambiguity, this solution may not be unique. This occurs when the smallest-magnitude singular value in $\mathbf{S}$ is zero. Due to small amounts of measurement noise the smallest singular value may not be identically zero. We consider the smallest singular value to be zero if it is less than 0.1 times the second-smallest singular value.

If there is no zero singular value then $\mathbf{M}^{+}\mathbf{b}$ is our final solution. Otherwise the ambiguity is resolved by finding the solution where $\|\mathbf{a}\| = 9.8\text{m/s}^2$. Let $\mathbf{u}^*$ be the column of $\mathbf{V}$ corresponding to the zero singular value. The final solution is $\mathbf{u} = \mathbf{M}^{+}\mathbf{b} + \alpha\mathbf{u}^*$ where $\alpha$ is selected by solving the quadratic equation:

$$\left\| [0\ 0\ a^x\ 0\ 0\ a^y\ 0\ 0\ a^z\ 0 \cdots 0]^{\mathsf{T}} \cdot (\mathbf{M}^{+}\mathbf{b} + \alpha\mathbf{u}^*) \right\|^2 = 9.8^2 \quad . \tag{7.25}$$

In the case of a static camera, this quadratic constraint on the acceleration will give us the scaling that corresponds to its size in the physical world. More importantly it also avoids the trivial solution of $\mathbf{u} = \mathbf{0}$ when $\mathbf{b} = \mathbf{0}$.

In the case of a moving camera that follows a parabolic path, this method will produce a parabolic path for the projectile that matches gravity, but it corresponds to the physical path through space only if the path of the camera was estimated with the correct scale. Even if the scale of the camera's path were estimated incorrectly, the solution will still correspond to a projectile path which produces the same image and that is scaled in a way that allows consistent error comparison in meaningful units.

### 7.2.7 Forensics

The estimation of the projectile motion described in the previous two sections yields two parametrizations of the projectile path. The first is the initial position, velocity and acceleration ($\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$) as specified in Equation (7.1). The second is a non-parametric representation specified by the variables $s_t$.

For authentic video of a ballistic projectile, these representations should be in agreement, while an imperfectly faked projectile motion will yield inconsistencies in these representations. For the static camera case, we quantify the error in these representations as the average Euclidean distance between the pair of representations of the projectile motion:

$$E = \frac{1}{n}\sum_{t=1}^{n} \|\mathbf{p}_t - \mathbf{l}_t\| \tag{7.26}$$

$$= \frac{1}{n}\sum_{t=1}^{n} \|(\mathbf{p}_0 + \mathbf{v}t + \tfrac{1}{2}\mathbf{a}t^2) - (\mathbf{c} + s_t(\mathbf{q}_t - \mathbf{c}))\|, \tag{7.27}$$

where $\|\cdot\|$ denotes a vector 2-norm. Specifically, an error $E$ above a specified threshold is taken to be evidence of tampering. For the moving camera case, this error takes on a similar form:

$$E = \frac{1}{n}\sum_{t=1}^{n} \|(\mathbf{p}_0 + \mathbf{v}t + \tfrac{1}{2}\mathbf{a}t^2) - ((\mathbf{c}_t + \mathbf{t}_t) + s_t(\mathbf{R}_t\mathbf{q}_t - \mathbf{c}_t))\|. \tag{7.28}$$

This average error will be large when an overall path does not correspond to a ballistic trajectory. For situations where only small parts of an otherwise correct motion have

been altered, the maximum error may also be informative. Furthermore, as shown for example in Figs. 7.6-7.23, a visual comparison can be made of the difference between the parametric and non-parametric solutions.

## 7.3 Results

We evaluate the efficacy of our forensic technique on authentic and fake videos of our creation, and on several videos obtained from video-sharing websites such as YouTube. In each example, the optional size constraint described in Section 7.2.5 was used. In order to be able to directly compare the error metric, Equation (7.27) or (7.28), across different videos, the size constraint is weighted to yield the same average absolute difference, Equation (7.22). An iterative approach was taken to determine this weighting. On each iteration the linear system is solved, and the weighting is decreased until the difference reaches a specified threshold. Because the trajectory estimation is scaled so that the acceleration term is $9.8\text{m/s}^2$, Equation (7.25), the resulting error metric is specified in meters.

### 7.3.1 Tracking

The projection of a projectile's center of mass $\mathbf{q}_t$ on each video frame is estimated using a combination of manual and automatic tracking. A user first manually selects the approximate center of mass of the projectile on each frame. A user also places a bounding box around the projectile in the first frame, which is subsequently used as a template to refine the initial manually selected projectile locations. The bounding box is selected such that its center of mass corresponds to the projectile's center of mass. For each frame, the cross-correlation between the template and a small region around the user selected location is computed. The final projectile location is the position that maximizes the cross-correlation. The template is updated on each frame by automatically moving the selected bounding box to the next frame. This contends with the possible changing appearance and size of the projectile over time.

A specialized tracking algorithm is employed when the projectile is spherical (a basketball or soccer ball). Specifically, a circular Hough transform [37] is used to refine the location of the initially selected projectile location. The technique is based on a voting procedure carried out in the parameter space of the shape. A set of features points in the image space (typically edges previously extracted with commonly used edge filter detectors, such as Canny or Sobel) are transformed in a set of accumulated votes in the parameter space. Local maxima in the accumulator array identify candidate objects, thus defining circle position and radius.

### 7.3.2 Simulations

We simulated ballistic trajectories of a projectile with varying initial positions relative to a fixed and stationary camera, varying velocities, small amounts of additive noise, and varying planes of motion relative to the camera. These trajectories had known world and image positions, so no tracking algorithm was required. In addition, the projectile was simulated as a point, so the optional size constraint was not used.

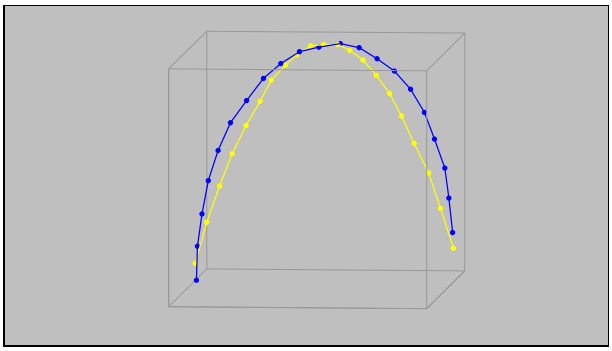


Figure 7.3: An authentic (yellow) and a visually plausible fake (blue) trajectories generated for simulations.

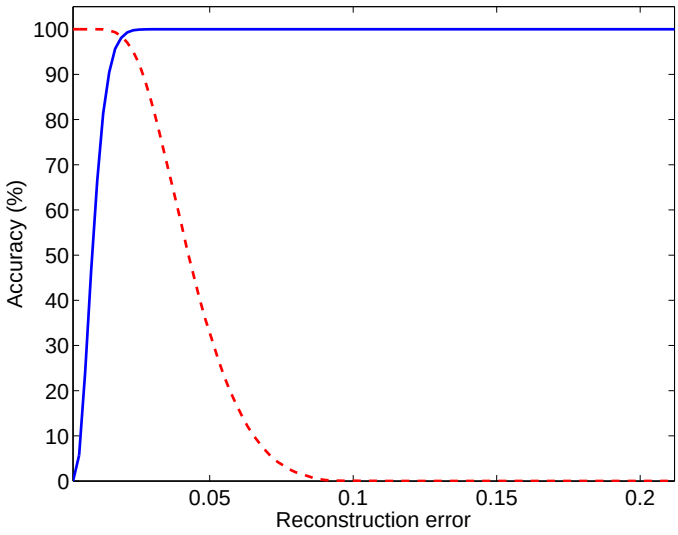



Figure 7.4: Shown is ROC curve for classification. The solid curve corresponds to the authentic trajectories, and the dashed curve corresponds to the inauthentic paths.

Ten thousand random trajectories were generated from which the error, Equation (7.27), between the estimated parametric and non-parametric parametrizations was computed. The mean error for the authentic ballistic trajectories is 0.0096 with a standard deviation of 0.0039.

Non-ballistic, yet visually plausible, trajectories were created by sampling a path along an ellipse. Ten thousand random trajectories were created. The mean error for the elliptical trajectories is 0.0446 with a standard deviation of 0.0150. Shown in Fig. 7.3 are two examples of simulated trajectories: in yellow an authentic projectile while in blue a visually plausible fake trajectory.

| | Fig | $\mu$ (m) | $\sigma$ (m) | max (m) |
|---|---|---|---|---|
| Authentic | 7.6 | 0.014 | 0.011 | 0.044 |
| | 7.7 | 0.029 | 0.021 | 0.093 |
| | 7.8 | 0.010 | 0.009 | 0.033 |
| | 7.9 | 0.038 | 0.029 | 0.121 |
| Fake | 7.10 | 0.151 | 0.093 | 0.382 |
| | 7.11 | 0.307 | 0.114 | 0.750 |
| | 7.12 | 0.175 | 0.077 | 0.420 |
| | 7.13 | 0.357 | 0.34 | 1.872 |

Table 7.1: The error between the parametric and non-parametric parametrizations of the estimated projectile trajectory for a static camera (Figs. 7.6-7.9 and 7.10-7.13), Equation (7.27). Shown are the mean $\mu$, standard deviation $\sigma$, and maximum error, specified in meters. The mean and maximum error for the fake videos are consistently and significantly higher than for the authentic videos.

Shown in Fig. 7.4 is the resulting ROC curve where the horizontal axis corresponds to the reconstruction error, and the vertical axis to the classification accuracy. The solid curve corresponds to the authentic trajectories, and the dashed curve corresponds to the inauthentic paths. 99% of the parabolic trajectories have an error less than 0.02, while 98% of the elliptical trajectories have an error greater than 0.02, revealing good discrimination.

### 7.3.3 Static Camera

Images in Fig. 7.6-7.9 show results for four authentic videos. The examples in Fig. 7.6 and Fig. 7.7 are of our recording, and the examples in Fig. 7.8 and 7.9 are from YouTube. The images in panels (a) show four frames taken from each of the video sequences. Panel (b) shows the two parametrizations of the estimated trajectory: the parametric trajectory specified by initial position $\mathbf{p}_0$, velocity $\mathbf{v}$, and acceleration $\mathbf{a}$ is denoted with filled yellow dots and solid line, and locations along the non-parametric trajectory specified by the variables $s_t$ are denoted with open blue circles. The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Panel (c) shows the estimated parametric trajectory, denoted with filled yellow dots, projected into the image plane, and the tracked position of the projectile denoted with small red dots. In each example, the different parametrizations of the projectile are in agreement, as would be expected for an authentic video. The mean errors from Equation (7.27) for these three sequences are respectively 0.014m, 0.029m, 0.010m, and 0.038m, Table 7.1.

Note that in Fig. 7.6 the ball temporarily disappears from the field of view. Regardless, we can still set up the linear system, skipping frames where the ball cannot be observed,

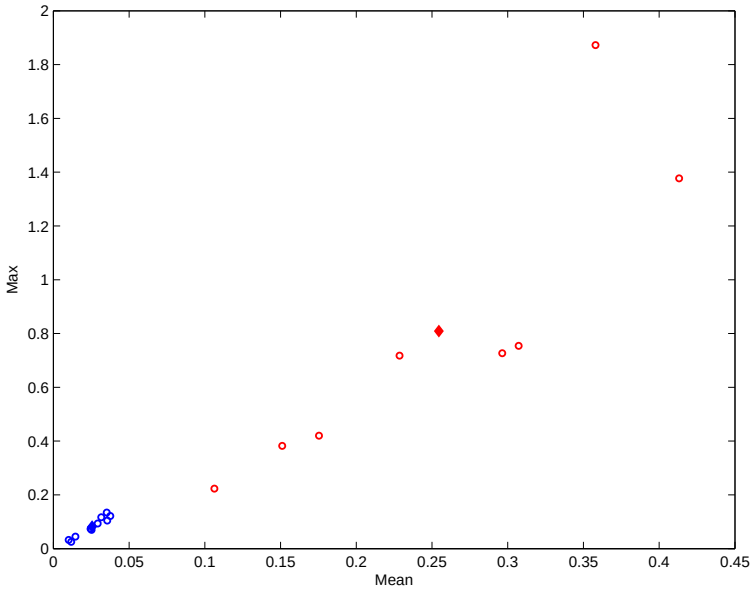
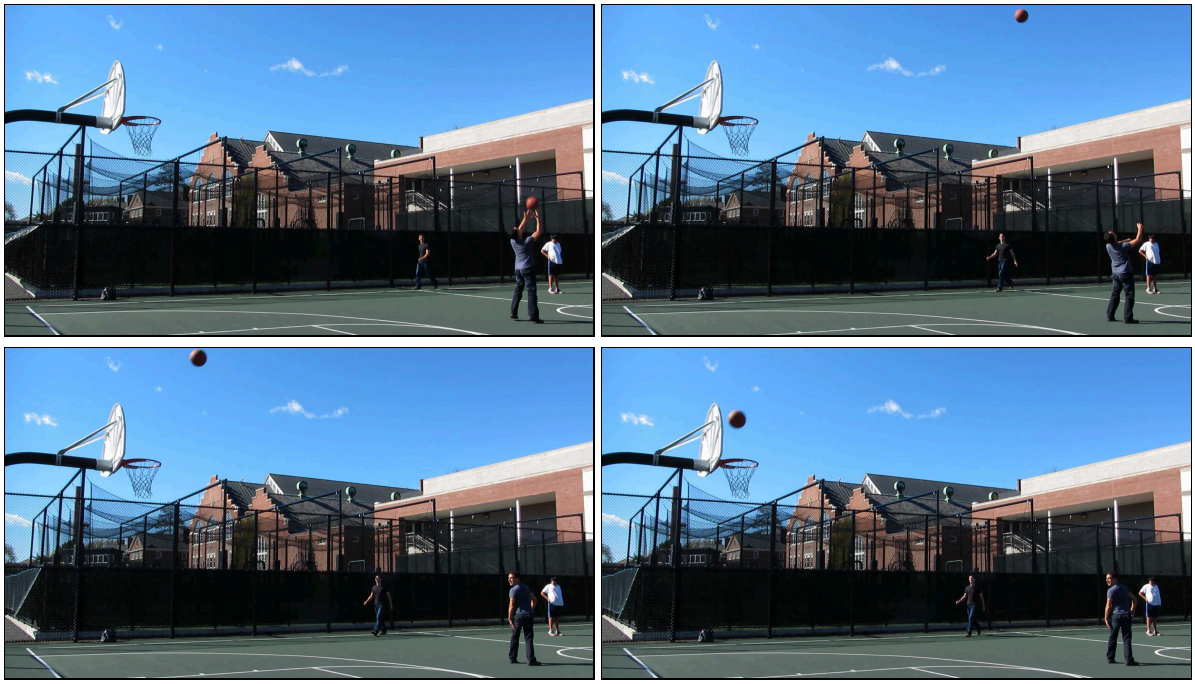



Figure 7.5: Error metric for authentic (blue) and fake (red) videos from a static camera, where the horizontal axis corresponds to the mean reconstruction error, and the vertical axis to the maximum error. The diamond identifies the mean value over all the authentic (blue) and fake videos (red).

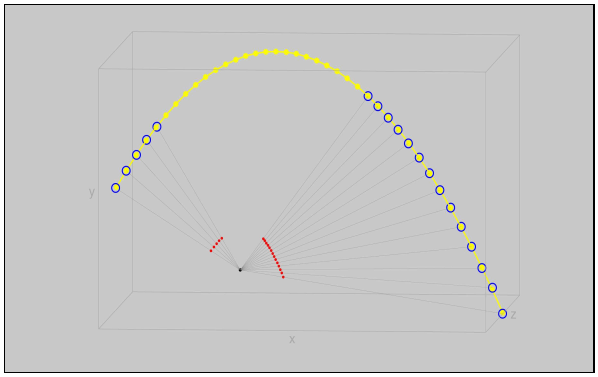
because the time-coding of the video establishes a consistent time parametrization. Once we have computed $\mathbf{p}_0$, $\mathbf{v}$, and $\mathbf{a}$, the full trajectory can then be estimated by extrapolating the ball's position during the time it is out of the field of view.

Images in Fig. 7.10-7.13 show results from four fake videos. The example in Fig. 7.10 is of our creation, the example inFig. 7.11 is from YouTube, and the examples in Fig. 7.12 and 7.13 are of a pendulum motion (a ball attached to a string). In each case, the different parametrizations are not in agreement, revealing these videos to be fake, or in the case of the pendulum motion, inconsistent with a ballistic motion. The mean errors from Equation (7.27) for these four sequences are 0.151m, 0.307m, 0.175m and 0.357m, Table 7.1. These errors are, on average, almost an order of magnitude larger than for the authentic videos in Fig. 7.6-7.9.

A total of ten authentic and eight fake videos with static cameras were analyzed, a subset of which are shown in Figs. 7.6-7.9 and 7.10-7.13. The mean error averaged over all the authentic videos is 0.026m and the largest mean error for any of the authentic videos is 0.038m. The mean error averaged over the fake videos is an order of magnitude larger at 0.255m, and the smallest of the mean errors for any of the fake videos is 0.106m. These errors show that authentic and fake videos are well separated in terms of the error metric used to assess authenticity, Fig. 7.5. There is also a very clear qualitative distinction that can be seen when comparing the trajectory visualizations in panels (c) of Figs. 7.6-7.9 and 7.10-7.13.



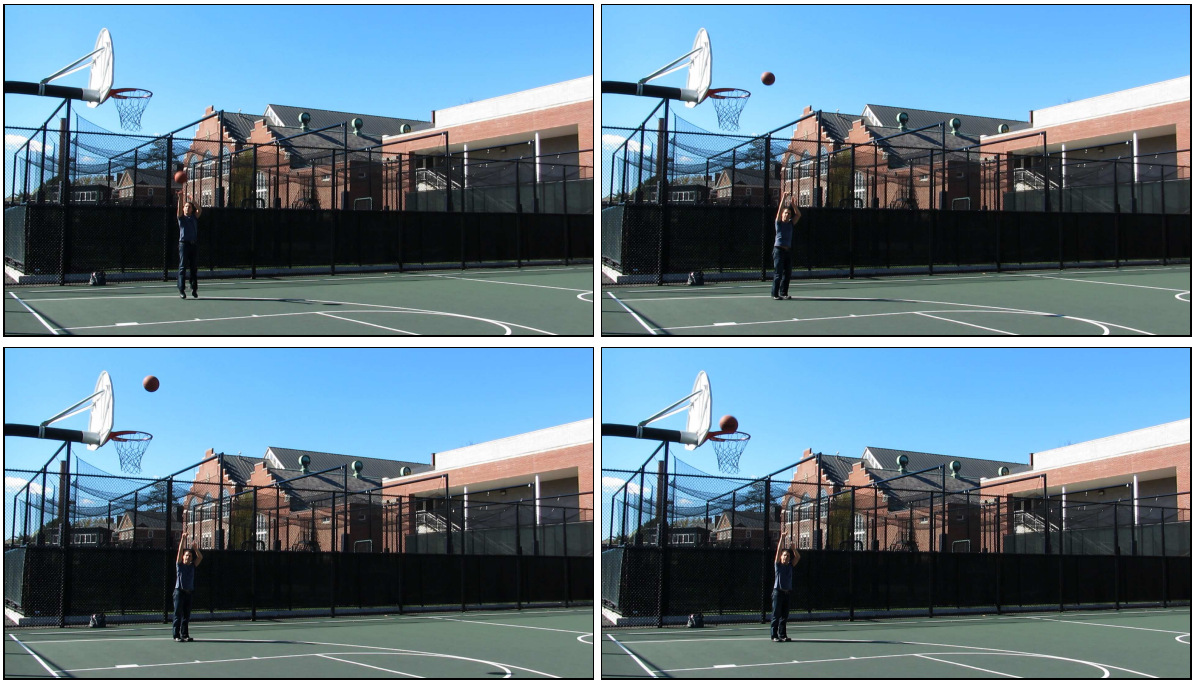
(a) *http://www.youtube.com/watch?v=z7NpkDCDYjg*



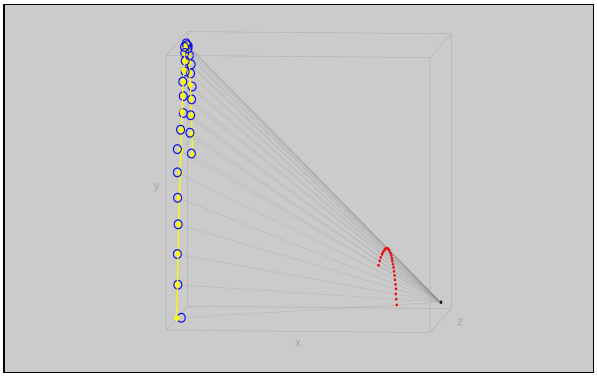
(b)

(c)

Figure 7.6: Authentic video from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.



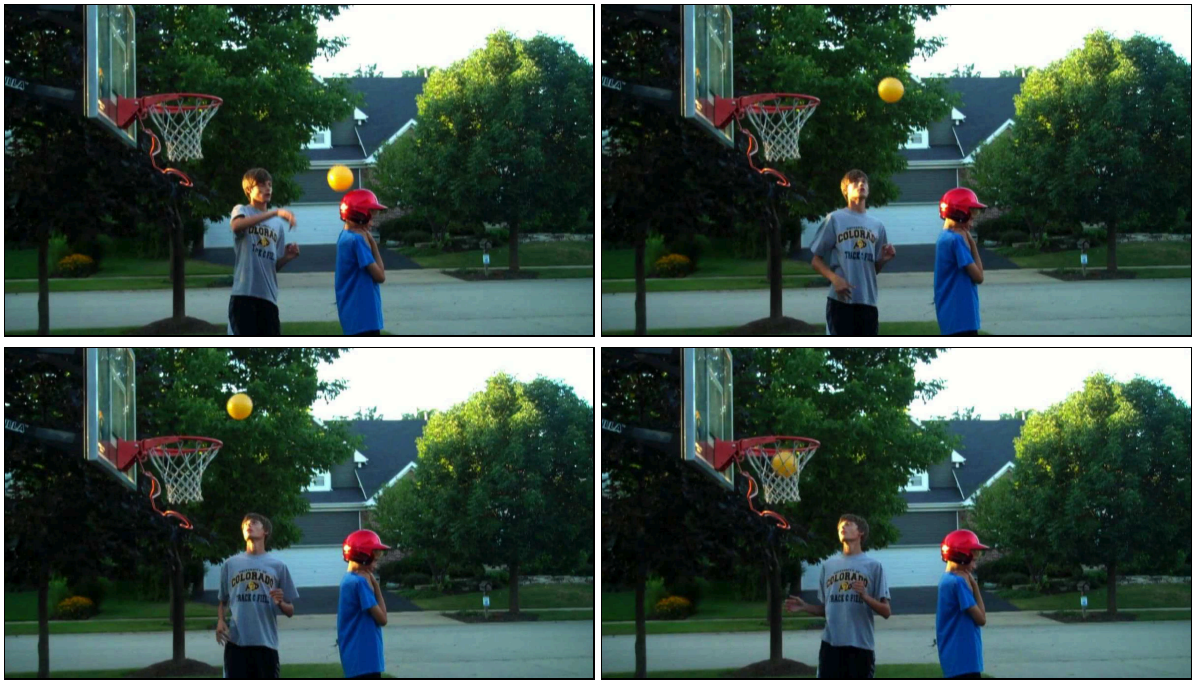
(a)*http://www.youtube.com/watch?v=GbKA7t79Wds*



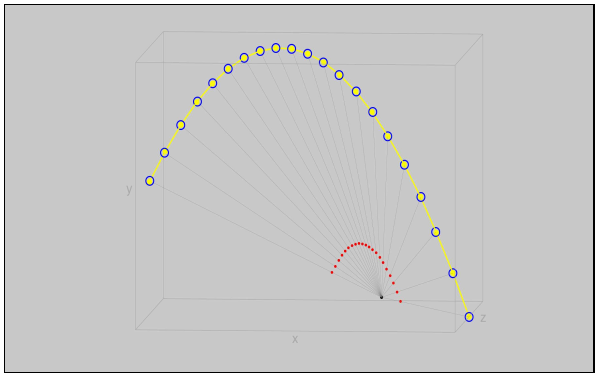
(b)

(c)

Figure 7.7: Authentic video from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.



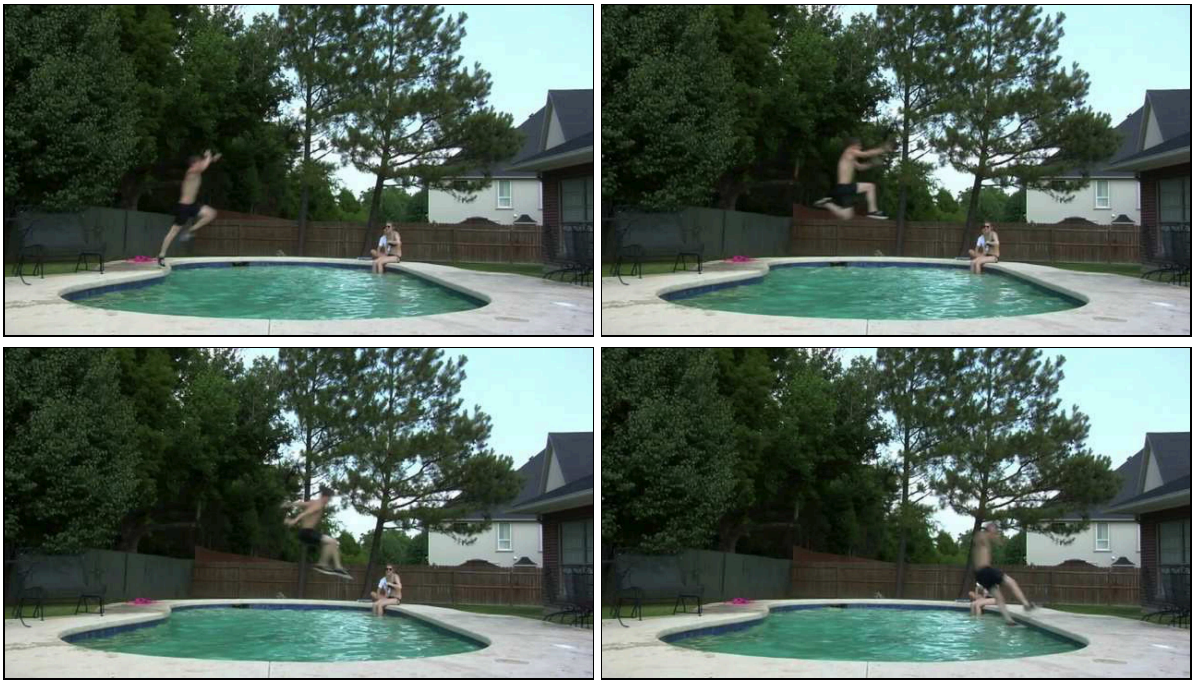
(a) http://www.youtube.com/watch?v=kDUHTioJi7A



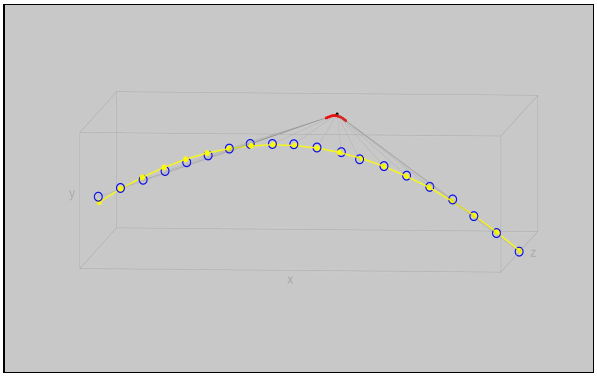
(b)

(c)

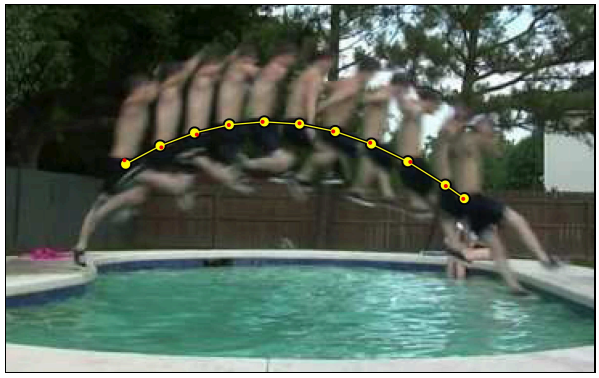
Figure 7.8: Authentic video from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.



(a) http://www.youtube.com/watch?v=wpYM3NkMYMs

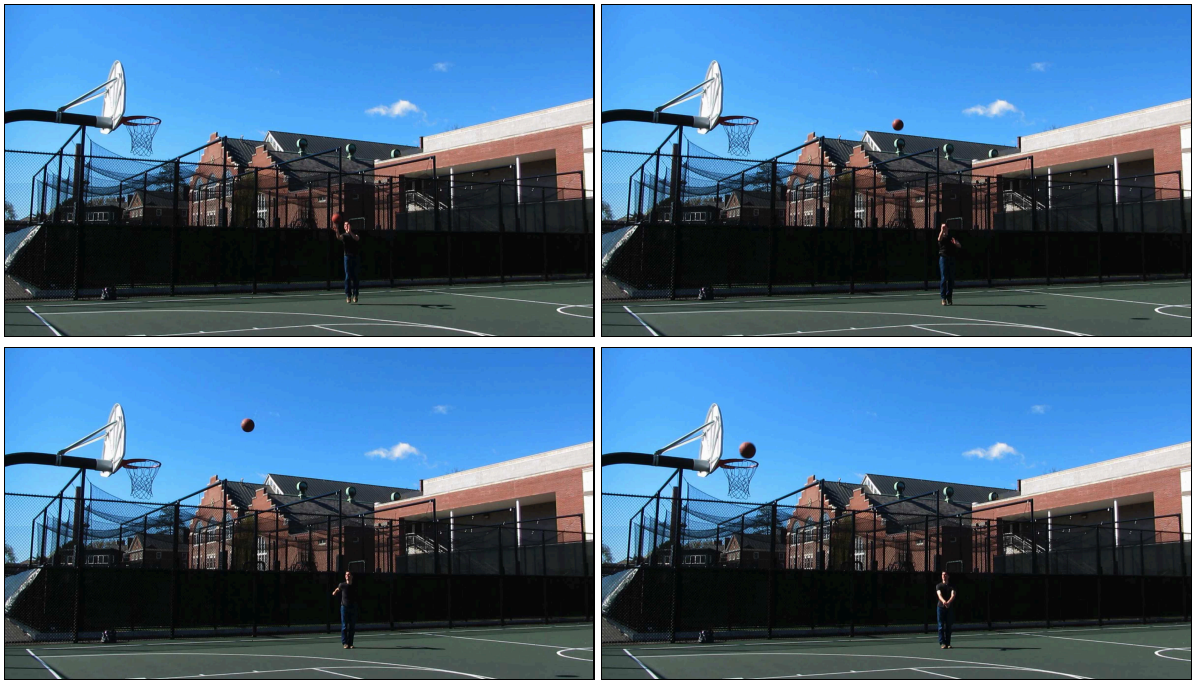


(b)

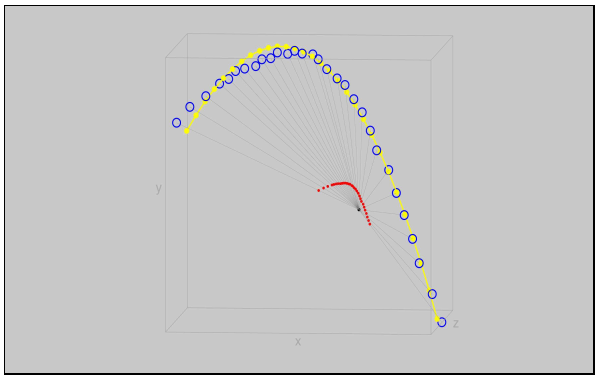


(c)

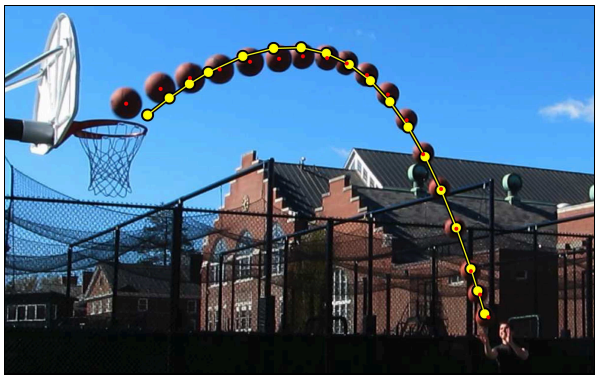
Figure 7.9: Authentic video from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.



(a) http://www.youtube.com/watch?v=RLJ7aNkIv7I

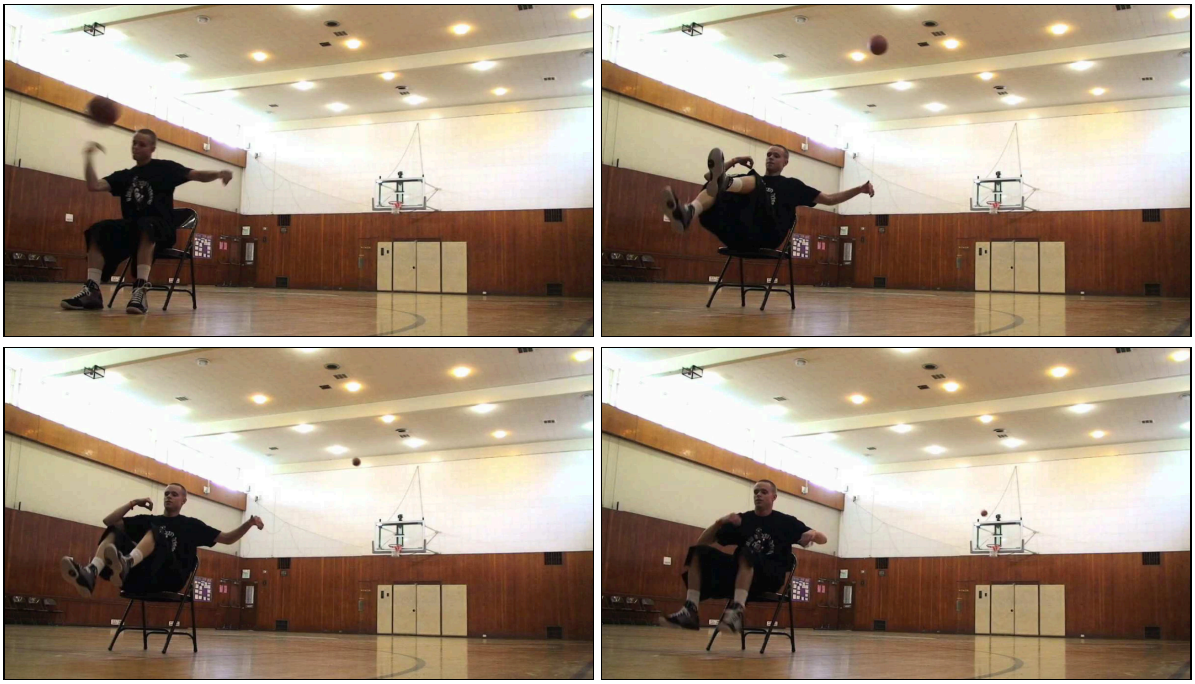


(b)

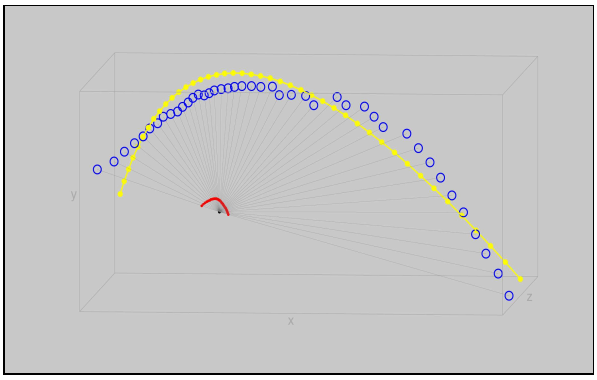


(c)

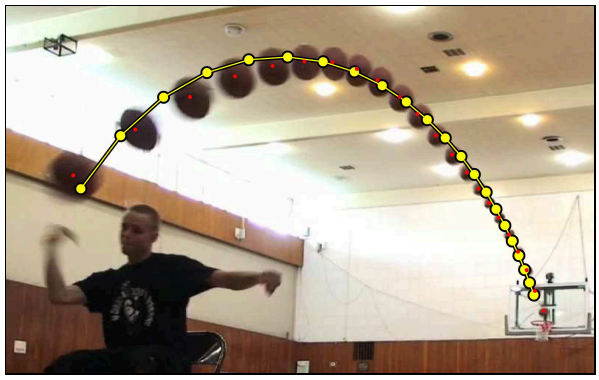
Figure 7.10: Fake videos from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.



(a) http://www.youtube.com/watch?v=WbaH52JI3So



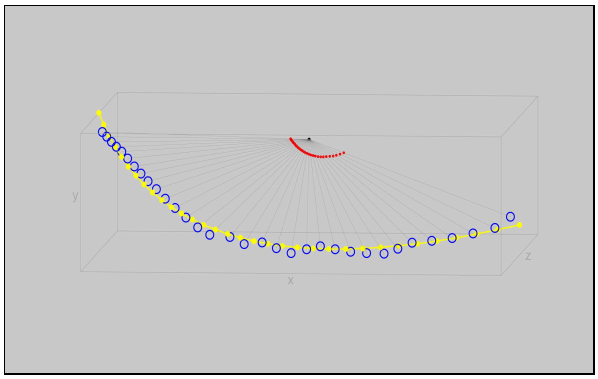
(b)



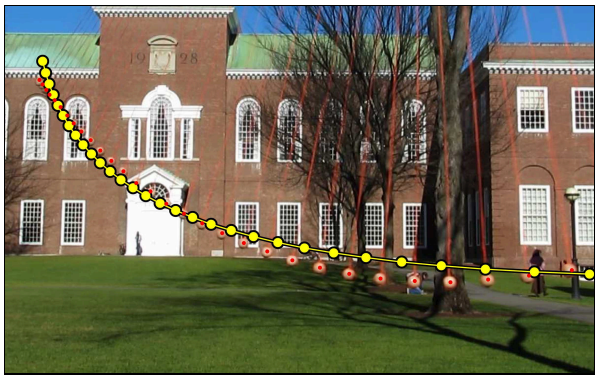
(c)

Figure 7.11: Fake videos from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

(a) http://www.youtube.com/watch?v=ucGLQ8wzxNM



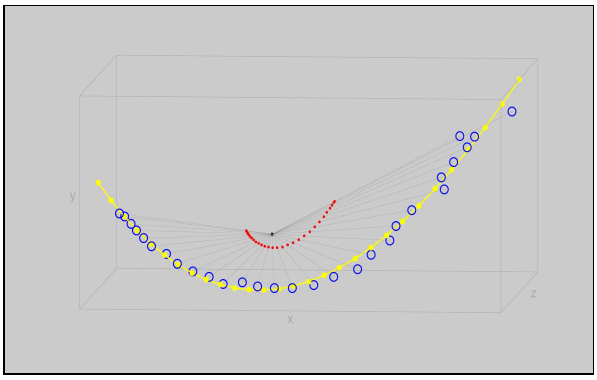
(b)



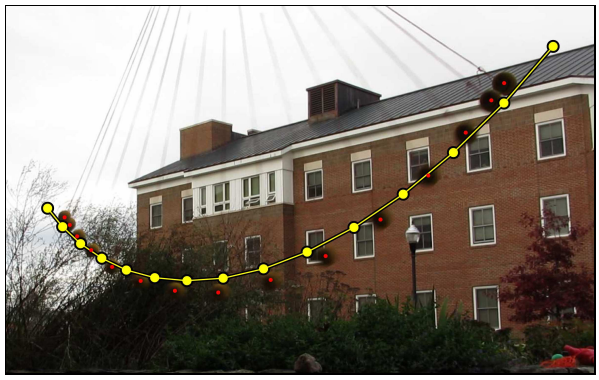
(c)

Figure 7.12: Fake videos from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

(a) http://www.youtube.com/watch?v=ChawQ_3CjbI



(b)



(c)

Figure 7.13: Fake videos from a static camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

| | Fig | $\mu$ (m) | $\sigma$ (m) | max (m) |
|---|---|---|---|---|
| Authentic | 7.16 | 0.031 | 0.017 | 0.096 |
| | 7.16 | 0.031 | 0.017 | 0.096 |
| | 7.17 | 0.056 | 0.035 | 0.162 |
| | 7.18 | 0.024 | 0.016 | 0.077 |
| Fake | 7.20 | 0.082 | 0.072 | 0.341 |
| | 7.21 | 0.119 | 0.064 | 0.365 |
| | 7.22 | 0.266 | 0.173 | 0.934 |
| | 7.23 | 0.166 | 0.091 | 0.557 |

Table 7.2: The error between the parametric and non-parametric parametrizations of the estimated projectile trajectory for a moving camera (Figs. 7.15-7.18 and 7.20-7.23), Equation (7.28). Shown are the mean $\mu$, standard deviation $\sigma$, and maximum error, specified in meters. The mean and maximum error for the fake videos are consistently and significantly higher than for the authentic videos.

### 7.3.4 Moving Camera

Images in Fig. 7.15-7.18 show results from authentic videos recorded with moving cameras. The example in Fig. 7.15 and 7.16 are of our recording, and the example in Fig.7.17 and 7.18 are two parts of one video from YouTube. Panels (a) show four video frames from each motion sequence. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by initial position $\mathbf{p}_0$, velocity $\mathbf{v}$, and acceleration $\mathbf{a}$ is denoted with filled yellow dots and solid line, and the non-parametric trajectory specified by the variables $s_t$ is denoted with open blue circles. The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Shown in panel (c) is the estimated parametric trajectory, denoted with filled yellow dots, projected into the image plane, and the tracked position of the projectile denoted with small red dots. The underlying image in panel (c) was constructed by compositing the video frames into a single composite. In each example the different parametrizations of the projectile are in agreement, as would be expected for an authentic video. The mean error, Equation (7.28), for these four sequences are 0.027m, 0.0315m, 0.056m, and 0.024m, Table 7.2. Moreover, in Fig. 7.19 the composition of all the sequence represented in Fig. 7.17 and 7.18 is shown.

Images in Fig. 7.20-7.23 show results from four fake videos taken with moving cameras. The example in Fig. 7.20 is of our creation, the example in Fig. 7.21 is from YouTube, and the examples in Fig. 7.22 and 7.23 is of a pendulum motion (a ball attached to a string). In each case, the different parametrizations are not in agreement, revealing these videos to be fake, or in the case of the pendulum motion, inconsistent with a ballistic motion. The error, Equation (7.27), for these four sequences are 0.082m, 0.119m, 0.266m and 0.161m, Table 7.2. These errors are, on average, four times larger than for the authentic videos in

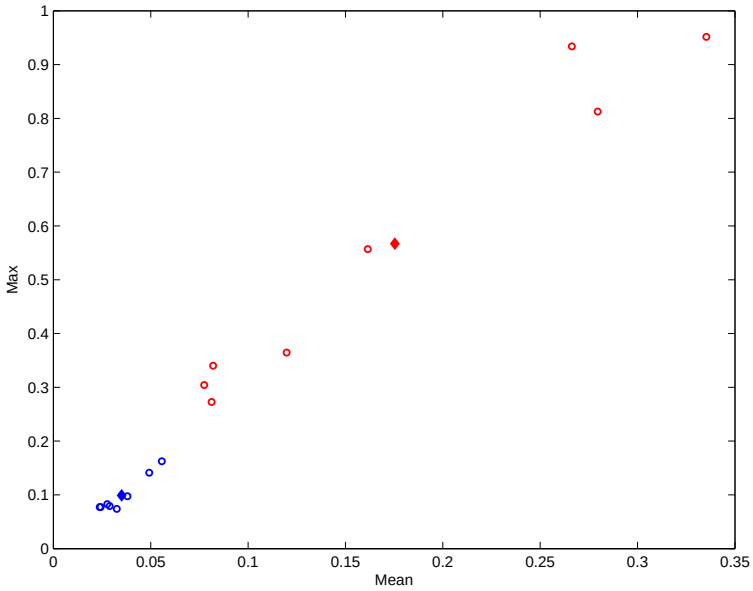
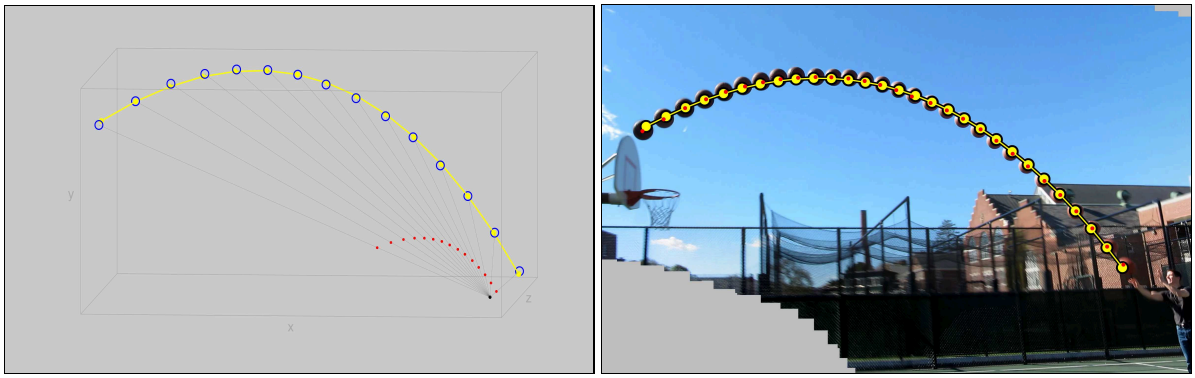



Figure 7.14: Error metric for authentic (blue) and fake (red) videos from a moving camera, where the horizontal axis corresponds to the mean reconstruction error, and the vertical axis to the maximum error. The diamond identifies the mean over over all the authentic (blue) and fake videos (red).

Fig. 7.15-7.18.

A total of eight authentic and eight fake videos with a moving camera were analyzed, a subset of which are shown in Figs. 7.15-7.18 and 7.20-7.23. The mean error averaged over the authentic videos is 0.035m and the largest mean error for any of the authentic videos is 0.056m. The mean error for the fake videos is several times larger at 0.175m, and the smallest of the mean errors for any of the fake videos is 0.077m. These errors show that the authentic and fake videos are well separated in terms of the error metric used to assess authenticity, Fig. 7.14. As with the static camera examples, there is also a very clear qualitative distinction that can be seen when comparing the trajectory visualizations.

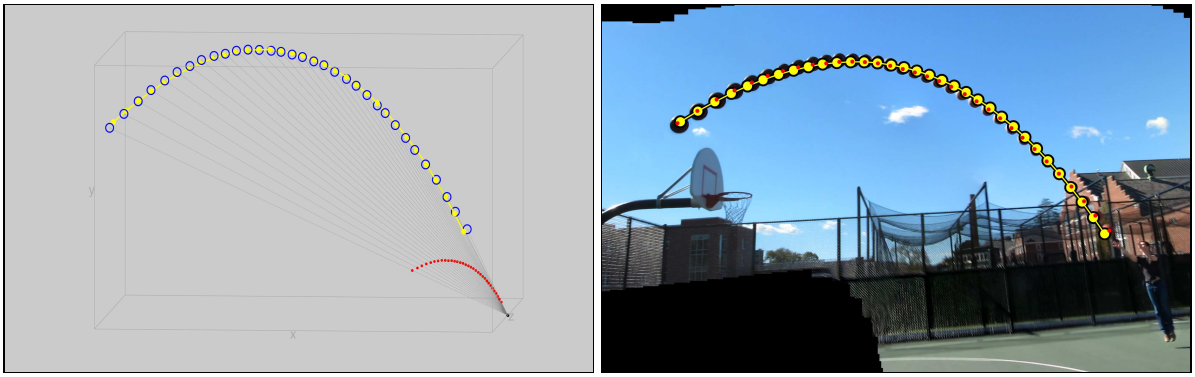
(a) http://www.youtube.com/watch?v=4n8B1jWjsq4



(b) (c)

Figure 7.15: Authentic video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.

(a) http://www.youtube.com/watch?v=c85qoQ8K6hQ



(b) (c)

Figure 7.16: Authentic video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.

(a) http://www.youtube.com/watch?v=2JwbubAX-Aw

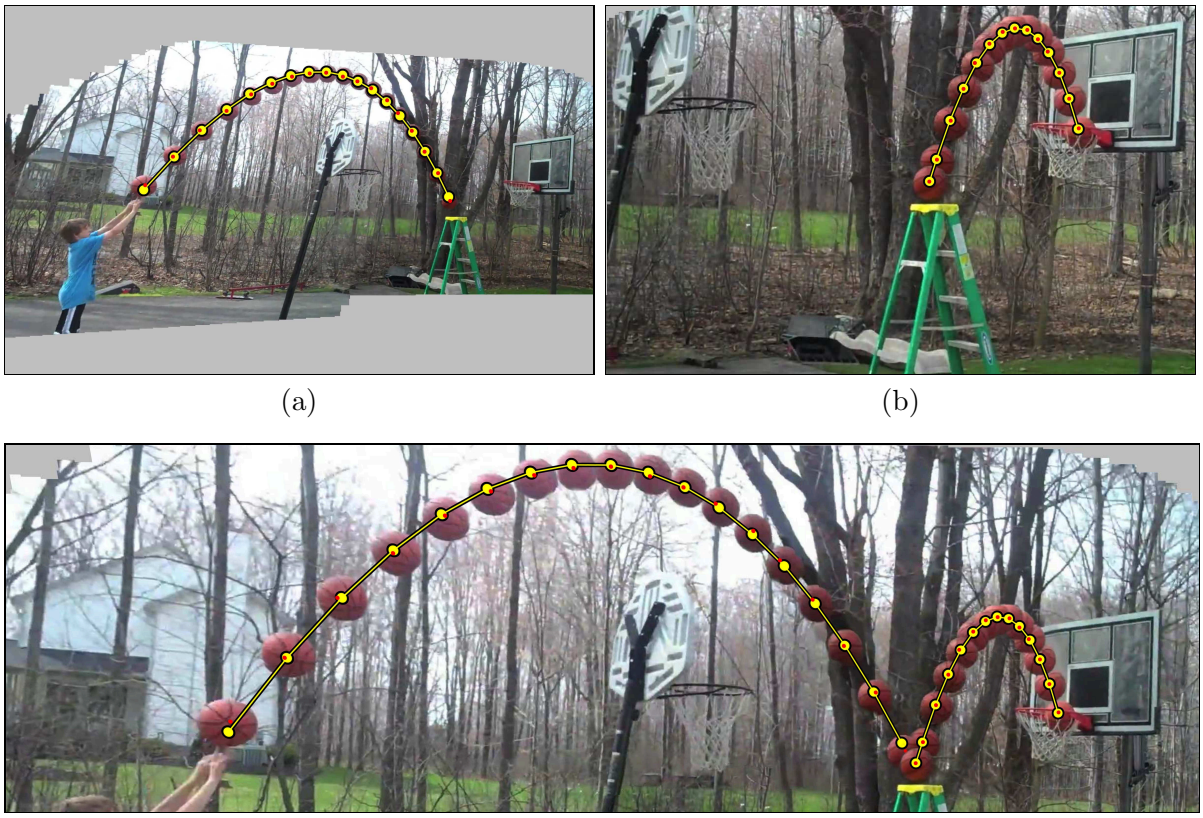
(b) (c)

Figure 7.17: Authentic video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.

(a) http://www.youtube.com/watch?v=2JwbubAX-Aw

(b) (c)

Figure 7.18: Authentic video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are in agreement, as expected for an authentic video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are in agreement, as expected for an authentic video.

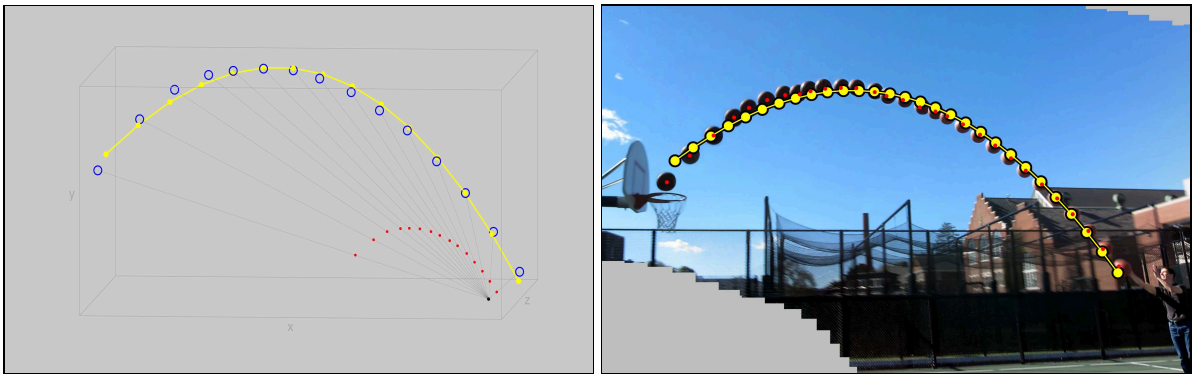


(a) (b)

(c)

Figure 7.19: Composite of the video frames of fig 7.17 (a) and 7.18 (b) into a single panorama (c).

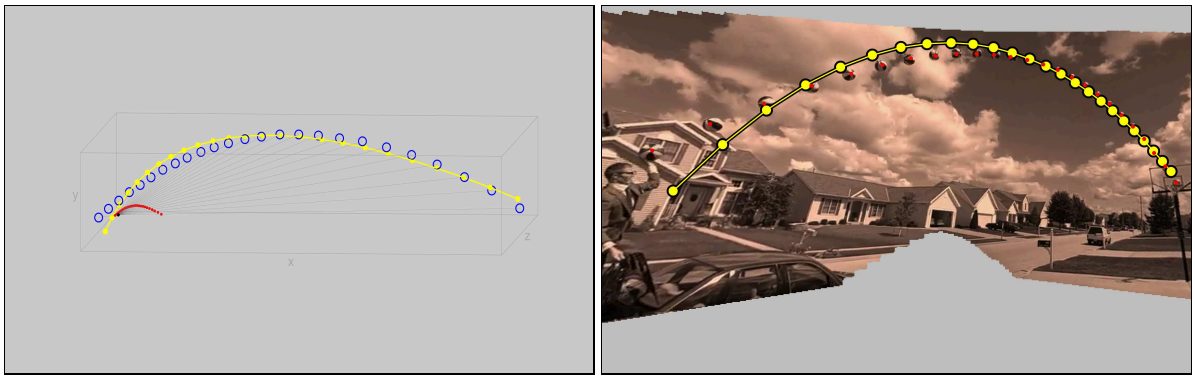
(a) http://www.youtube.com/watch?v=3v93eqwwkV0



(b) (c)

Figure 7.20: Fake video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

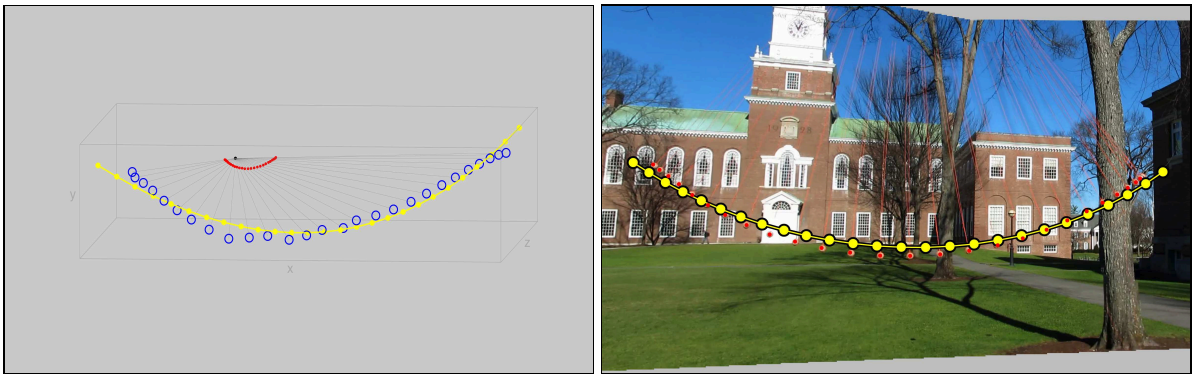
(a) http://www.youtube.com/watch?v=EXPOfStEllA



(b) (c)

Figure 7.21: Fake video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

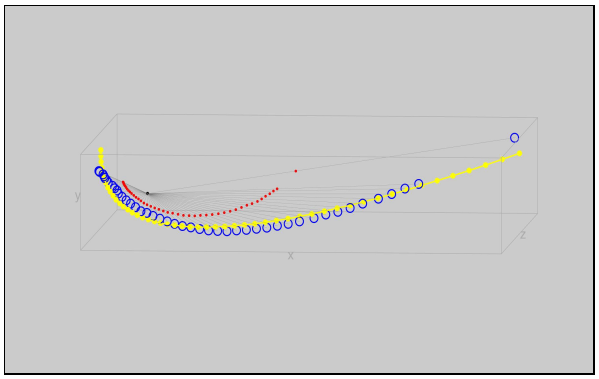
(a) http://www.youtube.com/watch?v=PumYuLlHDoc



(b) (c)

Figure 7.22: Fake video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

(a) *http://www.youtube.com/watch?v=TaXOmRzBvIg*



(b)

(c)

Figure 7.23: Fake video from a moving camera. Shown in panel (a) are four sample video frames. Shown in panel (b) are the two parametrizations of the estimated trajectory: the parametric trajectory specified by $\mathbf{p}_0$, $\mathbf{v}$, $\mathbf{a}$ (filled yellow dots and solid line), and the non-parametric trajectory specified by the variables $s_t$ (open blue circles). The small black dot corresponds to the camera center and the small red dots correspond to the projection of the projectile in each video frame. Note that the two parametrizations are not in agreement, as expected for a fake video. Shown in panel (c) is the estimated parametric trajectory projected into the image plane (filled yellow dots) and the tracked position of the projectile (small red dots). These locations are not in agreement, as expected for a fake video.

## 7.4 Discussion

We have described a geometric technique for detecting forged ballistic motion in video sequences. This technique explicitly models the three-dimensional trajectory of objects in free-flight and the two-dimensional imaging of the trajectory by a static or moving camera. We have shown that the three-dimensional trajectory can be directly and reliably estimated from a video sequence. Deviations from this model provide evidence of manipulation. This forensic analysis makes minimal assumptions, requires limited user input, and is computationally efficient. We have verified the efficacy of this technique in large-scale simulations, and on numerous real-world examples.

Although somewhat narrowly applicable, this forensic tool adds to a growing set of techniques for authenticating images and video. One of the advantages of geometric techniques over techniques based on low-level image statistics is that the modeling and estimation of geometry is less sensitive to resolution and compression that can easily confound the statistical analysis of images and video.

As with any forensic technique, it is important to consider the ease with which counter-measures can circumvent our forensic analysis of ballistic motion. Because our analysis considers the three-dimensional trajectory of a projectile, a two-dimensional image-based manipulation is unlikely to correctly generate a physically plausible three-dimensional trajectory. This is made even more unlikely when the video is filmed with a moving camera, since there is no frame of reference from which a forger can specify the overall trajectory. A determined forger could, of course, compute a desired 3-D parabolic trajectory, project it under the appropriate homography, and then alter a projectile's path accordingly. However, such a trajectory may not always exist for scenarios that a forger wishes to depict. In particular constraints, such as tying on to an initial segment of motion or connecting specific targets in a scene, can create situation where no physically possible ballistic motion wold meet the forgers requirements. Furthermore, for a moving camera the forger would have to first align all of the frames into a common coordinate system and project edits back into the original video sequence. This counter-measure, while not impossible, is currently outside of the capabilities of available video-editing software.

While we have focused only on ballistic motions, the basic framework presented here could be applied to analyzing the motion of cars, planes, rockets, or other moving objects. This would, of course, require more sophisticated models of motion to account for the particular type of accelerations experienced by these objects.

# Chapter 8

# Conclusions

The pervasive availability of the Internet, coupled with the development of affordable and high quality technologies, such as digital cameras and computers, has lead digital multimedia contents to be the primary source of visual information in several scenarios, replacing traditional film-based media. The digitalization of such contents has brought significant technical and economical benefits, but also problematic issues regarding their authenticity due to the fact that sophisticated and user-friendly graphics editing softwares have enabled forgers to easily manipulate digital images, or videos. Their trustworthiness is thus notably reduced and their reliability as a true representation of reality cannot be taken for granted, especially within those environment dealing with sensitive data, such as forensic investigation, science, medical imagery and press photographs. The objective of this thesis is the development of efficient techniques to cope with security issues related to multimedia data.

Digital watermarking has been firstly proposed as a valuable mean to cope with these problems, by imperceptibly embedding a message into a documents. Such message can later be detected and/or retrieved and used to disclose possible copyrights violations or manipulations. Since malevolent users may be willing to impair such methodology attempting to remove the embedded information, generally watermarks are required to be robust to resist to possible attacks, especially for specific application such as copyright protection. In Chapter 3 we presented an innovative benchmarking tool designed to evaluate robustness performances of image watermarking techniques under any combination of attacks, selected depending on the context requirements the method is designed for. From a developers point of view, a benchmark allows verifying the robustness of the own method with respect to existing algorithms, thus stimulating the research to look for better and better solutions. On the other hand, from a users point of view, evaluating performances of the plethora of watermarking techniques present in the literature is useful in order to select the most appropriate one for the intended application.

Although the main application of digital watermarking is copyright protection, this technology can be used for different purposes, such as multimedia indexing, authentication, fingerprinting, integrity verification, quality assessment and the like. In particular, applications dealing with sensitive imagery, such as medical, forensics or military, primarily require preserving the integrity of the data, rather than satisfying the robustness requirement. To cope with this issue, reversible watermarking schemes have been investigated. Their peculiarity is the possibility to restore the original host data after the detection or the extraction of the hidden data. Chapter 4 described the novel reversible algorithm we proposed. In particular, the presented technique extends existing prediction-based schemes, by exploiting local dependency, with non-local similarity information. Experimental results demonstrate that it successfully allow a complete recovery of the original host signal and over-performs state-of-the art algorithms in terms of embedding capacity, quality of the watermarked image and computational complexity.

The main limitation of digital watermarking is that it requires some information to be embedded at the time of recording or a person to embed it at the time of sending. Up to now, such processes are not standardized and the majority of contents available over the network is not protected, thus being subjected to unauthorized use and manipulation. In light of this, recently the scientific community started searching for new techniques able to verify integrity of digital data in absence of any watermark or special hardware. The solution has been found in passive forensics approaches. The basic idea is that the manipulation of a digital media, if performed properly, may not leave visual trace of its occurrence, but it alters the underlying statistics of the content. An accurate analysis can be carried out, without any prior knowledge about the content and reveal such alterations which can be taken as evidence of forgery. In this doctoral study we attempted to actively contribute to the research in the field, proposing a couple of innovative and efficient techniques for both image and video authentication. We focused our attention geometric-based forensic techniques, which exploit the principles of projection geometry of image formation. Such principles are very likely to be violated when creating a forgery, therefore their analysis can lead to evidence of tampering.

Chapter 4 presented the details of our forensic approach to authenticate photographs depicting text on sign and billboards. Because it is relatively easy to digitally insert text into a photo in a visually compelling manner, it can be difficult to determine if text is authentic just at visual inspection. Our forensic technique explicitly models the projection from the sign in the world to the image and determines if this projection satisfies the expected planar perspective mapping. We demonstrated that inauthentic text often violates the rules of perspective projection and can therefore be detected.

As far as video authentication is concerned, in Chapter 7 we reported details of the forensic technique we developed for determining if a projectile (e.g., a thrown ball) in a video is authentic. We explicitly described a plausible physical model for the expected trajectory of a projectile motion, and a basic imaging model for a static or moving camera. We then explained a technique to determine if the image of the trajectory of a projectile motion is consistent with this physical model.

Given the growing number of digital content available over the network, thanks to file sharing website and social networks, there is a constantly growing need for digital forensics techniques, and many promising and innovative techniques have been proposed so far to address authentication issues. However, together with all the efforts to disclose photographic frauds, also new techniques to create better forgeries that cannot be detected have been developed by determined forgers. In light of this, the trustworthiness of digital forensics have been recently questioned [51]. Anti-forensics is a new brand science which aims at identifying weaknesses in existing forensic techniques. To the best of our knowledge, the first paper dealing in this research area has been published in 2008 [79]. The authors demonstrate how traces left by re-sampling operations can be made undetectable by post-processing the content with a median filter. Little work has been done in the field up to now, [80, 147, 148, 14], but it is to be hoped that research in this direction will increase. As a matter of fact, even if someone may think that the development of anti-forensics techniques undermines the security of multimedia data, such studies may results very useful for researchers, pushing them towards improved techniques, able of both overcoming disclosed drawbacks and/or detecting when an anti-forensic operation has been used [53, 81].

Similarly to cryptography, multimedia content authentication, provided by both active and passive forensics, is a "cat-and-mouse" game, where advantage constantly shifts between researchers and forgers. Anyway, it is important to keep playing this game in order to increase the level of expertise required to violate and manipulate digital contents in a way that cannot be detected and to develop better and better solutions.

A challenge in the future of image forensics will be its application to multimedia opinion mining and bias analysis. In particular, the modification of a content may lead to an critical impact of its meaning, maybe altering the opinion of the viewer relative to the represented event. Understanding the perception of visual semantics could be important also to understand the maliciousness of a forgery. The work proposed in [32] represents the first step in this direction, proposing a theoretical framework to establish semantic dependencies among group of images. The work in this field is still in its infancy but the scientific community is recently focusing on this challenging issue [1].

# Acknowledgments

I barely believe I am at the end of my PhD and this is a real accomplishment for me. This could not have been possible without the precious support and valuable collaborations of all the people who believed in me and guided me over the past few years.

Firstly, I would like to deeply thank my advisor prof. Giulia Boato, for taking me as her first PhD student and all the challenges that this implied. Her priceless support, understanding and kindness added considerably to my graduate experience. She always provided me with complete encouragement, assistance and trust over the years, building a deep and respectful relationship, also from a personal point of view. I barely can express all my appreciation, and I owe her my most sincere gratitude.

A very special thanks goes to my co-advisor prof. Hany Farid, who truly made the difference in my PhD and my life. It is thanks to him if I have found a field of research that charms me and that I love. I had the priceless opportunity to work with him, within his group at Dartmouth College, and I have been learning a lot from him, both on a personal and professional level. He transmitted to me his love and commitment to work, highly motivating and encouraging me during my graduate experience. I thank him for his patience guidance, technical support, understanding and always positive attitude, and also for the wonderful experience of life I had living in the States. I doubt I will ever be able to fully convey my gratitude and appreciation, simply I owe him my eternal thankfulness.

I would like to warmly thank prof. Karen Egiazarian, for his valuable advice, guidance and technical support. I truly appreciate his always positive attitude and his precious teachings about the importance of hard work and love for what I do.

I would like to express my gratitude also to prof. Francesco De Natale, for his valuable advice and guidance of the multimedia group and understanding lab, to prof. Claudio Fontanari, whose valuable suggestions and collaboration have been fundamental, to prof. Marco Carli, for his important advices and collaboration and always friendly help, and to prof. James O'Brien, for his precious collaboration.

A warm thank goes to prof. Alessandro Piva and all the other members for accepting to be part of the judging committee and to all the technical stuff and the secretary for making things running smoothly over the past years.

Finally, I would like to eternally thank my mother and my family, I owe them my endless love and gratitude. Their support, encouragement and deep love in every single moment of my life have been and always will be the most important thing for me.

A special and warm thanks goes to all my friends, both close and far away, for sharing with me this wonderful experience that is life, supporting me in difficult moments and enjoying with me the good ones.

# Bibliography


[1] European project no. 231126 : Living knowledge. `http://livingknowledge-project.eu/`. Retrieved April 2011. 123

[2] European Project IST-1999 10987. Certimark : certification for watermarking techniques. `http://www.certimark.org`. Retrieved April 2011. 14

[3] A. M. Alattar. Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Transaction on Image Processing*, 13(8):1147–1156, 2004. 16

[4] J. Assfalg, M. Bertini, A.D. Bimbo, and W. Nunziati. Semantic annotation of soccer videos: Automatic highlights identification. *Computer Vision and Image Understanding*, 92(2-3):285–305, 2003. 67

[5] M. Barni and F. Bartolini. *Watermarking systems engineering. Enabling digital assets security and other applications*. CRC Press, Boca Raton, FL, USA, 2004. 2, 9, 20

[6] M. Barni, F. Bartolini, and A. Piva. Improved wavelet-based watermarking through pixel-wise masking. *IEEE Transaction on Image Processing*, 10(5):783–791, 2001. 26, 27, 28, 29

[7] S. Bayram, I. Avcibas, B. Sankur, and N. Memon. Image manipulation detection. *Journal of Electronic Imaging*, 15(4):1–17, 2006. 63

[8] S. Bayram, H. Sencar, and N. Memon. Video copy detection based on source device characteristics: A complementary approach to content-based methods. *ACM International Conference on Multimedia Information Retrieval*, pages 435–442, 2008. 86

[9] S. Bayram, H. T. Sencar, and N. Memon. A survey of copy-move forgery detection techniques. *IEEE Western New York Image Processing Workshop*, 2008. 63

[10] S. Bayram, H. T. Sencar, and N. Memon. An efficient and robust method for detecting copy-move forgery. *International Conference on Acoustics, Speech, and Signal Processing*, 0:1053–1056, 2009. 63

[11] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas. Source camera identification based on CFA interpolation. *IEEE International Conference on Image Processing*, 3:69–72, 2005. 62

[12] R. Böhme, F. Freiling, T. Gloe, and M. Kirchner. Multimedia forensics is not computer forensics. *Third International Workshop on Computational Forensics*, LNCS 5718:90–103, 2009. 63

[13] A. Buades, B. Coll, and J.M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008. 17

[14] G. Cao, Y. Zhao, R. Ni, and H. Tian. Anti-forensics of contrast enhancement in digital images. *ACM Multimedia and Security Workshop*, 0:25–34, 2010. 123

[15] H. Cao and A. C. Kot. Accurate detection of demosaicing regularity for digital image forensics. *IEEE Transactions on Information Forensics and Security*, 4(4):899–910, 2009. 62

[16] M.U. Celik, G. Sharma, A.M. Tekalp, and E. Saber. Lossless generalized lsb data embedding. *IEEE Transaction on Image Processing*, 14(2):253–266, 2005. 16, 44

[17] Y. L. Chang, K. T. Sun, and Y. H. Chen. Art2-based genetic watermarking. *International Conference on Advanced Information Networking*, 1:729–734, 2005. 15

[18] B. Chen and G. W. Wornell. Quantization index modulation methods for digital watermarking and information embedding of multimedia. *Journal of VLSI Signal Processing Systems - Special issue on multimedia*, 27(1):7–33, 2001. 33

[19] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš. Source digital camcorder identification using sensor photo response non-uniformity. *Proc. of SPIE*, 6505:65051G, 2007. 62

[20] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš. Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Forensics and Security*, 3(1):74–90, 2008. 62

[21] W. Chen and Y.Q. Shi. Detection of double MPEG video compression using first digits statistics. *International Workshop on Digital Watermarking*, 5450:16–30, 2008. 64, 86

[22] H. R. Chennamma and L. Rangarajan. Image splicing detection using inherent lens radial distortion. *International Journal of Computer Science*, 7(6):149–158, 2010. 65

[23] K.S. Choi, E.Y. Lam, and K. Wong. Source camera identification using footprints from lens aberration. *Proc. of SPIE*, 6069:60690J, 2006. 62

[24] I. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transaction on Image Processing*, 6(12):1673–1687, 1997. 11

[25] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2007. 2, 9, 10, 11, 12, 20, 44

[26] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung. Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications. *IEEE Journal On Selected Areas In Communications*, 16(4):573–586, 1998. 13

[27] S. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. Dean, and E. Felten. Reading between the lines: Lessons learned from the sdmi challenge. *Usenix Security Symposium*, pages 353–363, 2001. 2

[28] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1–13, 2004. 17

[29] S. Dehnie, T. Sencar, and N. Memon. Digital image forensics for identifying computer generated and digital camera images. *International Conference on Image processing*, 0:2313–2316, 2006. 62

[30] J. Delaigle, C. Devleeschouwer, B. Macq, and I. Lagendijk. Human visual system features enabling watermarking. *International Conference on Multimedia Expo*, 2:489–492, 2002. 21

[31] E. Delp and M. Wu. Digital forensics. *IEEE Signal Processing Magazine*, 2(26):14–15, 2009. 63

[32] A. DeRosa, F. Uccheddu, A. Costanzo, A. Piva, and M. Barni. Exploring image dependencies: a new challenge in image forensics. *Proc. of SPIE*, 7541:5410X–75410X–12, 2010. 123

[33] H. Ding, R. Bala, Z. Fan, R. Eschbach, C. A. Bouman, and J. P. Allebach. Semi-automatic object geometry estimation for image personalization. *SPIE Symposium on Electronic Imaging*, 7533, 2010. 84

[34] A. E. Dirik, S. Bayram, H. T. Sencar, and N. Memon. New features to identify computer generated images. *International Conference on Image processing*, 0:433–436, 2007. 62

[35] A. E. Dirik, H. T. Sencar, and N. Memon. Digital single lens reflex camera identification from traces of sensor dust. *IEEE Transactions on Information Forensics and Security*, 3(3):539 – 552, 2008. 62

[36] J. Dittmann, D. Megias, A. Lang, and J. Herrera-Joancomarti. Theoretical framework for a practical evaluation and comparison of audio watermarking schemes in the triangle of robustness, transparency and capacity. *Transaction on Data Hiding and Multimedia Security*, LNCS 4300:1–40, 2006. 14

[37] R.O. Duda and P.E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Comm. ACM*, 15:11–15, 1972. 97

[38] J. Eggers and B. Girod. *Informed Watermarking*. Kluwer Academic Publisher, Norwell,MA, USA, 2002. 2, 9, 20

[39] M. Fallahpour. Reversible image data hiding based on gradient adjusted prediction. *IECIE Electronic Express*, 5(20):870–876, 2008. 17, 44, 46, 51, 52

[40] Z. Fan and R. L. de Queiroz. Identification of bitmap compression history: Jpeg detection and quantizer estimation. *IEEE Transaction on Image Processing*, 14(2):230–235, 2003. 64

[41] H. Farid. Digital forensic database. `http://www.cs.dartmouth.edu/~farid/dfd/index.php/publications`. Retrieved April 2011. 63

[42] H. Farid. Photo tampering throughout history. `http://www.cs.dartmouth.edu/farid/research/digitaltampering`. Retrieved April 2011. 55, 70

[43] H. Farid. Exposing digital forgeries from JPEG ghosts. *IEEE Transactions on Information Forensics and Security*, 1(4):154–160, 2009. 64

[44] H. Farid. A survey of image forgery detection. *IEEE Signal Processing Magazine*, 2(26):16–25, 2009. 63

[45] H. Farid and M.J. Bravo. Image forensic analyses that elude the human visual system. *SPIE Symposium on Electronic Imaging*, 2010. 71

[46] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 74

[47] J. Fridrich. Digital image forensics: Introducing methods to estimate and detect sensor fingerprint. *IEEE Signal Processing Magazine*, 2(26):26–37, 2009. 62

[48] J. Fridrich, M. Goljan, and R. Du. Lossless data embedding: new paradigm in digital watermarking. *EURASIP Journal on Applied Signal Processing*, 2002(1):185–196, 2002. 16, 44

[49] G. Friedman. The trustworthy digital camera: Restoring credibility to the photographic image. *IEEE Transactions on Consumer Electronics*, 39(4):905–910, 1993. 2

[50] Z. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh. Methods for identification of images acquired with digital cameras. *Proc. of SPIE*, 4232:505–512, 2001. 62

[51] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme. Can we trust digital image forensics? *15th International Conference on Multimedia*, 0:78–86, 2007. 123

[52] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Inc., Boston MA, USA, 1999. 15, 22, 23, 27

[53] M. Goljan, J. Fridrich, and M. Chen. Sensor noise camera identification: Countering counter-forensics. *Proc. of SPIE*, 7541:0S1–0S12, 2010. 123

[54] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001. 44

[55] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 66, 73, 74, 93

[56] F. Hartung, J. K. Su, and B. Girod. Spread spectrum watermarking: Malicious attacks and counterattacks. *Proc. SPIE*, 3657:147–158, 1999. 13

[57] H. Hecht, M. K. Kaiser, and M. S. Banks. Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, 58(7):1066–1075, 1996. 86

[58] W. Hong, T. S. Chen, and C. W. Shiu. Reversible data hiding based on histogram shifting of prediction errors. *IEEE International Symposium on Intelligent Information Technology Application Workshops*, 0:292–295, 2008. 17, 44, 46, 51, 52

[59] W. Van Houten and Z.J. Geradts. Source video camera identification for multiply compressed videos originating from YouTube. *Digital Investigation*, 6(1-2):48–60, 2009. 86

[60] C. Hsu, T. Hungand C. Lin, and C. Hsu. Video forgery detection using correlation of noise residue. *IEEE Workshop on Multimedia Signal Processing*, pages 170–174, 2008. 86

[61] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu. Video forgery detection using correlation of noise residue. *IEEE International Workshop on Multimedia Signal Processing*, pages 170–174, 2008. 62

[62] Y.-F. Hsu and S.-F. Chang. Camera response functions for image forensics: An automatic algorithm for splicing detection. *IEEE Transactions on Information Forensics and Security*, 5(4):816–825, 2010. 65

[63] Y. Hu, H. K. Lee, K. Chen, and J. Li. Difference expansion based reversible data hiding using two embedding directions. *IEEE Transanction on Multimedia*, 10(8):1500–1512, 2008. 16, 44

[64] Y. Hu, H.-K. Lee, and J. Li. De-based reversible data hiding with improved overflow location map. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(2):250–260, 2009. 17

[65] C. H. Huang and J. L. Wu. A watermark optimization technique based on genetic algorithms. *Proc. of SPIE - Visual Communications Image Processing*, 3971:516–523, 2000. 15

[66] J. Hwang, J. Kim, and J. Choi. A reversible watermarking based on histogram shifting. *International Workshop on Digital Watermarking*, LNCS 4283:348–361, 2006. 16

[67] M.K. Johnson and H. Farid. Exposing digital forgeries by detecting inconsistencies in lighting. *ACM Multimedia and Security Workshop*, 0:1–10, 2005. 65

[68] M.K. Johnson and H. Farid. Exposing digital forgeries through chromatic aberration. *ACM Multimedia and Security Workshop*, 0:48–55, 2006. 64

[69] M.K. Johnson and H. Farid. Metric measurements on a plane from a single image. *TR2006-579*, 2006. 66

[70] M.K. Johnson and H. Farid. Detecting photographic composites of people. *6th International Workshop on Digital Watermarking*, 0:19–33, 2007. 65, 66, 71

[71] M.K. Johnson and H. Farid. Exposing digital forgeries in complex lighting environments. *IEEE Transactions on Information Forensics and Security*, 3(2):450–461, 2007. 65

[72] M.K. Johnson and H. Farid. Exposing digital forgeries through specular highlights on the eye. *9th International Workshop on Information Hiding*, 0:311–325, 2007. 65

[73] T. Kalker. Considerations on watermarking security. *IEEE Fourth Workshop on Multimedia Signal Processing*, 0:201–206, 2001. 10

[74] L. Kamstra and H. J. A. M. Heijmans. Reversible data embedding into images using wavelet techniques and sorting. *IEEE Transactions on Image Processing*, 4(12):2082–2090, 2005. 16

[75] E. Kee and H. Farid. Exposing digital forgeries from 3-d lighting environments. *IEEE International Workshop on Information Forensics and Security*, 0:1–6, 2010. 65

[76] S. Khan and A. Kulkarni. An efficient method for detection of copy-move forgery using discrete wavelet transform. *International Journal on Computer Science and Engineering*, 2(5):1801–1806, 2010. 63

[77] T. Kim, Y. Seo, and K.-S. Hong. Physics-based 3D position analysis of a soccer ball from monocular image sequences. *IEEE International Conference on Computer Vision*, 0:721–726, 1998. 67

[78] M. Kirchner. Linear row and column predictors for the analysis of resized images. *ACM Workshop on Multimedia and Security*, 0:13–18, 2010. 64

[79] M. Kirchner and R. Bhme. Hiding traces of resampling in digital images. *15th International Conference on Multimedia*, 3(4):582–592, 2008. 123

[80] M. Kirchner and R. Bhme. Synthesis of color filter array pattern in digital images. *Proc. of SPIE*, 7254:725421, 2009. 123

[81] M. Kirchner and J. Fridrich. On detection of median filtering in images. *Proc. of SPIE*, 7541:101–1012, 2010. 123

[82] J. D. Kornblum. Using JPEG quantization tables to identify imagery processed by software. *Digital Investigation*, 5(1):S21–S25, 2008. 64

[83] P. Kumsawat, K. Attakitmongcol, and A. Srikaew. A new approach for optimization in image watermarking by using genetic algorithm. *IEEE Transaction on Image Processing*, 53(12):4707–4719, 2005. 15

[84] M. Kuribayashi, M. Morii, and H. Tanaka. Reversible watermark with large capacity using the predictive coding. *IECIE Transactions on Fundamentals of Electronics, Communications, and Computer Science*, 7(7):1780–1790, 2008. 17

[85] K. Kurosawa, K. Kuroki, and N. Saitoh. CD fingerprint method-identification of a video camera from videotaped images. *IEEE Conference on Image processing*, 3:537–540, 1999. 62

[86] A. Lang. Stirmark benchmark for audio. `http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php`. Retrieved April 2011. 14

[87] T. Ngan Le, K. Hung Nguyen, and H. Bac Le. Literature survey on image watermarking tools, watermark attacks and benchmarking tools. *Second International Conference on Advances in Multimedia*, 0:67–73, 2010. 14

[88] S. Lee, C. D. Yoo, and T. Kalker. Reversible image watermarking based on integer-to-integer wavelet transform. *IEEE Transaction on Information Forensics and Security*, 2(3):321–330, 2007. 16

[89] S. J. Lee and S.H Jung. A survey of watermarking techniques applied to multimedia. *IEEE International Symposium on Industrial Electronics*, 1:272–277, 2001. 3

[90] D. Levicky, P. Foris, and N. Nikolaidis. Human visual system models in digital image watermarking. *Journal of Radio Engineering*, 13(4):38–43, 2004. 21

[91] A. Lewis. Multimedia forensics bibliography. `http://www.cl.cam.ac.uk/~abl26/bibliography`. Retrieved April 2011. 63

[92] C.-T. Li. Detection of block artifacts for digital forensic analysis. *Second International Conference on Forensics in Telecommunications, Information and Multimedia*, 8:173–178, 2009. 64

[93] Q. Li and I. J. Cox. Using perceptual models to improve fidelity and provide resistance to volumetric scaling for quantization index modulation watermarking. *IEEE Transaction on Information Forensics and Security*, 2(2):127–139, 2007. 26, 27, 33, 34, 35, 39

[94] C.-C. Lien, C.-L. Shih, and C.-H. Chou. Fast forgery detection with the intrinsic resampling properties. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 0:232–235, 2010. 64

[95] C.C. Lin and N.L. HSueh. A lossless data hiding scheme based on three-pixel block differences. *Pattern Recognition*, 41(4):1415–1425, 2008. 16

[96] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp. Advances in digital video content protection. *Proc. of IEEE*, 93(1):171–183, 2005. 20

[97] Y. Liu, D. Liang, Q. Huang, and W. Gao. Extracting 3D information from broadcast soccer video. *Image and Vision Computing*, 24(10):1146–1162, 2006. 67

[98] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110, 2004. 74, 75, 93

[99] J. Lukas and J. Fridrich. Estimation of primary quantization matrix in double compressed jpeg images. *Digital Forensic Research Workshop*, 2003. 64

[100] S. Lyu and H. Farid. How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 53(2):845–850, 2005. 62

[101] B. Macq, J. Dittman, and E. Delp. Benchmarking of image watermarking algorithms for digital rights management. *Proc. of IEEE*, 92(6):971–984, 2004. 3, 14

[102] B. Mahdian and S. Saic. Blind authentication using periodic properties of interpolation. *IEEE Transactions on Information Forensics and Security*, 3(3):529–538, 2008. 64

[103] B. Mahdian and S. Saic. Blind methods for detecting image fakery. *IEEE Aerospace and Electronic Systems Magazine*, 25(4):18–24, 2010. 63

[104] M. Mahmoudi and G. Sapiro. Fast image and video de-noising via nonlocal means of similar neighborhoods. *IEEE Signal Processing Letters*, 12(12):839–842, 2005. 17

[105] S. P. Maity, M. K. Kundu, and P. K. Nandi. Genetic algorithm for optimal imperceptibility in image communication through noisy channel. *International Conference on Neural Information Processing*, 3316:700–705, 2004. 15

[106] P. Meerwald and A. Uhl. A survey of wavelet-domain watermarking algorithms. *Proc. of SPIE*, 4314:505–516, 2001. 11

[107] N. Memon and P. Wong. Protecting digital media content. *Magazine Communications of ACM*, 41(7):35–43, 1998. 9

[108] M. Miller, I. Cox, J. Linnartz, and T. Kalker. A review of watermarking principles and practices. *Digital Signal Processing in Multimedia Systems*, Chapter 18:461–485, 1999. 11

[109] N. Mondaini, R. Caldelli, A. Piva, M. Barni, and V. Cappellini. Detection of malevolent changes in digital video for forensic applications. *Proc. of SPIE*, 6505:65051T, 2007. 62

[110] N. Mondaini, Roberto Caldelli, Alessandro Piva, Mauro Barni, and Vito Cappellini. Detection of malevolent changes in digital video for forensic applications. *SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents*, 6505, 2007. 86

[111] R. A. Nash and K. A. Wade. Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. *Memory and Cognition*, 37(4):414–424, 2009. 55

[112] T.-T. Ng and S.-F. Chang. A model for image splicing. *International Conference on Image Processing*, 2:1169–1172, 2004. 64

[113] T.-T. Ng and S.-F. Chang. Identifying and prefiltering images: distinguishing between natural photography and photorealistic computer graphics. *IEEE Signal Processing Magazine*, 2(26):49–58, 2009. 62

[114] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.-P. Tsui. Physics-motivated features for distinguishing photographic images and computer graphics. *13th ACM international conference on Multimedia*, 0:239–248, 2005. 62

[115] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su. Reversible data hiding. *IEEE Transaction on Circuits and Systems for Video Technology*, 16(3):354–362, 2006. 16, 44

[116] K. Nishiwaki, A. Konno, K. Nagashima, M. Inaba, , and H. Inoue. The humanoid saika that catches a thrown ball. *IEEE International Workshop on Robot and Human Communication*, 0:94–99, 1997. 67

[117] A. Pal and N. Memon. The evolution of file carving: the benefits and problems of forensics recovery. *IEEE Signal Processing Magazine*, 2(26):59–71, 2009. 63

[118] X. Pan and S. Lyu. Detecting image region duplication using sift features. *nternational Conference on Acoustics, Speech, and Signal Processing*, 0:1706–1709, 2010. 63

[119] S. Pereira. Checkmark benchmark. `http://watermarking.unige.ch/Checkmark`. Retrieved April 2011. 14

[120] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun. Second generation benchmarking and application oriented evaluation. *Workshop on Information Hiding*, 2137:340–353, 2001. 14

[121] F. Petitcolas. Stirmark benchmark. `http://www.petitcolas.net/fabien/watermarking/stirmark`. Retrieved April 2011. 14

[122] F. Petitcolas, R. J. Anderson, and M. Kuhn. Attacks on copyright marking systems. *Second International Workshop on Information Hiding*, 1525:218–238, 1998. 14

[123] A. Piva, F. Bartolini, and M. Barni. Managing copyright in open networks. *IEEE Transactions on Internet Computing*, 6(3):18–26, 2002. 9

[124] A.C. Popescu and H. Farid. Exposing digital forgeries by detecting duplicated image regions. Technical Report TR2004-515, Department of Computer Science, Dartmouth College, 2004. 63

[125] A.C. Popescu and H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, 53(10):3948–3959, 2005. 65

[126] R. Manjunatha Prasad and S. Koliwad. A comprehensive survey of contemporary researches in watermarking for copyright protection of digital images. *International Journal of Computer Science and Network Security*, 9:91–107, 2009. 11

[127] G. Qadir, X. Zhao, and A. T. S Ho. Estimating JPEG2000 compression for image forensics using Benford's law. *Proc. of SPIE*, 7723:77230J, 2010. 64

[128] Z. Qu, W. Luo, and J. Huang. Mixing model for shifted double jpeg compression with application to passive image authentication. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 0:1661–1664, 2003. 64

[129] S. D. Rane, G. Sapiro, and M. Bertalmio. Structure and texture filling-in of missing image blocks in wireless transmission and compression applications. *IEEE Transactions on Image Processing*, 12(3):296–303, 2003. 17

[130] I.D. Reid and A. North. 3D trajectories from a single viewpoint using shadows. *British Machine Vision Conference*, 0:721–726, 1998. 67

[131] J. Ren, J. Orwell, G.A. Jones, and M. Xu. A general framework for 3D soccer ball estimation and tracking. *IEEE International Conference on Image Processing*, 0:1935–1938, 2004. 67

[132] J. Rhedi, W. Taktak, and J-L. Dugelay. Digital image forensics: a booklet for beginners. *Multimedia Application Tools*, 51:133–162, 2011. 63

[133] E. Ribnikc, S. Atev, and N. Papanikolopoulos. Estimating 3D positions and velocities of projectiles from monocular views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):938–944, 2009. 67

[134] M. Riley and C.G. Atkeson. Robot catching: Towards engaging human-humanoid interaction. *Autonomous Robots*, 12(1):119–128, 2002. 67

[135] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein. Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Computing Surveys*. To appear in 2011. 63

[136] D. L. M. Sacchi, F. Agnoli, and E. F. Loftus. Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology*, 21(8):1005–1022, 2007. 55

[137] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y. Q. Shi. Reversible watermarking algorithm using sorting and prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(7):989–999, 2009. 17

[138] G. Sankar, H. V. Zhao, and Y.-H. Yang. Feature based classification of computer graphics and real images. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 0:1513–1516, 2009. 62

[139] R. Schyndel, A. Tirkel, and C. Osborne. A digital watermark. *IEEE International Conference on Image Processing*, 2:86–90, 1994. 11

[140] H. T. Sencar and N. Memon. Overview of state-of-the-art in digital image forensics. *Statistical Science and Interdisciplinary Research*, 3:325–347, 2008. 4, 62, 63

[141] M. Shehab, E. Bertino, and A. Ghafoor. Watermarking relational databases using optimization based techniques. *IEEE Transaction on Knowledge Data Engineering*, 20(1):116–129, 2008. 15

[142] C. S. Shieh, C. Huang, F. H. Wang, and J. S. Pan. Genetic watermarking based on transform-domain techniques. *Elsevier pattern Recognition*, 37(3):555–565, 2004. 15

[143] F. Y. Shih and F.-T. Wu. Enhancement of image watermark retrieval based on genetic algorithms. *Journal of Visual Communication and Image representation*, 16(2):115–133, 2005. 15

[144] V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, and I.Pitas. Optimark benchamrk. `http://poseidon.csd.auth.gr/optimark`. Retrieved April 2011. 14

[145] V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, and I. Pitas. A benchmarking protocol for watermarking methods. *IEEE International Conference on Image Processing*, 3:1023–1026, 2001. 15

[146] D. Soukal, J. Lukas, and J. Fridrich. Detection of copy-move forgery in digital images. *Proc. of Digital Forensic Research Workshop*, 2003. 63

[147] M. Stamm, S. Tjoa, W. S. Lin, and K. J. Ray Liu. Anti-forensics of jpeg compression. *International Conference on Acoustics, Speech, and Signal Processing*, 0:1694–1697, 2010. 123

[148] M. Stamm, S. Tjoa, W. S. Lin, and K. J. Ray Liu. Wavelet-based image compression anti-forensics. *International Conference on Image Processing*, 0:1737 – 1740, 2010. 123

[149] D. M. Thodi and J. Rodriguez. Expansion embedding techniques for reversible watermarking. *IEEE Transactions on Image Processing*, 16(3):721–730, 2007. 16

[150] D.M. Thodi and J.J. Rodriguez. Prediction-error based reversible watermarking. *IEEE International Conference on Image Processing*, 3:1549–1552, 2004. 16

[151] T. Thormählen and H. Broszio. Voodoo camera tracker: A tool for the integration of virtual and real scenes. `http://www.digilab.uni-hannover.de/docs/manual.html`. Retrieved April 2011. 93

[152] J. Tian. Reversible data embedding using a difference expansion. *IEEE Transaction on Circuits and Systems for Video Technology*, 13(8):890–896, 2003. 16, 44

[153] P. Tsai, Y.-C. Hu, and H.-L. Yeh. Reversible image hiding scheme using predictive coding and histogram shifting. *Signal Processing*, 89(6):1129–1143, 2009. 17

[154] H.-W. Tseng and C.-P Hsieh. Prediction-based reversible data hiding. *Elsevier, Information Science*, 179(14):2460–2469, 2009. 17

[155] W. van Houten and Z. J. M H. Geradts. Using sensor noise to identify low resolution compressed videos from youtube. *Third International Workshop on Computational Forensics*, LNCS 5718:104–115, 2009. 62

[156] T. van Lanh, K.S. Chong, S. Emmanuel, and M. Kankanhally. A survey on digital camera image forensic methods. *IEEE International Conference on Multimedia and Expo*, 0:16–19, 2007. 63

[157] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. *International Workshop on Information Hiding*, 0:211–236, 1999. 15, 21

[158] S. Voloshynovskiy, S. Pereira, A. Herrigel, N. Baumgartner, and T. Pun. Generalized watermarking attack based on watermark estimation and perceptual remodulation. *Proc. of SPIE*, 3971:358–370, 2000. 50

[159] G. Voyatzis and I. Pitas. The use of watermarks in the protection of digital multimedia products. *Proc. of IEEE*, 87(7):11971207, 1999. 9

[160] G. Wang, , Z. Hu, , and F. Wu andH.T. Tsui. Single view metrology from scene constraints. *Image and vision computing*, 23(9):831–840, 2005. 66

[161] K. Wang, G. Lavou, F. Denis, and A. Baskurt. A benchmark for 3d mesh watermarking. `http://liris.cnrs.fr/meshbenchmark/`. Retrieved April 2011. 15

[162] W. Wang, J. Dong, and T. Tan. Effective image splicing detection based on image chroma. *IEEE International Conference on Image Processing*, 0:1257–1260, 2009. 64

[163] W. Wang and H. Farid. Exposing digital forgeries in video by detecting double MPEG compression. *ACM Multimedia and Security Workshop*, pages 37–47, 2006. 86

[164] W. Wang and H. Farid. Exposing digital forgeries in interlaced and de-interlaced video. *IEEE Transactions on Information Forensics and Security*, 3(2):438–449, 2007. 65, 86

[165] W. Wang and H. Farid. Exposing digital forgeries in video by detecting duplication. *ACM Multimedia and Security Workshop*, 0:35–42, 2007. 63

[166] W. Wang and H. Farid. Exposing digital forgeries in video by detecting duplication. *ACM Multimedia and Security Workshop*, pages 35–42, 2007. 86

[167] W. Wang and H. Farid. Detecting re-projected video. *10th International Workshop on Information Hiding*, 0:72–86, 2008. 66, 71

[168] W. Wang and H. Farid. Exposing digital forgeries in video by detecting double quantization. *ACM Multimedia and Security Workshop*, 0:39–48, 2009. 64

[169] W. Wang and H. Farid. Exposing digital forgeries in video by detecting double quantization. *ACM Multimedia and Security Workshop*, pages 39–48, 2009. 86

[170] Z. Wang and A. C. Bovik. *Human visual system features enabling watermarking*. Morgan and Claypool, New York, NY, USA, 2006. 21

[171] A. B. Watson. Dct quantization matrices visually optimized for individual images. *Proc. of SPIE*, 1913:202–216, 1993. 33

[172] J. S. Watson, M. S. Banks, C. von Hofsten, and Constance S. Royden. Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception*, 21:69–76, 1992. 86

[173] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. *IEEE Computer Vision and Pattern Recognition*, 0:2464–2471, 2010. 17

[174] A. Wong and J. Orchard. A nonlocal-means approach to exemplar-based inpainting. *IEEE International Conference on Image Processing*, 0:2600–2603, 2008. 17

[175] Y.-T. Wu and F. Y. Shih. Genetic algorithm based methodology for breaking the steganalytic systems. *IEEE Transaction on System man and Applications*, 36(1):24–31, 2006. 15

[176] X. Xia, C. Boncelet, and G. Arce. A multiresolution watermark for digital images. *IEEE International Conference on Image Processing*, 0:548–551, 1997. 11

[177] G. Xuan, J. Zhu, J. Chen, Y. Q. Shi, Z. Ni, and W. Su. Distortionless data hiding based on integer wavelet transform. *Electronics Letters*, 38(25):1646–1648, 2002. 16

[178] J. Zhang, Y. Su, and M. Zhang. Exposing digital video forgery by ghost shadow artifact. *First ACM Workshop on Multimedia in Forensics*, pages 49–54, 2009. 64

[179] J. Zhang, H. Wang, and Y. Su. Detection of double-compression in jpeg2000 images. *International Symposium on Intelligent Information Technology Application*, 0:418–421, 2008. 64

[180] W. Zhang, X. Cao, Z. Feng, J. Zhang, and P. Wang. Detecting photographic composites using two-view geometrical constraints. *IEEE International Conference on Multimedia and Expo*, 0:1078–1081, 2009. 66, 71

[181] W. Zhang, X. Cao, J. Zhang, J. Zu, and P. Wang. Detecting photographic composites using shadows. *IEEE International Conference on Multimedia and Expo*, 0:1042–1045, 2009. 66

[182] N. Zhong, Z. He, J. Kuang, and Z. Zhuo. An optimal wavelet based image watermarking via genetic algorithm. *International Conference on Natural Computation*, 3:103–107, 2007. 15

# Publications

### Journals

J1 - V. Conotter, J. O'Brien and H. Farid, "Exposing Digital Forgeries in Ballistic Motion". Submitted to *IEEE Transaction on Information Forensics & Security*, 2011.

J2 - G. Boato, V. Conotter, F. De Natale, C. Fontanari, "Watermarking Robustness Evaluation based on Perceptual Quality via Genetic Algorithms". *IEEE Transaction on Information Forensics & Security*, 4:207-216, June 2009.

J3 - G. Boato, N. Conci, V. Conotter, F. De Natale, C. Fontanari, Multimedia asymmetric watermarking and encryption , *Electronic Letters: an international publication*, 44:601-603, April 2008.

### International Conferences and Workshops

C1 - V. Conotter, L. Cordin "Detecting Photographic and Computer Generated Composites", *Proc. of SPIE 2011*, San Francisco (CA, USA), January 2011.

C2 - A. Cortiana, V. Conotter, G. Boato, F.G.B. De Natale "Performance comparison of denoising filters for source camera identification", *Proc. of SPIE 2011*, San Francisco (CA, USA), January 2011.

C3 - V. Conotter, G. Boato, H. Farid "Detecting Photo Manipulation on Signs and Billboards", *IEEE International Conference on Image Processing 2010*, Hong Kong, September 2010.

C4 - V. Conotter, G. Boato, M. Carli, K. Egiazarian "Near Lossless Reversible Data Hiding Based On Adaptive Prediction", *IEEE International Conference on Image Processing 2010*, Hong Kong, September 2010.

C5 - V. Conotter, G. Boato, C. Fontanari, F. G. B. De Natale "Comparison of Watermarking Algorithms via a GA-based Benchmarking Tool", *IEEE International Conference on Image Processing 2009*, Cairo (Egypt), November 2009.

C6 - G. Boato, V. Conotter, F. G. B. De Natale, C. Fontanari "Towards Multimedia Opinion Mining", International Workshop Living Web 2009, Washington DC (USA), October 2009.

C7 - V. Conotter, G. Boato, M. Carli, K. Egiazarian "High Capacity Reversible Data Hiding based on Histogram Shifting and Non-local Means" (invited paper), *IEEE International Conference on Linear and Non-Linear Approximation 2009*, Tuusula (Finland), August 2009.

C8 - V. Conotter , G. Boato, C. Fontanari, F. De Natale, "Robustness and Security Assessment of Image Watermarking Techniques by a Stochastic Approach". , *Proc. of SPIE 2009*, San Jose (CA), January 2009.

C9 - G. Boato, V. Conotter, C. Fontanari, F. De Natale, A joint asymmetric watermarking and image encryption scheme , *Proc. of SPIE 2008*, San Jose (CA), January 2008.

C10 - G. Boato, V. Conotter , F. De Natale, GA-based robustness evaluation method for digital image watermarking. *International Workshop on Digital watermarking (IWDW2007)*, Guangzhou (Cina), December 2007.