International Journal of Sociotechnology and Knowledge Development
Volume 12 • Issue 1 • January-March 2020

Copy-Move Forgery Detection Based on Automatic Threshold Estimation

Aya Hegazi, Faculty of Computers and Informatics, Benha University, Benha, Egypt
Ahmed Taha, Faculty of Computers and Informatics, Benha University, Benha, Egypt
Mazen Mohamed Selim, Faculty of Computers and Informatics, Benha University, Benha, Egypt

ABSTRACT

Recently, users and news followers across websites face many fabricated images. Moreover, it goes far beyond that to the point of defaming or imprisoning a person. Hence, image authentication has become a significant issue. One of the most common tampering techniques is copy-move. Keypoint-based methods are considered as an effective method for detecting copy-move forgeries. In such methods, the feature extraction process is followed by applying a clustering technique to group spatially close keypoints. Most clustering techniques highly depend on the existence of a specific threshold to terminate the clustering. Determination of the most suitable threshold requires a huge amount of experiments. In this article, a copy-move forgery detection method is proposed. The proposed method is based on automatic estimation of the clustering threshold. The cutoff threshold of hierarchical clustering is estimated automatically based on clustering evaluation measures. Experimental results tested on various datasets show that the proposed method outperforms other relevant state-of-the-art methods.

KEYWORDS

Clustering Evaluation Measures, Copy-Move Detection, Image Forensics, Keypoint-Based Methods, Multiple-Copied Matching

1. INTRODUCTION

Digital images are everywhere, and they have the power to do infinitely more than a document. In the latter half of the last two decades, the internet, mobile technology, and obsession of social media have highly affected and changed people's lives (Katta & Patro, 2017; Mahajan et al., 2018; Muliawat et al., 2019). Recently, there is a rapid increase in images showing in the media as in social media and television that don't seem to be all as they appear. Authenticity of digital images is a critical issue. Day by day, it becomes easy for anyone to manipulate images even without leaving any visible clues. Wide availability of powerful image processing software like Photoshop and Gimp makes it more challenging for digital image authentication.

Digital image forensics is the science of detecting tampered regions in images. Identifying the authenticity of digital images is very important in digital forensics. The purpose of digital image manipulation is to conceal or hide information for several intentions therefore change their meaning. Many areas have been affected by digital forensics. The impact of image manipulation in media, journalism, digital cinema, news and in politics to mislead the public opinion. It could also be used in law for miscarrying justice. Manipulated images also have been found in academic papers. In a survey by Tijdink (Tijdink et al., 2014), in the past three years, 15% of offenders are involved in scientific

DOI: 10.4018/IJSKD.2020010101

misconduct such as fabricating, refutation or manipulating data. A study by (Farid, 2006) reported that in the Journal of Cell Biology about 20% of admitted manuscripts have at least one figure that must be restored due to unsuitable image manipulation, and about 1% are deceitful figures. These consequences make image authenticities less trustful.

Rapid growth of forged images and their influence in many areas have led to the development of tampering detection techniques. Tampering detection techniques fall under two categories: active authentication and passive authentication methods (Al-Qershi & Khoo, 2013) as shown in Figure 1. Active authentication methods require some preprocessing on digital images like watermarking, or signatures. Digital watermarking conceals a watermark into the image at the capturing end and extracts it at the authentication end to examine whether the image has been tampered with (Al-Qershi & Khoo, 2013). Inserting the watermark either at the capturing time of the image using a specially equipped camera or later by an authorized person is the main drawback of watermarking (Qureshi & Deriche, 2015). Moreover, most of the cameras today are not equipped with a watermark embedding technique. In addition, the subsequent processing of the original image could degrade the image visual quality. Moreover, digital signature is similar to digital watermarking. At the image-capturing end, unique features are extracted from the image as a signature. At the authentication and detection end, the signature is regenerated using the same method and the authenticity of the image can be identified and verified through comparison. Digital signatures have the same drawbacks of digital watermarking.

On the other hand, Passive (blind) authentication methods authenticate images while not requiring any previous information of it. It relies on the traces left on the image during manipulation by various processing operations. Therefore, passive authentication methods are considered as the most common (Lin et al., 2018). Passive detection techniques can be classified to forgery-type dependent or forgery-type independent. Forgery-type independent techniques detect forgeries regardless of the type of the forgery. To detect general tampering, the independent techniques exploit three diverse types of artifacts: traces of re-sampling, compression and inconsistencies (Redi et al., 2011). The forgery-type dependent techniques are used for certain types of forgeries. Copy-move and splicing are examples of forgery-type dependent (see Figure 1). Such techniques depend on copying and pasting image regions either from the same image (copy-move), or from different images (splicing). Image splicing is created from at least two different images (Sharma & Ghanekar, 2019; Walia & Kumar, 2018). An Example of image splicing is shown in Figure 2(d).

Copy-move or cloning is a technique of copying a region and pasting it in the same image. It contains at least two regions alike (see Figure 2(b)). Since the duplicated regions are from the same image, they inherit the same basic image properties such as color palette, illumination conditions and noise. Copy-move forgery is the most common type used for image manipulation due to its simplicity and effectiveness (Al-Qershi & Khoo, 2013; Bakiah et al., 2016). Although this technique is easy to implement, it is hard to detect. Often in practice, forgery is not just limited to copying and pasting the regions, some processing operations are applied to these regions. These operations can be classified to intermediate operations (geometric transformations) and post-processing operations. Intermediate operations are used to provide a spatial synchronization and homogeneity between the copied region and its neighbors (Al-Qershi & Khoo, 2013; Bakiah et al., 2016). Examples of intermediate operations are rotation and scaling. Post-processing operations are used to remove traces left from forgery and to make it unnoticeable. Additive noise, JPEG compression and blurring are examples of post-processing operation (Liu et al., 2010). Since all those operations make detecting copy-move forgery more challenging, numerous methods have been proposed for Copy-Move Forgery Detection (CMFD). Most of them can be classified either into block-based methods or keypoint-based methods (Bakiah et al., 2016). In block-based methods, the image is divided into overlapping or non-overlapping blocks of fixed size. On the other hand, keypoint-based methods calculate local interest points (keypoints) from the whole image without any subdivisions.

Generally, CMFD techniques follow a common pipeline as shown in Figure 3 (Christlein et al., 2012). Both block-based and keypoint-based methods follow the same pipeline steps except for