ISO/IEC JTC 1/SC 29/WG 1

(ITU-T SG16)

Coding of Still Pictures

JBIG
Joint Bi-level Image Experts Group

JPEG
Joint Photographic Experts Group

TITLE: **JPEG Fake Media: Context, Use Cases and Requirements Draft**

SOURCE: JPEG Fake Media AHG

PROJECT: Requirements

STATUS:

REQUESTED ACTION: For public distribution

DISTRIBUTION: WG1

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

JPEG Fake Media: Context Use Cases and Requirements

1. Executive Summary

Recent advances in media manipulation, particularly deep learning based approaches, can produce near realistic media content that is almost indistinguishable from authentic content to the human eye. While these developments open opportunities for production of new types of media contents that are useful for the entertainment industry, it also risks the spread of doctored media (e.g., 'deepfake') leading to copyright infringement, social unrest, spread rumours for political gain or encouraging hate crimes.

Declaration of media manipulation is considered to be important in many usage scenarios including version controlling or traceability which however is not always the case where the intention is to 'hide' the mere existence of such manipulations. This is already leading various Governmental organizations to plan new legislation or companies (especially social media platforms or news outlets) to develop mechanisms that would clearly detect and annotate manipulated media contents when they are shared. Thus, there is a clear need for standardisation related to media content and associated metadata. The JPEG Committee is interested in exploring if a JPEG standard can facilitate a secure and reliable annotation of media modifications, both in good faith and malicious usage scenarios.

2. Introduction

2.1. Context & Motivation

Recent advances in media manipulation, particularly deep learning based approaches, can produce near realistic media content that is almost indistinguishable from authentic content to the human eye. These developments open opportunities for production of new types of media contents that are useful for the entertainment industry and other business usage, e.g., creation of special effects or artificial natural scene production with actors in the studio. However, this also leads to issues relating to fake media generation defying the integrity of the media (e.g., deepfakes), copyright infringements and defamation to mention some. Misuse of manipulated media can cause social unrest, spread rumours for political gain or encourage hate crimes. In this context, the term 'Fake Media' is used here to refer to any generated or modified media, independently of its 'good' or 'bad' intention.

In many application domains, the 'media manipulators' may want to declare the type of manipulations performed, in opposition to other situations where the intention is to 'hide' the mere existence of such manipulations. This is already leading various Governmental organizations to plan new legislation or companies (especially social media platforms or news outlets) to develop mechanisms that would clearly detect and annotate manipulated media contents when they are shared. While growing efforts are noticeable in developing technologies, there is a need to have a standard for the media content and associated metadata, e.g., a JPEG standard that facilitates a secure and reliable annotation of fake media, both in good faith and malicious usage scenarios. Therefore, it is important for the JPEG Committee to better understand the fake media ecosystem and needs in terms of standardization through an in-depth analysis of fake media use cases, naturally independently of the 'intentions'.

2.2. Objectives

It is envisaged that JPEG initiates a standardization activity in order to ensure interoperability between a wide range of applications dealing with fake media. To reach this goal, and as a first step, stakeholders are invited to join this effort by helping to better understand applications and scenarios relevant to fake media use cases. This will allow the JPEG Committee to then identify key requirements for a standard in fake media. Initial findings suggest that a set of standard metadata to signal fake media content along with relevant information on the latter are needed. In addition, standard mechanisms for security and protection of integrity both metadata and fake media content are desired. The latter is closely related to issues highlighted in media blockchain under progress in the last two years in JPEG and therefore is considered as a natural continuation of that effort.

2.3. Scope of standardization

There is a clear need for standardisation related to media content and associated metadata. The JPEG Committee is interested in exploring if a JPEG standard can facilitate a secure and reliable annotation of media generation and modifications, both in good faith and malicious usage scenarios. It is also important to understand in more depth the usage scenarios which will require input from relevant industries, public bodies (responsible for legislations), technology providers and end-users. Therefore, the JPEG committee has the intention to engage with stakeholders in this use case in order to develop a clearly defined roadmap for standardization.

3. Use cases

The JPEG committee currently identified use cases related to the following topics:

- Misinformation and fake news
    - Deepfakes
    - Manipulated media
    - Authentic media used out of context
- Forgery / Media forensics
    - Document forgery (e.g. IDs and passports)
    - Insurance fraud (e.g. pictures of accidents)
    - KYC (Know Your Customer) (e.g. fake identity)
    - Impostoring (e.g. impersonating a celebrity)
- Media modification
    - Image editing software
    - Movie preservation
    - Film enhancement
    - Restoration of old movies
- Media creation
    - Use of deep fakes for special effects
    - Green screens, media processing and composition
    - Short content bursts

- UGC (User Generated Content) e.g. TikTok, Triller, Adobe Spark
- Media tracing, e.g. provenance, content versioning, context
- Picture and movies production

Based on these topics, the following sections provide a preliminary overview of relevant use cases. Both the topics and use cases will be extended in the future based on feedback from stakeholders.

3.1. Misinformation and fake news

Journalism

In his coverage, a journalist wants to use images from a social media post depicting police violence during protests. The journalist has to make a fast decision but of course he wants to be sure the image in the post is genuine, unaltered and taken at the mentioned place and time.

Deep fake detection

A news host wants to double check if a video he received of the president making questionable claims is genuine and not a deep fake.

Content authenticity checking

An historian wants to verify if an image depicting past atrocities is actually from that era and place.

Content usage tracing

A photographer wants to find out where and how some of the images from his portfolio have been used.

Academic research

An academic journal reviewer might want to know if an image used as evidence for a successful experiment hasn't been altered.

Photographic framing

A journalist received images of the Grand Place in Brussels in the aftermath of the terroristic attacks. Due to the specific framing the images give a frightening impression of the situation. Therefore, the journalist wants to compare with other images taken at the same place and time but from different perspectives to better evaluate the actual situation.

3.2. Media modification

Image colorization and restoration

A developer has created an algorithm that uses deep learning to colorize grayscale images and enhances the image quality. The output images are assigned a label to allow consumers to identify these images have been processed and might diverge from reality.

Photo editing

A photographer uses Photoshop to edit model pictures for a magazine. The final images are labeled to indicate that they are post-processed. The labels allow to indicate how "severe" the changes are to distinguish simple contrast and tone enhancements from changes where content has been added, removed or altered in shape.

3.3. Media creation

Movie special effects

A creative movie production company has created several shots for a movie that are computer generated but almost indistinguishable from real footage. The generated footage is labelled to allow consumers to identify the content is computer generated. Since the final movie is a composition of generated and real footage, the entire movie can be labelled frame by frame.

Green screen

Using green screens a reporter can be virtually placed in a different location. By labelling the content, consumers can identify whether the shots were actually taken at the location or not.

3.4. Forgery/media forensics

Insurance fraud

In the context of insurance fraud, an insurer might want to check whether an image used as evidence has not been manipulated.

4. Requirements

Although this is still preliminary, so far requirements in two main categories have been identified: modification description and secure signalling of authenticity information. The sections below list identified requirements for each of these categories.

4.1. Modification Description

- The standard shall provide means to **describe** how the content was generated and/or modified.
- The standard shall provide means to describe the **type of modification**, e.g., no modifications, enhanced, restored, colorized, edited, composed, deep fake, ...
- The standard shall provide means to describe the **purpose of a modification**.
- The standard shall provide means to describe (algorithmically or human) the **likelihood of a modification.**
- The standard shall provide means to describe the **region** where the media was modified.
- The standard shall provide means to **attach provenance information** to media content.
- The standard shall provide means to **keep track of the history of media modifications**.

4.2. Secure linking of modification descriptions and media content

- The standard shall provide means to **restrict access to** metadata.
- The standard shall provide means to **identify** if the media content or associated metadata has been modified.
- The standard shall provide means to record and protect **IPR and provenance** information.
- The standard shall provide means to **identify the source** of the media content.
- The standard shall provide means to **verify the integrity** of the media content.