Vasileios Mezaris · Lyndon Nixon ·
Symeon Papadopoulos · Denis Teyssou
Editors

Video Verification in the Fake News Era

Springer

Video Verification in the Fake News Era

Vasileios Mezaris · Lyndon Nixon ·
Symeon Papadopoulos · Denis Teyssou
Editors

Video Verification in the Fake News Era

Springer

Editors

Vasileios Mezaris
Centre for Research and Technology Hellas
Information Technologies Institute
Thermi, Thessaloniki, Greece

Lyndon Nixon
MODUL Technology GmbH
MODUL University Vienna
Vienna, Austria

Symeon Papadopoulos
Centre for Research and Technology Hellas
Information Technologies Institute
Thermi, Thessaloniki, Greece

Denis Teyssou
Agence France-Presse
Paris, France

ISBN 978-3-030-26751-3 ISBN 978-3-030-26752-0 (eBook)
https://doi.org/10.1007/978-3-030-26752-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The digital media revolution is bringing breaking news to online video platforms, and news organizations often rely on user-generated recordings of breaking and developing events shared in social media to illustrate the story. However, in video there is also deception. In today's 'fake news' era, access to increasingly sophisticated editing and content management tools and the ease in which fake information spreads in electronic networks require the entire news and media industries to carefully verify third-party content before publishing it. This book presents the latest technological advances and practical tools for discovering, verifying and visualizing social media video content, and managing related rights. These are expected to be of interest to computer scientists and researchers, news and media professionals, and even policymakers and data-savvy media consumers.

The book is organized in four main parts. Part I presents the necessary Problem Statement, Part II covers the various Technologies that can contribute to video verification, Part III introduces three complete Applications that integrate several verification technologies and Part IV presents some Concluding Remarks.

Part I Problem Statement

The first step in addressing the problem of 'fake news', or disinformation, is to understand the problem. Chapter 1, 'Video Verification: Motivation and Requirements', attempts to introduce us to the peculiarities of the video verification problem by initially presenting the motivations of those involved in video verification, showcasing the respective requirements and highlighting the importance and relevance of tackling disinformation on social networks. Then, this chapter provides an overview of the state of the art of techniques and technologies for video verification. It also highlights the emergence of new threats, such as the so-called 'deep fakes'. Finally, the chapter concludes by formulating an empirical typology of false videos spreading online.

Part II Technologies

In this part of the book, Chaps. 2 through 8 present in-depth analyses of different technologies that contribute to video verification. Chapter 2, ‘Real-Time Story Detection and Video Retrieval from Social Media Streams’, starts one step before coming to verifying a specific video: it discusses how a journalist can detect emerging news stories online and find videos around that story, which may then require verification. The chapter starts by reviewing the prior research in the area of topic detection, and then presents a keyword-graph-based method for news story discovery out of Twitter streams. Subsequently, it presents a technique for the selection of online videos that are candidates for news stories, by using the detected stories to form a query against social networks. This enables relevant information retrieval at web scale for news-story-associated videos. These techniques are evaluated by observation of the detected stories and of the news videos that are presented for those stories, demonstrating how journalists can quickly identify videos for verification and reuse.

Chapter 3 focuses on ‘Video Fragmentation and Reverse Search on the Web’. Such search is a first and simple, yet often very valuable, means for checking if a video under examination, or a slightly modified version of it, has appeared in previous times in the web and social sphere. Video reuse is in fact the ‘easy fake’: it does not take elaborate editing tools and effort to fake an event in this way; it suffices to fetch some older footage of, e.g. a terrorist attack or a plane crash from the web, and repost it claiming that this is happening right now, right before your eyes. Chapter 3 presents technologies for the fragmentation of a video into visually and temporally coherent parts and the extraction of a representative keyframe for each defined fragment that enables the provision of a complete and concise keyframe-based summary of the video; these keyframes can then be used for performing a fragment-level search for the video on the web. Following a literature survey on the topic, the chapter describes two state-of-the-art methods for video subshot fragmentation—one relying on the assessment of the visual coherence over sequences of frames, and another one that is based on the identification of camera activity during the video recording. It then goes on to present a web application that enables the fine-grained (at the fragment-level) reverse search for near-duplicates of a given video on the web, and evaluation results and conclusions about the effectiveness of these technologies as well as some thoughts on future developments.

Chapter 4, ‘Finding Near-Duplicate Videos in Large-Scale Collections’, sticks to the topic of detecting video reuse for combating the ‘easy fakes’ that we started to deal with in Chap. 3, but views this from a different—and complementary—perspective. In Chap. 3, we discussed web-scale search, which inevitably relies on the reverse image search functionalities that are offered by popular web search engines. The latter provide excellent coverage of the whole web, but on the other hand only allow us to deal with reverse video search in a ‘quick and dirty’ way: by searching for and matching just isolated keyframes. In Chap. 4, we deal with finding duplicate or near-duplicate videos (Near-Duplicate Video Retrieval—NDVR) in our own, closed collections of videos. This means that the coverage of the search is more limited (since we cannot index the whole web in the way that major web search engines do), but on the other hand we can do a much more elaborate and accurate search, because we have full control of the indexing and searching process. Thus, having indexed, for instance, a large number of web videos that show previous terrorist attacks and related content, if we want to check the possible prior use of an (allegedly) new terrorist attack video we can complement the web-scale search of Chap. 3 with a more accurate search in our own collection of such videos. As the main objective of a typical NDVR approach is, given a query video, to retrieve all near-duplicate videos in a video repository and rank them based on their similarity to the query, the chapter starts by reviewing the literature on this topic, and then goes on to present two methods for video-level matching. Extensive evaluation on publicly available benchmark datasets documents the merits of these approaches, and their complementarity to keyframe-based web-scale search.

Chapter 5, 'Finding Semantically-Related Videos in Closed Collections', takes the search for similar video content one step further. When trying to verify a video, and an associated news story, important cues can come from looking at the greater picture: what other videos out there (and thus also in our closed collection of videos, as long as we keep collecting videos related to specific, potentially newsworthy events, such as terrorist attacks) can support (or disprove) the claims made with the help of the specific video in question? For this, besides any near-duplicate videos (as discussed in Chaps. 3 and 4), we would like to detect semantically similar videos. That is, videos showing the same event/actors/activities from a different viewpoint or videos coming from the same source ('channel'—in the broad sense). In other words, we need to be able to organize any content that we collect from the web and social media sources. For this, we discuss two classes of techniques in this chapter: the detection of semantic concepts in video (a.k.a. the annotation of the video with semantic labels) and the detection of logos that are visible in videos and can help us to identify their provenance. Both classes of techniques rely on deep learning (deep neural networks), which is a learning paradigm that is considered to be a key element of Artificial Intelligence (AI). The chapter discusses the state of the art in these two sub-problems of video understanding and presents two techniques developed by the authors of the chapter and their experimental results.

Chapter 6, 'Detecting Manipulations in Video', discusses another fundamental problem related to video verification: can we trust what we see? If an event really unfolded before our eyes, the answer would be yes. But if it is shown on video, how can we assess if the video is an accurate depiction of (some) reality or an alternate 'reality' whose capture in video was only made possible with the help of digital video editing tools? To answer this question, this chapter presents the techniques researched and developed within InVID for the forensic analysis of videos, and the detection and localization of forgeries. Following an overview of state-of-the-art video tampering detection techniques, the chapter documents that the bulk of current research is mainly dedicated to frame-based tampering analysis or encoding-based inconsistency characterization. The authors built upon this existing research, by designing forensics filters aimed to highlight any traces left behind by video tampering, with a focus on identifying disruptions in the temporal aspects of a video. Subsequently, they proceeded to develop a deep learning approach aimed to analyse the outputs of these forensics filters and automatically detect tampered videos. Experimental results on benchmark and real-world data, and analyses of the results, show that the proposed deep-learning-based method yields promising results compared to the state of the art, especially with respect to the algorithm's ability to generalize to unknown data taken from the real world. On the other hand, the same analyses also show that this problem is far from being resolved, and further research on it is in order.

Chapter 7, 'Verification of Web Videos Through Analysis of Their Online Context', continues in the direction of previous chapters, most notably Chap. 5, of looking at the greater picture for verifying a specific video. Contrary (and complementarily) to Chap. 5, though, we are not examining here other related videos that can help debunk the video under examination; instead, we are looking at the online 'context' of this video. The goal is to extract clues that can help us with the video verification process. As video context, we refer to information surrounding the video in the web and/or the social media platforms where it resides, i.e. information about the video itself, user comments below the video, information about the video publisher and any dissemination of the same video through other video platforms or social media. As a starting point, the authors present the Fake Video Corpus, a dataset of debunked and verified UGVs that aim at serving as reference for qualitative and quantitative analysis and evaluation. Next, they present a web-based service, called Context Aggregation and Analysis, which supports the collection, filtering and mining of contextual pieces of information that can serve as verification signals.

Chapter 8, 'Copyright Management of User Generated Video for Journalistic Reuse', concludes this part of the book on technologies, by considering what comes after a newsworthy piece of user-generated video is verified: how can the journalist use it in a legal way? For this, i.e. for reviewing the copyright scope of reuse of user-generated videos usually found in social media, for journalistic purposes, the starting point of this chapter is the analysis of current practices in the news industry. Based on this analysis, the authors provide a set of recommendations for social media reuse under copyright law and social networks terms of use. Moreover, they describe how these recommendations have been used to guide the development of the InVID Rights Management module, focusing on EU copyright law given the context of the InVID EU project.

Part III Applications

Chapter 9, ‘Applying Design Thinking Methodology: The InVID Verification Plugin’, kick-starts the presentation of integrated, complete tools for journalists who what to verify user-generated videos. It describes the methodology used to develop and release a browser extension which has become one of the major tools to debunk disinformation and verify videos and images, in a period of less than 18 months. This is a tool that combines several of the technologies discussed in Chaps. 2 through 8 in a free, easy-to-use package, which has attracted more than 12,000 users worldwide from media newsrooms, fact-checkers, the media literacy community, human rights defenders and emergency response workers dealing with false rumours and content.

Chapter 10, ‘Multimodal Analytics Dashboard for Story Detection and Visualization’, is the second tool presented in this part of the book. The InVID Multimodal Analytics Dashboard is a visual content exploration and retrieval system to analyse user-generated video content from social media platforms including YouTube, Twitter, Facebook, Reddit, Vimeo and Dailymotion. That is, it is not a tool for video verification, but rather a tool for discovering emerging newsworthy stories and related video content, which then may be verified (either using the InVID Verification plugin, presented in the previous chapter; or by directly transferring the video in question, with a click of a button, to the InVID Verification Application that will be discussed in the following chapter). The InVID Multimodal Analytics Dashboard uses automated knowledge extraction methods to analyse each of the collected postings and stores the extracted metadata for later analyses. The real-time synchronization mechanisms of the dashboard help to track information flows within the resulting information space. Cluster analysis is used to group related postings and detect evolving stories, which can be analysed along multiple semantic dimensions—e.g. sentiment, geographic location, opinion leaders (persons or organizations) as well as the relations among these opinion leaders. The result can be used by data journalists to analyse and visualize online developments within and across news stories.

Chapter 11, ‘Video Verification in the Newsroom’, comes as a natural extension of both Chap. 9 (which presented a first tool for the exact same problem: video verification) and Chap. 10, whose Multimodal Analytics Dashboard provides a direct, one-click link for importing newsworthy videos detected with the latter tool into the newsroom’s video verification pipeline. The chapter starts by describing the integration of a video verification process into newsrooms of TV broadcasters or news agencies. The authors discuss the organizational integration concerning the workflow, responsibility and preparations as well as the inclusion of innovative verification tools and services into an existing IT environment. Then the authors present the InVID Video Verification Application or Verification App for short. This can be considered to be an ‘InVID Verification plugin on steroids’, i.e. a more complete and professional application for video verification, which can serve as a blueprint for introducing video verification processes in professional newsroom systems. This verification application, similarly to the InVID Verification plugin, combines several of the technologies discussed in Chaps. 2 through 8.

Part IV Concluding Remarks

The book concludes with Chap. 12, ‘Disinformation: the Force of Falsity’, which departs a bit from the primarily technology-oriented presentation in previous chapters, to engage in a more forward-looking discussion on how can we avoid the proliferation of fake videos, and stop them from spreading over and over again. This final chapter borrows the concept of force of falsity from the famous Italian semiotician and novelist Umberto Eco, to describe how manipulated information remains visible and accessible despite efforts to debunk it. It illustrates, with the help of real-life examples, how search engine indexes are getting confused by disinformation and they too often fail to retrieve the authentic pieces of content, the ones which are neither manipulated nor decontextualized. The chapter concludes with some further thoughts on how to address this problem.

Thessaloniki, Greece — Vasileios Mezaris
Vienna, Austria — Lyndon Nixon
Thessaloniki, Greece — Symeon Papadopoulos
Paris, France — Denis Teyssou
May 2019

Acknowledgements Most of the work reported throughout this book was supported by the European Unions Horizon 2020 research and innovation programme under grant agreement No 687786 ‘InVID: In Video Veritas—Verification of Social Media Video Content for the News Industry’, 2016–2018.

Contents

Part I Problem Statement

Part II Technologies

Part III Applications

Part IV Concluding Remarks

Contributors

Evlampios Apostolidis Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece;
School of Electronic Engineering and Computer Science, Queen Mary University, London, UK

Konstantinos Apostolidis Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Albert Berga Universitat de Lleida, Lleida, Spain

Roger Cozien eXo maKina, Paris, France

Paloma de Barrón Private Law Department, Universitat de Lleida, Lleida, Spain

Daniel Fischl MODUL Technology GmbH, Vienna, Austria

Rolf Fricke Condat AG, Berlin, Germany

Roberto García Computer Science and Engineering Department, Universitat de Lleida, Lleida, Spain

Rosa Gil Computer Science and Engineering Department, Universitat de Lleida, Lleida, Spain

Max Göbel webLyzard technology gmbh, Vienna, Austria

Alexander Hubmann-Haidvogel webLyzard technology gmbh, Vienna, Austria

Ioannis Kompatsiaris Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Giorgos Kordopatis-Zilos Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece;
School of Electronic Engineering and Computer Science, Queen Mary University, London, UK

Foteini Markatopoulou Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Grégoire Mercier eXo maKina, Paris, France

Alexandros I. Metsai Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Vasileios Mezaris Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Lyndon Nixon MODUL Technology GmbH, Vienna, Austria

Symeon Papadopoulos Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Olga Papadopoulou Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Ioannis Patras School of Electronic Engineering and Computer Science, Queen Mary University, London, UK

Gerard Rovira Universitat de Lleida, Lleida, Spain

Tobi Schäfer webLyzard technology gmbh, Vienna, Austria

Arno Scharl webLyzard technology gmbh, Vienna, Austria

Jochen Spangenberg Deutsche Welle, Berlin, Germany

Maria Teixidor Universitat de Lleida, Lleida, Spain

Denis Teyssou Agence France-Presse, Paris, France

Jan Thomsen Condat AG, Berlin, Germany

Markos Zampoglou Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece

Part I
Problem Statement

Chapter 1
Video Verification: Motivation and Requirements

Denis Teyssou and Jochen Spangenberg

Abstract The production and spreading of manipulated videos have been on the rise over the past years, and is expected to continue and increase further. Manipulating videos have become easier from a technological perspective, and can be done with freely available tools that require less expert knowledge and fewer resources than in the past. All this poses new challenges for those who aim to tackle the spreading of false, manipulated or misleading video content. This chapter covers many of the aspects raised above. It deals with the motivations of those involved in video verification, showcases respective requirements and highlights the importance and relevance of tackling disinformation on social networks. Furthermore, an overview of the state of the art of available techniques and technologies is provided. The chapter then describes the emergence of new threats like so-called 'deep fakes' created with the help of artificial intelligence. Finally, we formulate an empirical typology of false videos spreading online.

1.1 Introduction

The rise of the smartphone and the convergence of affordable devices, powerful camera technology, ubiquitous mobile Internet access and social media platforms have enabled a massive growth of citizen and eyewitness-powered news coverage. As Malachy Browne, Senior Story Producer at the New York Times stated some years ago: 'one byproduct of this is an enormous amount of video being uploaded and shared every minute, every hour' [1].

Indeed, although Facebook has been facing criticism and regular complaints from advertisers for allegedly failing to disclose errors in its video viewership metrics [2], there is no doubt that video consumption, especially on mobile devices, is steadily

D. Teyssou (✉)
Agence France-Presse, Paris, France
e-mail: denis.teyssou@afp.com

J. Spangenberg
Deutsche Welle, Berlin, Germany
e-mail: jochen.spangenberg@dw.com

V. Mezaris et al. (eds.), *Video Verification in the Fake News Era*,
https://doi.org/10.1007/978-3-030-26752-0_1

on the rise. YouTube statistics of October 2018 reports more than one billion users of its video platform, with one billion hours watched daily, 70% of them on mobile devices.[1]

'The revolution in information technology is not over, and the volume of newsworthy user-generated content will only grow. Journalists have a new responsibility—to quickly gather, verify and ascertain the usage rights of UGC.[2] Traditional values of investigation apply, but a new skillset is required for media such as video', predicted Malachy Browne in the above-mentioned Verification Handbook, edited by Craig Silverman and published by the European Journalism Centre in 2014. He was right!

In journalism, the verification of information has been a commonplace activity for decades. It is part of the craft of journalism for everyone who takes the profession seriously. What is relatively new, however, for both newsrooms and journalists alike, is the rather systematic verification of third-party eyewitnesses digital information such as images or videos that are being shared and distributed via social platforms.

For media organisations, this meant 'crossing a Rubicon', as stated by the former Director of BBC News—now Professor of Journalism at Cardiff University—Richard Sambrook [3] during the London bombings of July 2005. 'Trapped in the London Underground, witnesses used their mobile phones to take pictures and make video recordings of the events as they unfolded. Unable to deploy journalists to the bombing sites, the BBC relied on eye-witness accounts and survivors' stories', as media scholar Valerie Belair-Gagnon pointed out [4].

Speaking on the same matter Helen Boaden, Director of BBC News from 2004 until 2013, said that during the London bombings, the scope and reach of user-generated content was greater than ever before. She stated that within 24 h, the Corporation received more than 1,000 pictures and videos, 3,000 text messages and 20,000 e-mails [5]. The amount of user-generated content in news stories was unprecedented: 'Twenty-four hour television was sustained as never before by contributions from the audience', explained Belair-Gagnon [4].

As hinted above: the digitisation of information, and especially the emergence of social networks, has resulted in fundamental changes when it comes to the gathering and spreading of information. Nowadays, it is possible (at least technically) for anyone to publish and distribute digital content with ease and speed to a potentially worldwide audience, and reach, potentially, millions of people. In turn, the so-called 'legacy media' (news agencies, traditional newspapers, magazines, broadcasters, etc.) are no longer the exclusive gatekeepers who decide what is being circulated to a wider audience, and what is not.

All this poses new challenges and brings up numerous issues that need to be addressed both by established and emerging media organisations as well as newsrooms, independent journalists and fact-checkers. The challenges are not an exclusive domain of the journalistic profession though: for example, human rights workers and emergency response personnel are also confronted with related tasks when it comes to assessing whether material that is being circulated digitally is right or wrong.

[1] https://www.youtube.com/intl/en/yt/about/press/.

[2] User-generated content.

To start with, let us outline some of the challenges that exist when it comes to the verification of digital content, focusing on video. According to sources such as the Verification Handbook [6] or the First Draft Visual Verification Guide for Video [7], issues include the following:

1. Identifying the video provenance (is it the original version of the video? Who captured the video?).
2. Verifying the source (who is the producer/uploader, and what can we find out about him/her?).
3. Locating the video (where was the video captured?).
4. Verifying the date (when was the video captured?).
5. Identifying what the video shows (why was the video captured?).

The above list provides a selection of aspects that need to be considered and addressed in the context of verification of digital content. In this introductory chapter, we will focus on a variety of aspects regarding the verification of digital video content. In particular, we will look at incentives for video verification, and look at corresponding needs and requirements, as well as solutions. Furthermore, we will portray some of the actions and activities that are being undertaken in the sphere. All this is to lay some foundations for what is to come in subsequent chapters, while we approach the issue from a broader perspective in this introductory chapter.

1.1.1 Who Verifies Video, Which Groups Are We Dealing With and Why Is This Important?

While anybody with an interest in video content is a potential verifier of such content, here we are limiting ourselves to a list of five groups of professionals (excluding the military, intelligence, law and enforcement sectors), namely:

1. Journalists (working in media organisations or on their own);
2. Fact-checkers (from media organisations, associations and fact-checking groups, e.g. from civil society, NGOs, etc.);
3. People working in the human rights sector (e.g. to document war crimes, crimes against humanity and such like);
4. People working in the emergency response sector (e.g. people working for relief agencies after natural disasters, accidents or catastrophes);
5. Media education scholars (people working on media literacy projects and respectively research, teach and lecture on the subject).

For all the above stakeholder communities, the verification of video content and dealing with related consequences are becoming more and more important for a variety of reasons. As also indicated above in the introduction, the consumption and popularity of online video content with consumers, in general, have been steadily on the rise, and is expected to grow further.

Next, most people these days are equipped with smartphones that can record, capture and share video content in steadily increasing quality. Then, data tariffs for uploading/sharing as well consuming online video content have gone down steadily as well, and are likely to go down even further ('flat fees', new networks from LTE to 5G, etc.).

All this means that what has happened in the text and image sector before will increasingly apply to video, too. And it has huge potential due to the attractiveness of video content. In other words, we will see more and more (high quality) video content being shared as well as consumed by ordinary citizens in the years to come. All this material, obviously, also has the potential to play an increasing role for

1. the newsgathering and reporting process (in the field of journalism);
2. the documenting of human rights abuses (in the human rights sector);
3. other sectors like the emergency response field.

1.1.2 *The Value of Video Analysis*

Having at your disposal (verified) video material can have tremendous value: it is a great asset for newsgathering and reporting, it can detect human rights abuses and violations, while it can also play a life-saving role in guiding emergency response operations.

Video material captured on any device and shared via social networks or by direct communication has thus opened up new avenues for all the above sectors and domains.

As stated previously, being able to verify and authenticate digital material that is being circulated (e.g. on social networks) is of paramount importance in order to avoid the spreading of false information and judgements or assessments based on the wrong foundations.

The availability of so-called 'Open Source Intelligence' (OSINT)[3] tools have particularly changed the verification sphere and respective possibilities profoundly.

The example below[4] is just one of many examples in which OSINT tools and meticulous desk research have exposed the truth behind a crime and human rights violation, which made headline news in numerous media outlets.

In the above case of the murder of two women and two young children in Cameroon, researchers/journalists of the BBC, Bellingcat and Amnesty International

[3] By OSINT we mean information and data that is gathered from public (or open) sources. OSINT tools are digital tools (software, platforms) that are freely available and facilitate research and investigations, such as satellite imagery to identify or cross-check a particular geographic location.

[4] ATTENTION: the video that is linked here includes graphic imagery that some people may find disturbing and/or painful to watch. Do not view unless you are prepared accordingly. In case of viewing the video (this also applies to other disturbing/traumatic imagery) and this having negative effects on you, do not hesitate to seek professional help.

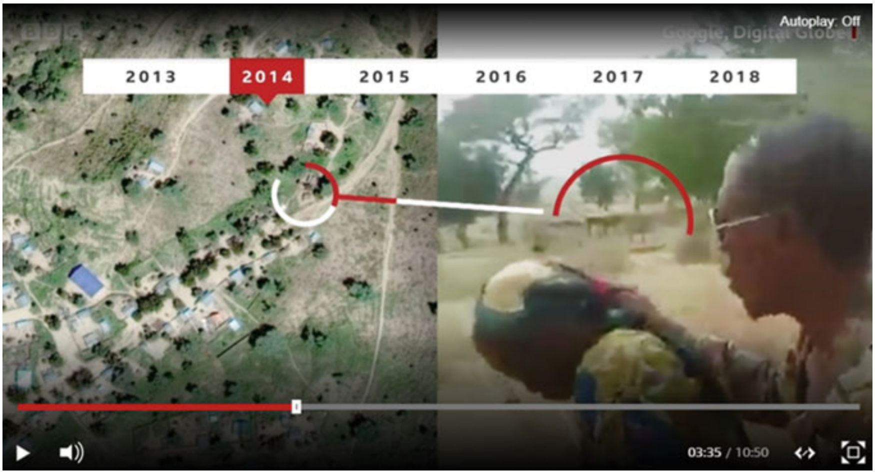

Fig. 1.1 Screenshot of a BBC Africa Eye documentary in Cameroon; 24 Sep. 2018. Video available on YouTube: https://www.youtube.com/watch?v=XbnLkc6r3yc

supported by others were able to identify where exactly the crime took place, when this happened, and who was involved (and largely responsible) (Fig. 1.1).

Doing all the research using OSINT tools like satellite imagery, geo-tools or software that calculates the level of the sun at a given time anywhere in the world (here: suncalc), and other aids took considerable time and effort, as reporters dealing with the case have repeatedly pointed out.[5]

This is one use case in which technology significantly supports (and helps enable) the work of journalists as well as human rights personnel.

1.1.3 *Tools/Platforms 'Made for Video Verification'*

To date (in the first half of 2019, at the time of writing) there are not many tools available on the market that have been designed and developed with the specific purpose of verifying videos. Rather, people who are trying to determine the veracity of digital online videos (especially user-generated video) use a variety of different tools and platforms in combination, such as satellite imagery (e.g. Wikimapia, Google Earth Pro, Terra Server), reverse image search (e.g. Google Images, Tineye,

[5] See, for example, an interview with one of the reporters of the story, Aliaume Leroy, on CBC Radio Canada. The site includes a link to an 8-minute audio interview with Leroy about the team's work on the case: https://www.cbc.ca/radio/asithappens/as-it-happens-Monday-edition-1.4836241/how-bbc-africa-uncovered-the-story-behind-an-execution-video-of-women-and-children-1.4836248.

Yandex), geo-based search that helps in finding out more about particular locations (e.g. Echosec, Liveuamap, GeoNames), tools that provide more information about people/sources (e.g. Pipl, various directories), image forensic tools (e.g. FotoForensics and Forensically), metadata /Exif data viewers (e.g. Jeffrey's Exif Viewer) and so on.

A very comprehensive list of tools, called online investigation toolkit is maintained by investigative platform Bellingcat[6] and Christiaan Triebert[7] as a shared document[8] that is frequently updated by the user community.

One exception, i.e. a (basic) platform that has been built specifically for video analysis and verification, is Amnesty International's YouTube Data Viewer, launched in mid-2014. In a nutshell, it enables users to key in the URL of a YouTube video, and then let it automatically extract the correct upload time and all thumbnails associated with the video. The thumbnails can then be used to perform a reverse image search to see if identical images exist elsewhere online, especially with earlier upload times.

According to Christoph Koettl, Amnesty International's main architect behind the YouTube Data Viewer,[9] 'Many videos are scraped, and popular videos are re-uploaded to YouTube several times on the same day, so having the exact upload time helps to distinguish these videos … and a reverse image search is a powerful way of finding other/older versions of the same video'.[10]

The YouTube Data Viewer thus became part of the standard kit of online verification experts and investigators dealing with assessing the credibility of online videos. It is however limited to the YouTube platform, and the thumbnails provided by this platform can be changed by the uploaders.

Then came InVID with its Verification Plugin. Obviously, the authors of this chapter are biased, as we have been involved in its development (with one of the authors, Denis Teyssou of AFP, being the main driver behind its development). Nevertheless, based on feedback obtained from the verification community,[11] it can be stated that the InVID Verification Plugin with its added features and functionalities nowadays provides journalists and human rights investigators with a very powerful and useful tool for video analysis. How it looks like exactly, and what functionalities are included, will be portrayed in Chap. 9 of this book.

[6] https://www.bellingcat.com/.

[7] https://twitter.com/trbrtc.

[8] https://docs.google.com/document/d/1BfLPJpRtyq4RFtHJoNpvWQjmGnyVkfE2HYoICKOGguA/edit.

[9] See https://citizenevidence.amnestyusa.org/.

[10] Christoph Koettl, quoted on Poynter in the article Amnesty International launches video verification tool, website, by Craig Silverman, 8 July 2014. https://www.poynter.org/news/amnesty-international-launches-video-verification-tool-website.

[11] Some user feedback has been collected in Twitter Moments here: https://twitter.com/i/moments/888495915610275840.

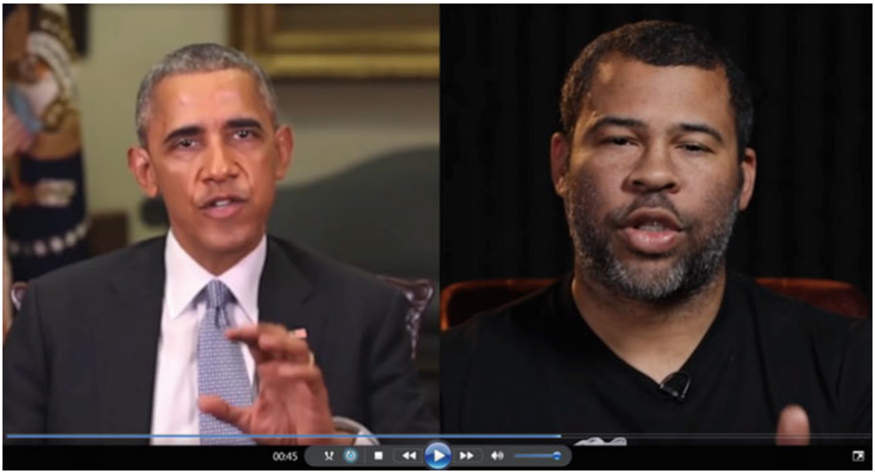

Fig. 1.2 Screenshot of a deep fake video by BuzzFeedVideo, 17 April 2018. Video available on YouTube: https://www.youtube.com/watch?v=cQ54GDm1eL0

1.1.4 A New Challenge: 'Deep Fakes'

Sadly, methods and ways to manipulate digital content (including video) and mislead audiences are becoming more and more sophisticated. A relatively new concept of faking (or manipulating) video is called 'deep fakes'.

Deep fakes involve the use of algorithms and artificial intelligence for face re-enactment, face swapping, lip-syncing and the synthetic mimicking of voices. In other words, movements of the mouth are transferred from one person (e.g. an actor) to the face and facial expressions of another person (in most cases, a celebrity). The same goes for voice. To fully understand this, it helps a lot to see deep fakes in action (as in the video example below, in which the words and lip movements of actor Jordan Peele are 'transported' to the face of Barack Obama) (Fig. 1.2).[12,13]

Researchers from the University of Erlangen-Nuremberg and the Max-Planck-Institute for Informatics in Europe, from Washington and Stanford University in the USA, in turn, demonstrated how comparatively easy it is to make a person say something convincingly—that is difficult to detect—that he or she never uttered in

[12] Deep fakes are particularly popular on Reddit. Apps (such as FakeApp) exist that allow users with little to no technical knowledge to create video manipulations—some of it amusing and rather creative. Inserting the actor Nicolas Cage into deep fake videos, in turn, has almost become its own sub-genre. Other usage 'scenarios'—besides celebrity manipulations—are (child) pornography, revenge porn, extortion and mis-/disinformation of various types.

[13] In the referenced video, Barrack Obama's mouth is lip-synced automatically. The words that seem to be spoken by Obama come from actor and film-maker Jordan Peele, who does an impression of the former US President. Obama (aka Peele) says things like 'This is a dangerous time. Moving forward, we need to be more vigilant with what we trust from the internet. This is a time when we need to rely on trusted news sources'. https://www.youtube.com/watch?v=cQ54GDm1eL0.

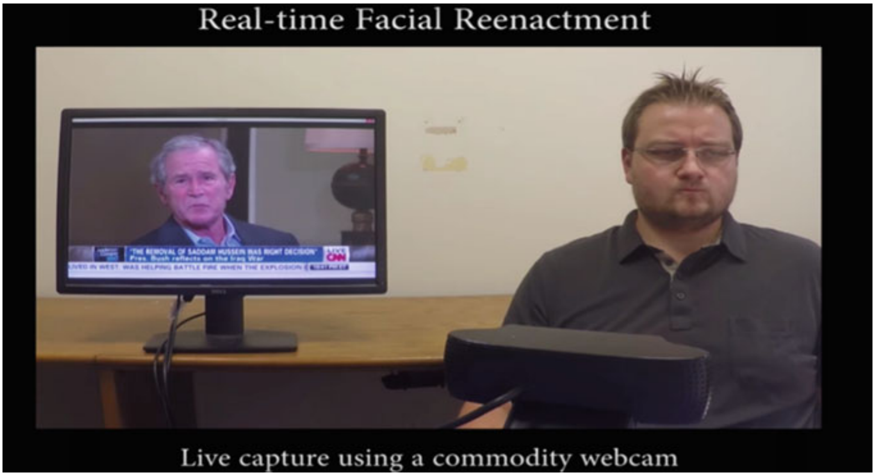
Real-time Facial Reenactment
"THE REMOVAL OF SADDAM HUSSEIN WAS RIGHT DECISION"
Pres. Bush reflects on the Iraq War
CNN
Live capture using a commodity webcam

Fig. 1.3 Real-time facial re-enactment. Demonstrated by the University of Erlangen-Nuremberg, Max-Planck-Institute for Informatics, presented at CVPR 2016. Video available on YouTube: https://www.youtube.com/watch?v=ohmajJTcpNk

real life. All it takes is enough original video footage of that person (which exists in abundance with, e.g. prominent politicians or celebrities), train the respective algorithms accordingly, and set up the system and actors/speakers/'modellers' as required. How this is done is demonstrated in the video (Fig. 1.3).

Obviously, if manipulated videos like the ones showcased above do the rounds, and technology to alter them becomes even more sophisticated (and easily available), this will create even more suspicion with everything we watch. The dangers are obvious, especially as many would say that 'if something has been caught on (video) tape, it must be real'. To put it differently, sophistically manipulated videos will create a new kind of suspicion about everything we watch. Politicians as well as others with vested interests are likely to exploit this. At its worst, deep fakes can damage brands and careers, manipulate and impact the political process, and go as far as destroy people's lives or even cause wars.

This raises a whole set of questions, such as: will technical advances that allow for the fabrication of videos with ease destroy one of our remaining beliefs that we cannot even trust our eyes and ears any longer? Is there a need for (and chance of) a coordinated international effort to agree on norms of behaviour when it comes to this kind of manipulation technology?

On the other hand, it becomes obvious that what has been outlined above makes the need to develop and supply tools and technologies that allow for the detection of manipulations more important than ever.

Furthermore, it is not only technology that should be at the centre of attention here: what about regulation, ethics and a sort of—at least to some extent—common agreement about what we label and commonly refer to as 'reality'?

Sam Dubberley, Special Advisor to Amnesty International's Evidence Lab, adds another dimension. Asked about what worries him when it comes to verifying and analysing information, chasing the truth and what keeps him awake at night, Dubberley stated in an interview with one of the authors of this paper that it is 'the people who build these [deep fake] tools are doing so without an ethical framework'.[14] So apart from developing tools and services that help in detecting fakes and manipulations (as we do in InVID), it may be time to also bring this to the highest levels of the political agenda. Whether agreements on a supra-national level can be reached remains doubtful. It is nevertheless vital that everything possible is done in the fight against manipulations and distortion. Nothing less than democracy is at stake!

Although deep fakes were not part of the scope of the InVID project, we applied some of the technologies (especially fragmentation and reverse image search to be discussed in Chap. 3) to tackle this specific problem. We showed, for instance, that a deep fake video of Donald Trump allegedly asking Belgian people to withdraw from the Paris climate agreement (made by the Flemish Socialist Party in a campaign to mobilise citizens for the environment) was made by manipulating a previous video of Mr. Trump announcing strikes against Syria in retaliation for the use of chemical weapons against civilians by the Syrian regime, in a public address from the White House, on the 13 April 2018.[15]

1.1.5 *Typology of Fake Videos*

There are a variety of ways in which videos can be manipulated, or—in other (better) words—ways in which the audience can be deceived with video content.

To start with, a video can be actively manipulated. For example, video frames can be added or deleted, components can be inserted (e.g. via an overlay/insert) or even created by a computer as we just saw above. Examples abound,[16] from the Pope apparently performing a 'magic table cloth trick' (Fig. 1.4) during a visit to the US[17]

[14] Sam Dubberley, Special Advisor to Amnesty International's Evidence Lab (and an InVID project reviewer), interviewed by Jochen Spangenberg on 27 September 2018.

[15] https://www.slideshare.net/InVID_EU/invid-at-ifcn-global-fact-v.

[16] Here is a selection: https://youtu.be/ccENfRThXOk.

[17] Watch the (fabricated) video sequence here: https://youtu.be/1KFj6b1Xfe8?t=14. It was created for the US-based 'The Ellen Show'. Here https://youtu.be/ABy_1sL-R3s you can see the original video footage next to the manipulated version.

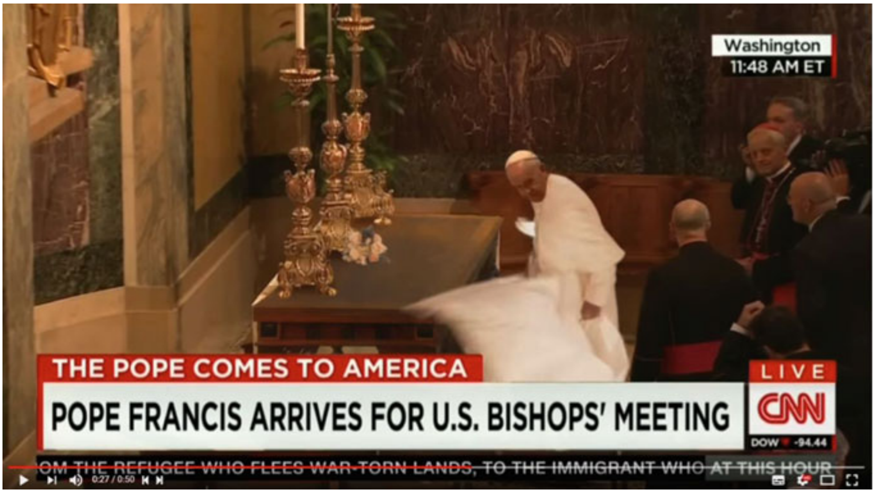
Washington
11:48 AM ET
THE POPE COMES TO AMERICA
POPE FRANCIS ARRIVES FOR U.S. BISHOPS' MEETING
LIVE
CNN
DOW -94.44
OM THE REFUGEE WHO FLEES WAR-TORN LANDS, TO THE IMMIGRANT WHO AT THIS HOUR
0:27 / 0:50

Fig. 1.4 Screenshot of manipulated video from the satirical ‘The Ellen Show’, showing Pope Francis miraculously ‘pulling the cloth’

to an eagle apparently snatching a baby from a pram and ‘taking it up high’[18] or a snowboarder being chased by a bear while going down a slope in Japan.[19]

Another way of deceiving audiences is by taking a video out of context, and distributing it with misleading captions. In such cases, the video is not actively manipulated, but instead presented as being something that it is not. For example, whenever there is a plane crash, videos (and images) from previous plane crashes very quickly make the rounds on social networks or certain websites. As video footage of plane crashes is still relatively rare, it is usually the same footage that appears again and again, whenever there is such an accident.[20]

Polarising topics like migration are often used for disinformation campaigns against migrants: a Russian news item about a drunk man assaulting nurses in a

[18] See the video here: https://youtu.be/Xb0P5t5NQWM. The video is the product of three students studying animation and digital design at Montreal’s Centre NAD, as part of a digitised class project, not a real snatching. See also NPR update: ‘Eagle Snatches Kid’ Video Makers Admit Hoax, by Mark Memmott, 19 December 2012. https://www.npr.org/sections/thetwo-way/2012/12/19/167610327/eagle-snatches-kid-video-the-debunking-begins?t=1537977241104.

[19] See https://youtu.be/vT_PNKg3v7s for the video. Debunks here: https://youtu.be/lXmk-BxXXhY, https://www.nationalgeographic.com/adventure/adventure-blog/2016/04/12/bears-really-do-chase-skiers-but-this-video-is-fake/ and https://www.snopes.com/fact-check/snowboarder-girl-chased-by-bear/.

[20] This video https://youtu.be/A-8Ig67O3ck of a plane crash of a US cargo Boeing 747 near Bagram Airfield, Afghanistan, in May 2013 is one such video. Whenever there is a new plane crash, this video (and others that capture previous plane crashes—of which there are not that many) are being shared and distributed pretending to be of the current plane crash. WARNING: some people may find watching the video distressing or disturbing, as it portrays a real plane crash that resulted in the loss of lives.

Novgorod hospital in February 2017 started to circulate 1 month later in various countries (France, Italy, Spain, Belgium, Turkey), copied and widely distributed on Facebook and Twitter, as allegedly being an assault perpetrated by a migrant in those respective countries.[21] More recently, an amateur video taken by a Czech tourist in Crete (Greece) was at the origin of a conspiracy theory accusing mainstream media of staging migrant arrivals on European beaches, while the scene was actually the making of a film about the 1922 Greek exodus from Asia Minor.[22]

The challenges for those working with video are thus to

1. detect fakes and manipulations as quickly and accurately as possible;
2. do so using traditional verification techniques (from direct contact with eyewitnesses to counterchecks and such like);
3. have tools at their disposal to assist in the process of digital content/video verification;
4. familiarise themselves with tools and techniques;
5. integrate the required activities into respective workflows;
6. have available the respective infrastructure as well as (organisational) support mechanisms;
7. be clear about and aware of possible effects on mental well-being (having psychological support is vital, and learning about respective coping mechanisms something that everyone dealing with user-generated video should be familiar with).[23]

Our work in InVID helped us to draw up the following typology of fake video content:

1. Decontextualised videos (unchanged or almost unchanged with high level of similarity, including low-quality copies for clickbait purposes);
2. Decontextualised videos altered (cut in length to one or several fragments of the original video, or cropped to remove, e.g. a timestamp in a CCTV camera footage);
3. Staged videos (e.g. produced on purpose by a video production company);
4. Tampered videos (through editing software to remove, hide, duplicate or add some visual or audio content);
5. Computer-Generated Imagery (CGI) including deep fakes (false images generated by artificial intelligence) made entirely from a computer or mixed with a blend of previous footage.

[21] https://observers.france24.com/en/20170323-no-internet-man-hitting-nurse-not-migrant.

[22] https://factcheck.afp.com/no-not-video-journalists-staging-migrants-drowning.

[23] Dealing with traumatic imagery and gruesome material is not the focus of this chapter. Nevertheless, it is of vital importance in this context and deserves utmost attention in order to avoid trauma. Equally important: to learn and develop respective coping mechanisms. For more on the dealings with digital material and its possible psychological effects see, for example, the work of Eyewitness Media Hub (now incorporated into First Draft News) and the Dart Center. A very useful read is the study [8].

In the chapters that follow, we will review the technologies mobilised to discover newsworthy user-generated videos and to verify them. These include video fragmentation and reverse image search on the Web, an approach giving good practical results to tackle decontextualised videos, techniques for finding duplicate or near-duplicate videos in closed collections, techniques for detecting digital manipulations within the videos and methods for the contextual analysis of the videos' surrounding metadata. Then, we will focus on the current state of user-generated video copyright management. These technology-oriented chapters will be followed by a detailed presentation of the complete applications developed within the InVID project: the Verification Plugin, the multimodal analytics dashboard and the InVID Verification Application tailored for newsrooms. We will conclude with some findings on the current state of disinformation spreading on social networks and some directions for future research to increase the efficiency of dealing with it.

References

1. Browne M (2014) Verifying video. European Journalism Center, EJC
2. Vranica S (2018) Advertisers allege Facebook failed to disclose key metric error for more than a year. Wall Street J. https://www.wsj.com/articles/advertisers-allege-facebook-failed-to-disclose-key-metric-error-for-more-than-a-year-1539720524
3. Sambrook R (2005) Citizen journalism and the BBC. Verification handbook. https://niemanreports.org/articles/citizen-journalism-and-the-bbc/
4. Belair-Gagnon V (2015) Social media at BBC news: the re-making of crisis reporting. Routledge
5. Boaden H (2008) The role of citizen journalism in modern democracy. http://www.bbc.co.uk/blogs/theeditors/2008/11/the_role_of_citizen_journalism.html
6. Silverman C et al (2014) Verification handbook. European Journalism Center, EJC
7. First Draft News. Visual verification guide for video. https://firstdraftnews.org/wp-content/uploads/2017/03/FDN_verificationguide_videos.pdf
8. Dubberley S, Griffin E, Mert Bal H Making secondary trauma a primary issue: a study of eyewitness media and vicarious trauma on the digital frontline. https://eyewitnessmediahub.com/research/vicarious-trauma

Part II
Technologies

Chapter 2
Real-Time Story Detection and Video Retrieval from Social Media Streams

Lyndon Nixon, Daniel Fischl and Arno Scharl

Abstract This chapter introduces two key tools for journalists. Before being able to initiate the process of verification of an online video, they need to be able to determine the news story that is the subject of online video, and they need to be able to find candidate online videos around that story. To do this, we have assessed prior research in the area of topic detection and developed a keyword graph-based method for news story discovery out of Twitter streams. Then we have developed a technique for selection of online videos which are candidates for news stories by using the detected stories to form a query against social networks. This enables relevant information retrieval at Web scale for news story-associated videos. We present these techniques and results of their evaluations by observation of the detected stories and of the news videos which are presented for those stories, demonstrating state-of-the-art precision and recall for journalists to quickly identify videos for verification and re-use.

2.1 Introduction

The starting point for any journalist before any content verification will be to find online content that purports to show events that are in relationship with the news story. In this age of 24 h news channels and online breaking news, sometimes the starting point is actually asking what are the current news stories and choosing which one would be relevant for reporting, necessitating the identification and verification of potentially relevant online media. The InVID solution has therefore also considered

L. Nixon (✉) · D. Fischl
MODUL Technology GmbH, Vienna, Austria
e-mail: nixon@modultech.eu

D. Fischl
e-mail: daniel.fischl@modul.ac.at

A. Scharl
webLyzard technology gmbh, Vienna, Austria
e-mail: scharl@weblyzard.com

V. Mezaris et al. (eds.), *Video Verification in the Fake News Era*,
https://doi.org/10.1007/978-3-030-26752-0_2

the pre-verification step, since no-one can easily verify online video purporting to show a news story without being first able to find suitable candidate videos.

A Web dashboard (see Chap. 10) will provide the results of the pre-verification analysis to the user: the automatic identification of news stories out of social media streams and the ranked listing of relevant online video postings associated to that news story. In this chapter, we will explain how this has been implemented.

Traditionally, newswires have been responsible for identifying and spreading the news of a new event or story to the global news institutions—newspapers and radio/TV stations. The speed in which a newswire would pick up on and report breaking news was directly related to the presence of their journalists at the news event, since media professionalism dictated that there was an independent, trustworthy source for news reporting. The Web, and especially social media, has rapidly changed how news is reported in the few decades it has existed, as it has given a globally accessible voice to any person who wants to announce something as breaking news. Journalists at professional news organizations are now faced with a much broader range of sources of potential news, much quicker in reaction than the traditional newswires but potentially uncorroborated.

Identifying a news story in social media is advantageous to journalists as it can point them to previously unknown information, indicate eyewitnesses who could potentially be interviewed and corroborate events, and link to eyewitness media, i.e., photos or videos taken as the news story took place. Since the journalists themselves cannot be present at every news event, especially when they happen in a previously unannounced fashion, nor arrive in a timely manner when they have been informed of an event, eyewitness media is becoming ever more significant in news reporting. However, they first must identify the stories that are happening and could be relevant for their news reporting cycle. While eyewitness reporting on social media means a news story can be announced there very shortly after its occurrence, the bulk and variety of social media content means such emerging news stories can be easily lost in the noise. Furthermore, the lack of secure controls regarding who is posting social media about a news story or why (e.g., political motivation) leads to the growing problem of "fake news"—both "misinformation" (possibly unintended false or inaccurate information) and "disinformation" (deliberately published false or inaccurate information, specifically intended to deceive). Hence, the discovery of social media from non-professional news sources is intricately linked to the task of content verification.

The SNOW 2014 Data Challenge [1] was the first challenge to tackle newsworthy topic detection. It confirmed that it is a challenging task: the top F-score of the competing solutions was only 0.4 (precision: 0.56, recall: 0.36). To develop an effective tool for news story detection in InVID, we had to consider:

1. Data acquisition through a real-time news monitor: we chose the Twitter Streaming API.
2. Data selection and preprocessing mechanism including spam/noise filtering.
3. Content modeling (e.g., from the state of the art: bag-of-words model, n-grams, TF-IDF) including semantic enrichment and (cross-lingual) linking/disambiguation via a knowledge base.