Anais da Academia Brasileira de Ciências (2019) 91(Suppl. 1) e20180149
(Annals of the Brazilian Academy of Sciences)
Printed version ISSN 0001-3765 / Online version ISSN 1678-2690
http://dx.doi.org/10.1590/0001-3765201820180149
www.scielo.br/aabc | www.fb.com/aabcjournal

# Counteracting the contemporaneous proliferation of digital forgeries and fake news

**ALEXANDRE FERREIRA[1], TIAGO CARVALHO[2], FERNANDA ANDALÓ[1] and ANDERSON ROCHA[1]**

[1]Institute of Computing, University of Campinas (Unicamp),
Av. Albert Einstein, 1251, 13083-852 Campinas, SP, Brazil
[2]Instituto Federal de São Paulo (IFSP), Av. Comendador Aladino Selmi, s/n,
13069-901 Campinas, SP, Brazil

*Manuscript received on February 13, 2018; accepted for publication on September 16, 2018*

**How to cite:** FERREIRA A, CARVALHO T, ANDALÓ F AND ROCHA A. 2019. Counteracting the contemporaneous proliferation of digital forgeries and fake news. An Acad Bras Cienc 91: e20180149. DOI 10.1590/0001-3765201820180149.

**Abstract:** Fake news has been certainly the expression of the moment: from political round table discussions to newspapers to social and mainstream media. It is everywhere. With such an intense discussion and yet few effective ways to combat it, what can be done? Providing methods to fight back even the least harming hoax is a social responsibility. To look for authenticity in a wide sea of fake news, every detail is a lead. Image appearance and semantic content of text and images are some of the main properties, which can be analyzed to reveal even the slightest lie. In this vein, this work overviews some recent methods applicable to the verification of dubious content in text and images, and discusses how we can put them together as an option to curb away the proliferation of unverified and phony "facts". We briefly present the main idea behind each method, highlighting real situations where they can be applied and discussing expected results. Ultimately, we show how new research areas are working to seamlessly stitch together all these methods so as to provide a unified analysis and to establish the synchronization in space and time — the *X-Coherence* — of heterogeneous sources of information documenting real-world events.

**Key words:** Digital Forensics, fake news, visual analysis, semantic analysis, X-Coherence, DéjàVu.

## INTRODUCTION

In a scenario where fake news is in every corner trying to convince readership that the most unlikely fact is an authentic truth, it is really difficult to tell apart genuine from phony facts. Cases such as the 2016 USA presidential election, when communication vehicles (from social to mainstream media) overwhelmed people with an astonishing amount of fake news stories, are a red alert to collateral damages caused by this kind of behavior. In a work focused on the 2016 USA elections, Allcott and Gentzkow (2017) indicated that a number of commentators suggested Donald Trump would not have been elected president were it not for the influence of fake news. This statement is based on a very specific series of facts: (1) 62% US adults get news from social media; (2) the most

Correspondence to: Anderson Rocha
E-mail: anderson.rocha@ic.unicamp.br
ORCid: http://orcid.org/0000-0002-4236-8212

∗ Contribution to the centenary of the Brazilian Academy of Sciences.

popular fake news stories were more widely shared on Facebook than the most popular mainstream news stories; (3) many people who see fake news stories report they believe them; and (4) the most discussed fake news stories tended to favor Donald Trump over Hillary Clinton.

In Brazil, according to Biller (2018), the combination of political polarization and passion for social media offers fertile ground for fake news in the run-up to the 2018 general elections, leading to results that could set the Brazilian society on a backward path or even favor the appearance of Fascist movements.

Furthermore, the broadcast of sensitive content, mainly pornographic, through the Internet is as dangerous as fake news. The situation is further complicated when fake news meets pornographic content and both become entangled. If 2016 was the year of "post-truth"[1] – further consolidated in 2017 as the year of fake news – this year will probably be the year of DeepFakes (Morris 2018). This new form of content falsification makes use of deep learning algorithms (from the Artificial Intelligence field) to produce convincing face-swapping videos, in particular for replacing – in movie sequences – faces of porn stars with mainstream celebrities. Although the term is new, it is swiftly spreading out most likely propelled by a user-friendly and controversial (Farokhmanesh 2018) application called FakeApp, which allows anyone to create this kind of fake videos effortlessly.

But how can we fight back situations involving the broadcast of fake news? In recent years, scientists have been developing research in the field of *Digital Forensics* to prevent or to aid the investigation of such problems. Differently from the Information Security field, whose focus is on aspects concerning system's violation and unauthorized system access, *Digital Forensics* targets the development and deployment of methods for digital document analysis (images, videos, audio, and text), in order to evaluate, among other aspects, their authenticity.

Aiming at discussing possible ways of facing the aforementioned problems, this work brings an overview of recent research in an effort to combat fake news. We start with a discussion targeting document, image and text analysis. We then move to the study of methods for video verification and textual authorship detection. As we evolve in the presentation of ideas, we examine the rationale behind each method, highlight real situations where it could be applied, and discuss expected results. For more complex cases, when it is necessary to understand how a set of transformed images are related to each other, we review the image phylogeny framework setting the stage for the final part of the paper. Ultimately, we show how new research areas are working to provide more than a basal disjoint source analysis for a given situation, allowing the synchronization, in space and time (*the X-Coherence*), of heterogeneous sources documenting and describing real-world events, leading to a thorough understanding of facts.

## METHODS FOR DOCUMENT ANALYSIS

Social media platforms, such as Facebook, Twitter, and Instagram, have been revolutionizing the way people communicate with each other. They are designed to enable users to interact, collaborate, and share anything they want in the process of creating as well as consuming content (Obar and Wildman 2015). Notwithstanding, users of these engaging platforms can easily, sometimes inadvertently, consume and broadcast dubious and sensitive content, establishing grounds for fake news proliferation.

In this complex and fast-paced setup, how to discern between pristine and fake content? How to

[1] According to the Oxford Dictionary, it refers to "circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief".

evaluate whether certain pieces of text, images, and videos are factual? Computational methods covered in this section can aid the task in different ways. In this section, we first present literature related to the authenticity evaluation of digital images and printed text. Following, the described methods are able to detect fake videos and associate textual authorship, by taking content semantics into consideration.

## FAKE IMAGE DETECTION

When analyzing suspicious documents, a suitable starting point is to evaluate the authenticity of images therein. In the digital era, images are particularly easy to manipulate using commonplace image editing software suites, such as Adobe Photoshop and Gimp, leading to an astonishing number of fake images reaching us everyday through the Internet. Moreover, some studies (Nightingale et al. 2017, Schetinger et al. 2017) suggest that human beings are notably limited in the task of distinguishing original and manipulated images, even when presented with photometric or geometric inconsistencies.

Image forgery can be classified roughly into two main groups: *image splicing*, which refers to situations in which parts of two or more images are used to compose a new one depicting an event that never happened, as shown in Figure 1(a); and *copy-paste*, which takes place when parts of the image itself are replicated (often with modifications) to hide content in the same image or to increase/decrease the importance of a specific aspect, as shown in Figure 1(b).

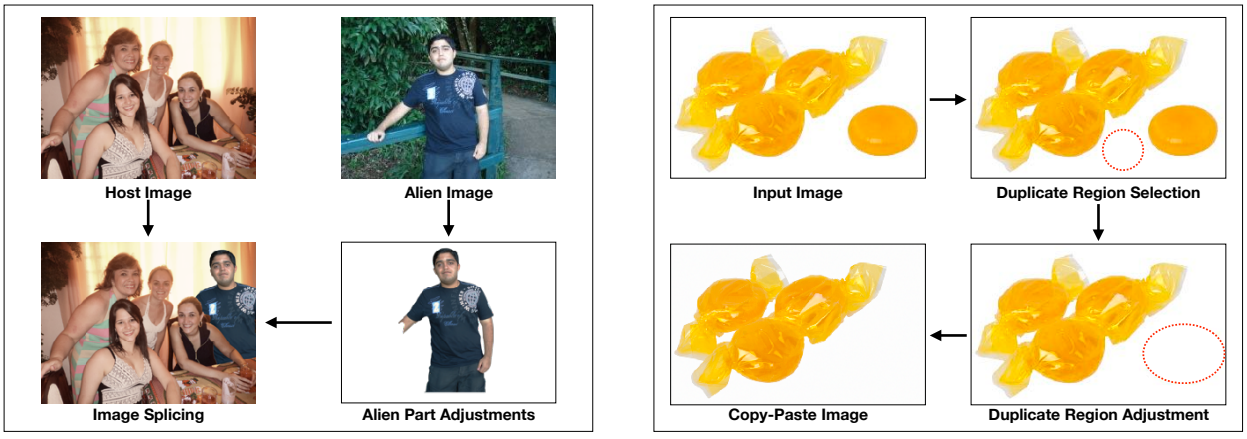
To detect such forgeries, experts look for traces of inconsistencies in different properties, such as illumination, compression, and noise. However, according to Rocha et al. (2011), illumination inconsistencies are potentially effective, mainly when dealing with image splicing: from the viewpoint of a manipulator, proper adjustment of the illumination conditions is hard to achieve when creating a composite image.

Considering techniques that deal with illumination inconsistencies, we can highlight two categories: (1) methods based on the light setting, which aims at finding inconsistencies in the light source position, and (2) methods based on light color, which look for inconsistencies in the color of illuminants of the scene. In a very simplistic form of putting it, the former group analyzes where is the source of illumination of a photograph while the latter investigates how objects are illuminated according to a given light source.

Methods based on the light setting are useful when evaluating the authenticity of outdoor images. One example where such methods could be applied is depicted in Figure 2, in which Marilyn Monroe and Elizabeth Taylor are side-by-side. This fake image, which circulated in social media in 2017, is the result of a composition (splicing) of at least two different photographs.

The method proposed by Carvalho et al. (2015) is appropriate to analyze this kind of image. It is based in the concept of *normal vectors*, often simply named as "normals", which are vectors perpendicular to a surface at a given point. Given some user intervention to mark points in the suspected region, approximately corresponding to normal vectors, the method uses the normal vectors' direction and illumination at the chosen marks to estimate the light source position for the object.

However, even a knowledgeable forensic expert is not 100% precise in the task of setting up normal marks. A proposed workaround is to estimate the light source position several times, always with a random and small distortion in the original normals, to find not a single position, but a region with a certain degree of confidence. In this case, if two or more objects yield inconsistent light source regions (without intersection), as depicted in Figure 2, it is an indicative of an image splicing as it denotes inconsistent light sources.




(a) Image splicing. (b) Copy-paste.

**Figure 1 -** Simplified scheme of the image forgery process: (**a**) image splicing creation, and (**b**) copy-paste image creation.


(a) (b)

**Figure 2 -** (**a**) Example of image splicing, which circulated in social media in 2017. The left figure is the result of an image splicing while the right one is the host image used to construct the forged one. Source: Gizmodo (Novak 2017a). (**b**) Application of the method by Carvalho et al. (2015). Top row: input image, and normal marks for Marilyn Monroe's and Elizabeth Taylor's faces. Bottom row: estimated light source regions. The presence of two regions without intersection is an indicative of image splicing/composition.

In a similar line of research, Seuffert et al. (2018) also explored 3D light in the fake image detection context. The authors presented a new method that evolves from Kee and Farid (2010) model, which proposed that, under the assumption of Lambertian reflectance, the observed intensity can be represented by second order spherical harmonics. Starting from this previous model, the authors proposed a more stable method for real scenario forensics applications. Using an "intensity binning sphere" (IBS), the intensities are binned by their surface normals, avoiding extrapolation over surface normals without observations. The authors also proposed a new error score, instead of a hard threshold, which is learned from training data, i.e., from face images that are acquired under known lighting.

In the second category, techniques that deal with inconsistencies in light color are very useful in more complex setups, often involving not just a single light source. One example where such methods could be applied is depicted in Figure 3, a fake image showing Putin surrounded by other world leaders, which circulated in social media in 2017.

The method by Carvalho et al. (2016) can be properly applied in this case. It explores image transformed spaces to capture artifacts and pinpoint


**Figure 3 -** Example of image splicing, which circulated in social media in 2017. The left figure depicts an image composition while the right one depicts the host image used to construct the fake one. Source: Gizmodo (Novak 2017b).

possible forgeries. Illumination inconsistencies in objects with similar materials (such as human skin) become more pronounced when projecting the fake image onto illuminant maps. An illuminant map is a transformed color space which reproduces the illuminant color (the color of the light that appears to be emitted during the capture) in each region of the image.

There are different color constancy methods able to estimate scene illuminants and this work relied upon two of them: a statistical one (van de Weijer et al. 2007) used to estimate the illuminant from pixels; and a physics-based one (Riess and Angelopoulou 2010), which is a variant of the original inverse-intensity chromaticity space estimation proposed by Tan et al. (2004) to deal with local illuminant estimation.

Illuminant maps can be characterized by different statistical features: texture, color, and shape. Then, for each pair of selected faces in an image, these features are used to train different pattern classifiers, which vote to decide whether or not an image presents traces of illumination inconsistencies. If at least one pair of faces is classified as fake, the method provides a hint that the image may have been manipulated and that the examiner should look for other traces of doctoring. Figure 4 depicts a simplified overview of the method, which classifies an input as being genuine or fake.

The method proposed by Huh et al. (2018) takes advantage of the automatically recorded photo EXIF metadata as supervisory signal for training a model to determine whether an image is, or not, produced by image splicing. The authors apply a Siamese network to measure the consistency $c_{ij}$ of EXIF metadata attributes between two patches $i$ and $j$. The Siamese network uses shared ResNet-50 (He et al. 2016) sub-networks, each one producing feature vectors with 4096 dimensions. Such vectors are concatenated and passed through a four-layer neural network with 4096, 2048, 1024 units, followed by the final output layer. Despite the interesting results, mainly on tampering maps generation, scenarios involving compressed test images have not been tested, which is an inconvenient drawback since this kind of operation tends to degrade accuracy.

The aforementioned methods are effective when dealing with images generated by image composition/splicing. Nevertheless, when the forgery operation involves just parts of the same image, as in copy-move images, there are more appropriate solutions.

Silva et al. (2015) proposed a method tailored for copy-move image forgery detection based on a multi-scale analysis of the input image. The input

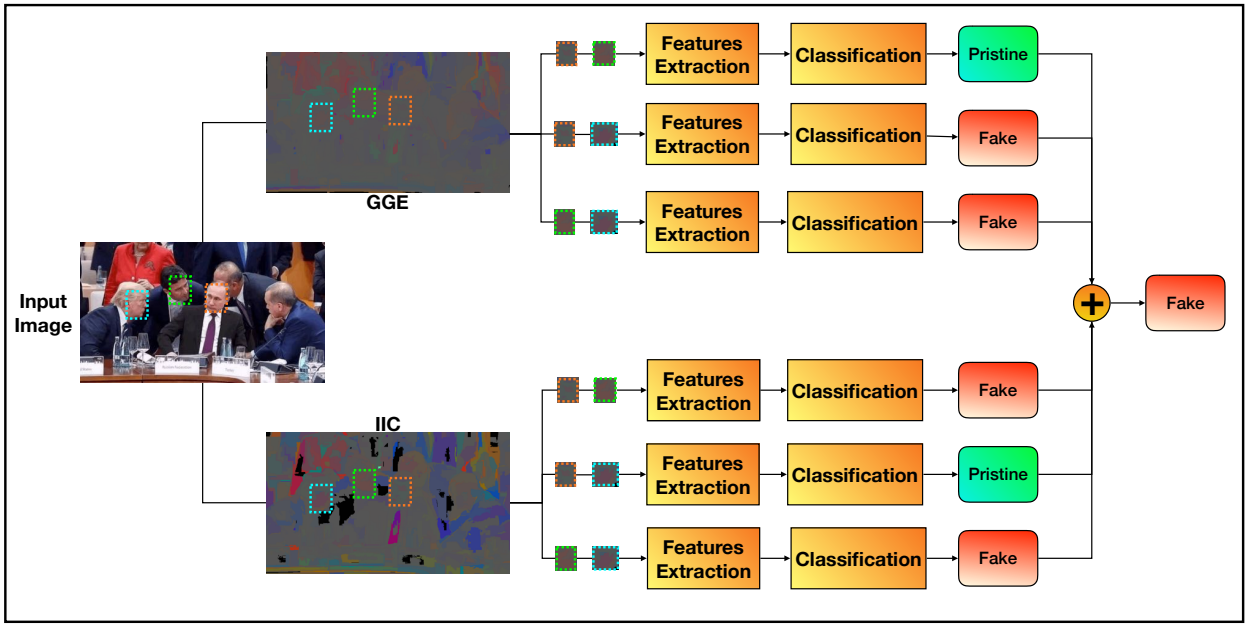



**Figure 4 -** Overview of the method proposed by Carvalho et al. (2016). An image is analyzed by segmenting all suspected faces. Each face is represented in the transformed illuminant space and further characterized through some image description methods such as the ones involving patterns of texture, color, and shape. Then, the descriptors for all pairs of faces in the image are concatenated. Ultimately, a classifier is trained based on the concatenated feature vectors to issue an authenticity decision upon receiving each pair of faces.

image is converted into the HSV color space to decrease possible false positive matches of similar regions. Then, the Speeded-Up Robust Features (SURF) algorithm (Bay et al. 2006) is used to detect a set of keypoints (representative regions of orientation change in the image), which are matched against each other. The method always associates keypoints in pairs by using the Nearest Neighbor Distance Ratio (NNDR) policy (Mikolajczyk and Schmid 2005). The next step consists in clustering keypoints into two groups, based on two specific constraints: (1) spatial proximity between keypoints assigned to the same group; and (2) similarity between the angles of the connection line of keypoints in the same group. The next two steps are, respectively, a Gaussian Pyramidal Decomposition to generate the image's scale-space, and a multi-scale lexicographical analysis, looking for candidate cloned regions in each scale of the pyramid. Finally, the method performs a voting process through the different pyramidal levels to find the final detection map.

An example of the usefulness of such techniques can be found in Figure 5. The image depicts a case which made the news years ago, when Iran was conducting missile tests (Shachtman 2008, Nizza and Lyons 2008). The image is a result of copy-pasting portions of the successful missile launches to the failing ones.

Presenting results comparable to the state-of-the-art methods, but with the drawback of a high complexity implementation framework, Wang et al. (2018) proposed a method based on color invariance model and quaternion polar complex exponential transform (QPCET), for the detection and localization of copy-move forgeries. The proposed method consists of five main steps: (1) extraction of stable color image interest points using a detector composed by SURF (Bay et al. 2008) features and a color invariance model; (2)

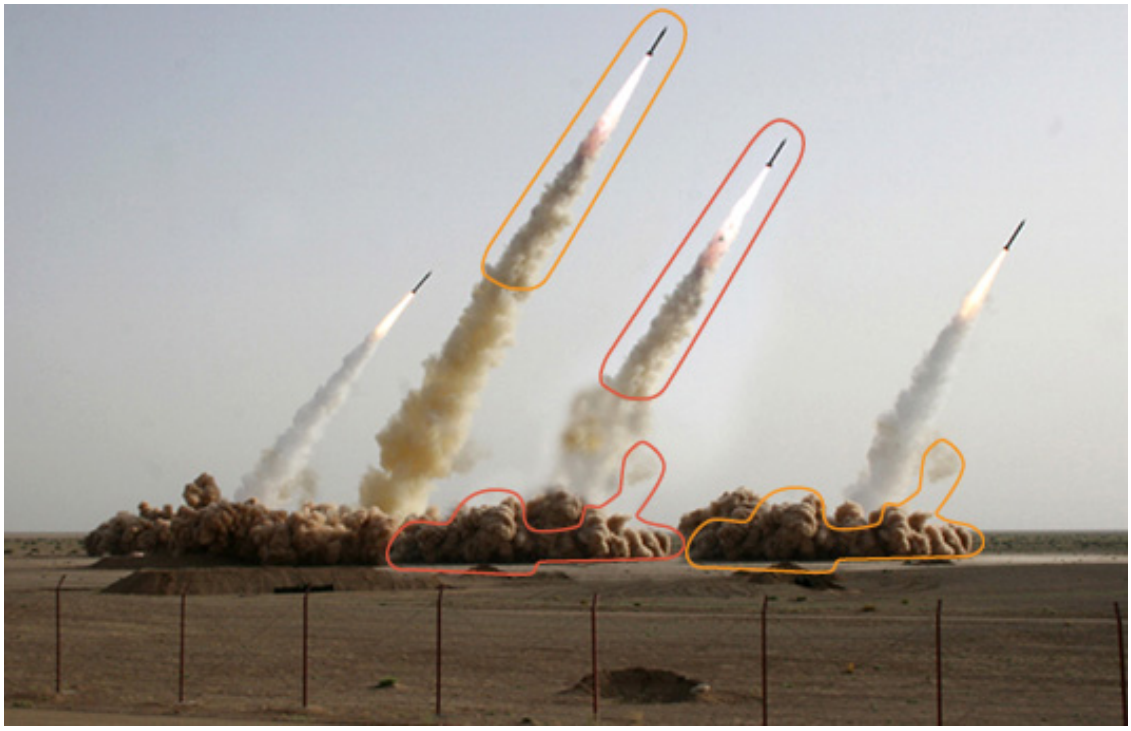


**Figure 5 -** The Iranian Missile Case (Shachtman 2008, Nizza and Lyons 2008). Patterns of a successful missile launch were replicated to the failing one, a clear case of copy-pasting portions of an image so as to change its meaning. Illustration by The New York Times; photo via Agence France-Presse.

generation of connected Delaunay triangles based on the extracted interest points; (3) computation of suitable local visual features of the triangle mesh using the quaternion polar complex exponential transform (QPCET); (4) match of triangular meshes by using local visual features associated with reversed-generalized 2 nearest-neighbor (Rg2NN) and best bin first (BBF); and (5) removal of the falsely matched triangular meshes by random sample consensus.

Also focusing on copy-move detection, Mahmood et al. (2018) proposed a technique based on stationary wavelet and discrete cosine transform. The method first converts the input image into the $YC_bC_r$ color space. For feature extraction, the authors rely on two main steps: (1) using stationary wavelet transform (SWT), the method decomposes the suspicious image into four sub-bands (approximation, horizontal, vertical and diagonal); (2) it divides the approximation sub-band into overlapping blocks, and uses discrete cosine transform (DCT) to reduce them to six dimensions. This combination of SWT and DCT makes the representation of features more diverse and also appears as a better choice for copy-move detection. Using a lexicographical sorting algorithm, features are sorted and the similarity of close blocks are calculated. As a last step, a morphological opening operation with a structural element is applied over resulting maps for eliminating falsely detected areas.

Aiming to improve the detection of forgery localization in fake images, Zhou et al. (2018) proposed a method based on a two-stream Faster R-CNN (Ren et al. 2017) network, which is independent of the forgery process creation (splicing, copy-paste, etc.) The first is an RGB stream that uses a ResNet101 network (He et al.

2016) to learn features from the RGB image input, which are feed into a Region Proposal Network (RPN), in order to find tampering artifacts, such as strong contrast difference and unnatural tampered boundaries. In the second, which is a noise stream, the input RGB image goes through a steganalysis rich model (SRM) filter layer to discover the noise inconsistency between authentic and tampered regions. Features from both streams are fused through bilinear pooling to detect manipulation.

## PRINTED DOCUMENT SOURCE DETECTION

Although we are living in the digital era, in which we are highly connected through many digital devices, printed paperwork is (still) everywhere. Due to the decreasing costs of printer devices (matrix dot, thermal, ink-jet, or laser printers) and the increasing number of digital documents, it is difficult to ensure the authenticity of printed documents against criminal intentions. Identifying the source of a printed document might prove beneficial in investigations involving forged contractual clauses, threatening letters, illegal correspondence, fake currency and documents, among others.

In this vein, it is pivotal to recognize the device signature based on the different characteristics left by its mechanical nuances. Shang et al. (2014) proposed a method to distinguish text documents from laser printer, ink-jet printer, and copier, using features such as noise energy, contour roughness of the character, and average gradient of the character edge region. A SVM classifier is applied for each character and a voting mechanism provides the final result, with a reported 90% accuracy. Similar approaches can be seen in (Joshi and Khanna 2018, Ferreira et al. 2015, Bertrand et al. 2013, Tsai and Liu 2013), where hand-crafted features from characters are extracted and combined for single classification. Although similar, the feature extraction may differ, ensuring some advantages for each technique in different scenarios.

Looking for a more general solution, Ferreira et al. (2017) developed a set of tools to analyze and to recognize document ownership based on clues left behind by printer devices using a data-driven approach. In this approach, several parallel Convolutional Neural Networks (CNNs) extract meaningful discriminative patterns from the analyzed documents. The method is capable of learning distinctive attribution features directly from available training data, a remarkable advance when compared to prior art. By representing these patterns in different ways, it is possible to better identify printing artifacts based on printed characters and, therefore, enhance the document-printer attribution task.

Figure 6 shows the document attribution pipeline where documents are scanned and characters are identified. The approach is based on the analysis of small patches or regions of the analyzed document, represented by text characters. Multiple representations of the same character are used as complementary features, increasing the overall accuracy. These representations are formed by raw data (characters image pixels), media filter residual (subtracting the raw image from the media filtered version provides high frequency imperfections), and average filter residual (subtracting the raw image from the average filtered version isolates border effects). All these three representations are used as input by shallow CNNs, which learn the most relevant discriminant characteristics. The created feature vectors are concatenated and used as input by a set of linear classifiers, called “early fusion”. This step is represented by the middle block in the figure. The classification results at character level are combined by a majority voting mechanism, called “late fusion”, represented by the next block in the figure. The final decision-making process states

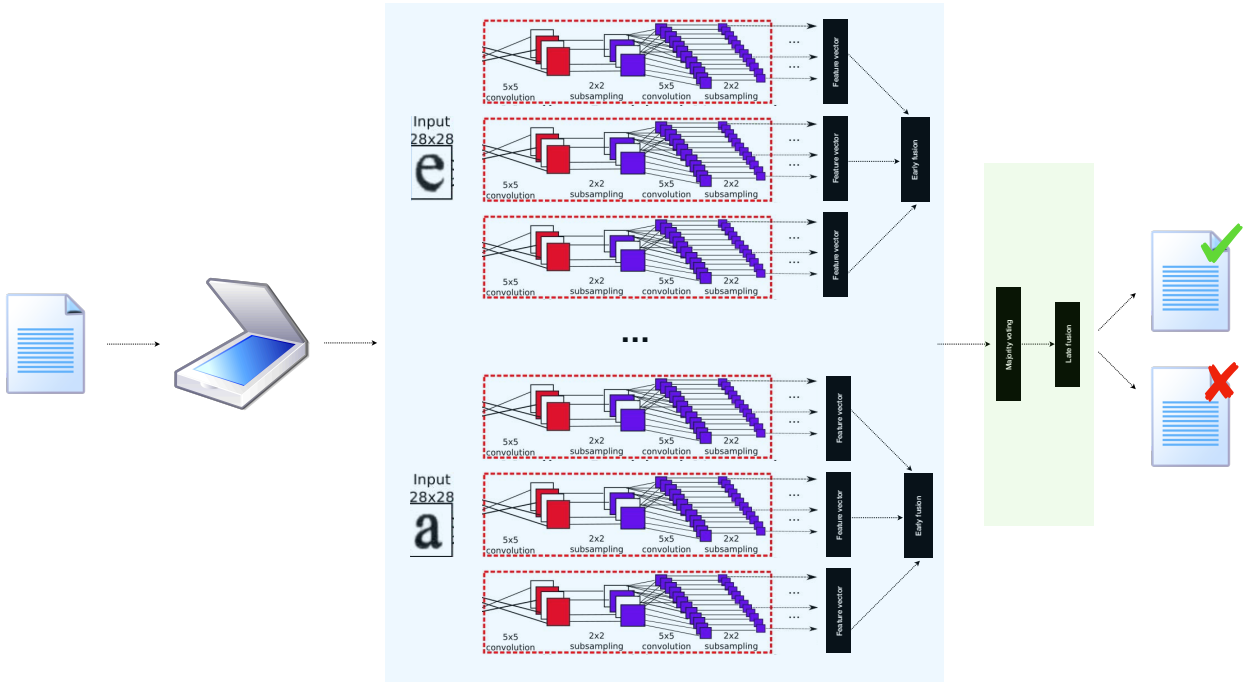



**Figure 6 -** Document attribution pipeline overview. Adapted from (Ferreira et al. 2017). A suspected document is analyzed through a series of non-linear transformations by means of convolutional neural networks in order to extract distinctive attribution patterns. Ultimately, all features are combined for decision-making.

which printer, from a set of suspect printers, was used to print the document under investigation.

## FAKE VIDEO AND SENSITIVE MEDIA DETECTION

With the spreading availability of user-friendly applications to generate fake videos, some of which are used solely to generate pornographic content effortlessly (Farokhmanesh 2018) — the so-called deep fakes — it comes the alarming necessity of methods to reliably detect fake videos. Due to the incredibly realistic generated videos, it can be extremely difficult, for humans and computers, to discern between original and synthetic content, which becomes even more challenging when those videos are shared in low resolution, with various compression artifacts.

Some recent methods in literature have gained momentum due to the their ability to generate compelling manipulated content, by face reenactment in real-time (Thies et al. 2016), by learning lip-sync from audio (Suwajanakorn et al. 2017), or by animating static images (Averbuch-Elor et al. 2017). Being able to accurately detect such content would aid in ceasing the proliferation of fake news by, for instance, blocking or tagging the manipulated images and videos disseminated in social media.

Mainly due to the lack of data (Rössler et al. 2018), which is a requirement for training modern machine learning methods, research in video manipulation detection is rather limited, in opposition to the research scenario discussed for images. The literature focuses on some simple clues, often found in carelessly generated videos, such as insertion and deletion of frames (Gironi et al. 2014, Smith et al. 2017), copy-move manipulations (Bestagini et al. 2013), and green-screen splice (Mullan et al. 2017).

The presented scenario indicates that more research is necessary, considering newly proposed datasets (Rössler et al. 2018) which, for instance, enables the study of video compression on the detection task, a problem often overlooked in literature. One research area that might help in this effort is pornography and violence detection, which are common themes in fake news, specially in videos. Such methods could be used, for instance, to tag content prior to the forgery detection.

Considering pornography and violence detection, there are some works in the literature targeting broader contexts. Moreira et al. (2016) proposed a detection and localization method for general pornography and violence scenes in videos. In a parallel work, Perez et al. (2017) have a similar solution but using deep-learning techniques.

Child pornography is a serious unfolding from general pornography, which only recently has gained proper attention. Automatically distinguishing child pornography from adult pornography and regular everyday images/videos is the main goal of a work conducted by our research group in collaboration with the Brazilian Federal Police and several universities (Vitorino et al. 2018). Based on data-driven strategies, the approach consists of first training CNNs to address different tasks for which there is a massive amount of available training examples, such as general image classification (objects, persons, cars, etc.). Then, through transfer learning techniques, the networks are fine-tuned first to general pornography detection and then further refined for child pornography content detection, outperforming different off-the-shelf solutions.

## TEXTUAL AUTHORSHIP DETECTION – WHO DIT IT?

According to the statistics company Statista,[2] Twitter is currently among the most popular social networks worldwide, with some 330 million active users, who are able to read and post short messages, the so-called *tweets*. In this sea of users and tweets, fans can happily interact with their idols, such as the pop-band Coldplay, or the soccer player Cristiano Ronaldo. However, in the same way images are forged to generate fake news, this technology can also be used for shady purposes.

In 2015, the New York Times documented the case of a Russian media agency that allegedly ran organized disinformation campaigns on social media using pseudonyms and virtual identities (Chen 2015). Ruling an office full of media professionals, the agency achieved success in promoting fake news stories, influencing public opinion on politics. Cases such as this one are examples of how online anonymity can encourage less accountability, being powerful triggers for fake news. Early on in 2018, another full-coverage of fake profiles on social media has broken the news. A Times report delved into the social media's black market in which fake profiles can be bought to boost online popularity (Confessore et al. 2018). Equally alarming are estimates that some 48 million of Twitter's reported active users are automated accounts seeking to simulate real people, according to the article.

The problem of text authorship attribution based on short sequences of text is not new. Sanderson and Guenter (2006) evaluated the usage of word sequence kernels based on Markov chains for words and characters. The considered short text varies from 300 to 5,000 words, which is much more than Twitter's 280 characters limitation (historically 140 characters). Stamatatos (2009) highlighted the difficulties of short text scenarios and its associated challenges. Even considering an accumulative representation, which is considered best when only short text is available, the text length is still a major issue.

[2]https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Focusing on cybercrime identification from short texts shared on Twitter, Layton et al. (2010) adopted the Source-Code Author Profile (SCAP) methodology, introducing new preprocessing methods for text of 140 characters or less. Schwartz et al. (2013) described the concept of a *k-signature* for an author, which is formed by character and word n-grams. They also described a new feature, called flexible patterns, to capture fine-grained nuances in an author's style. Looking to identify an author's style from *tweets*, Bhargava et al. (2013) blended several syntactical, lexical, and tweet specific metrics. These metrics were later evaluated by Overdorf and Greenstadt (2016) in a cross-domain scenario, for which the authors proposed specific feature selection methods.

To aid the fight against this lack of accountability in social networks, Rocha et al. (2017) discussed a general framework, which has the advantage of being scalable to a high number of suspects. It is composed of training and testing stages.

In the training stage, messages associated with suspects' accounts are collected from social media and pre-processed in order to remove sparse features, such as numbers, dates, times, URLs, very short messages with only a few characters, and non-English messages, which enforces the consistency required in the subsequent feature extraction step. All features are then combined into feature sets based on the common bag-of-words models (Salton and McGill 1986). The authors implemented different strategies for this step: character-level n-grams, word-level n-grams, part-of-speech n-grams, and diverse lexical and syntactic statistics as features. This form of characterization captures stylistic features of an author (for instance, a tendency for using capital letters over lowercase ones), patterns of use of Emojis and other social media conventions, as well as vocabulary richness and user-specific grammar constructions. The feature sets are used to train a classifier, such as Power Mean SVM (Wu 2012), W-SVM (Scheirer et al. 2014), Random Forests (Breiman 1996), SCAP (Frantzeskou et al. 2007), and compression-based attribution (Teahan and Harper 2003).

The test stage starts with a message of unknown authorship, which proceeds through the exact same feature extraction process as the training messages. The resulting feature vector is submitted to the pre-trained classifiers, which produces a prediction of its authorship. This result points out the most probable suspect from a set of possible ones. Although it represents an important step towards understanding the difficult problem of authorship attribution for very short messages in social networks, this work also highlights the necessity of developing more informative features capable of capturing stylistic nuances of each person in order to achieve a better classification. Figure 7 depicts an overview of the method.

## MULTIMEDIA PHYLOGENY: UNDERSTADING THE INTERPLAY OF DIGITAL OBJECTS

The efficient techniques presented thus far can be applied to specific forgery cases. Taking a wider perspective, more complicated situations can be easily found, involving several doctored images, in which the original (source) image is replicated and a set of transformations is applied to generate new images. Although all images share common characteristics, they might transmit a completely different message. Considering our highly-connected world through social networks and the universal language of images and videos, visual content can go viral into a worldwide scale very rapidly. Understanding the relationship of digital objects and their interplay is at the core of *Multimedia Phylogeny*, a research area focused on understanding the history and evolution over time of digital objects as a group rather than whether or not isolated objects are authentic.

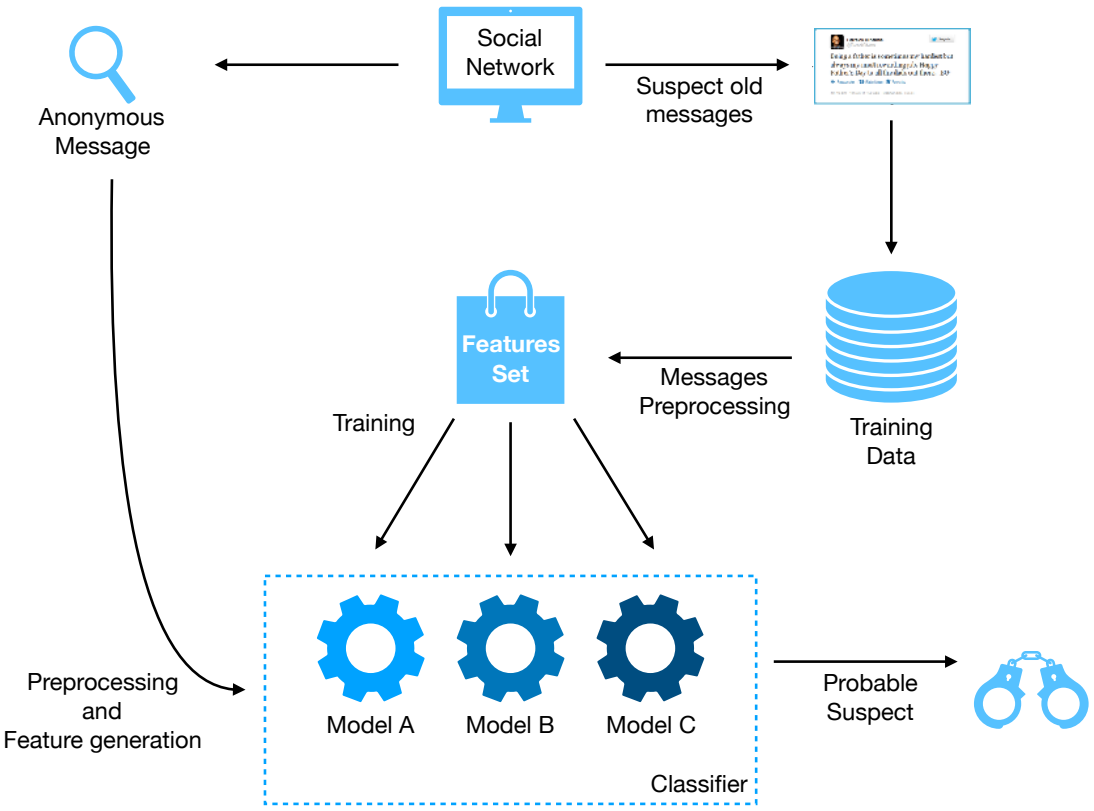



**Figure 7 -** Overview of the method proposed by Rocha et al. (2017) for authorship attribution of very short messages online. Adapted from (Rocha et al. 2017).

A well-known example of such a problem is "The Situation Room" photo, taken by the White House photographer Pete Souza, on May 1st 2011. It depicts former US President Barack Obama and some officers receiving updates from the operation that aimed at capturing Al Qaeda's terrorist Osama bin Laden. The image went viral on the Internet with several transformations, such as text overlay, face swap, insertion of new elements, among others (Figure 8).

There are countless examples of fake news stories spread either through image modification or simple composition. Many of them have political guise, but they can assume any other vantage point. Some other famous examples include: the Iranian missiles case (Figure 5); Brazilian former president Dilma Rousseff's criminal records (Folha de S.Paulo 2009); Sarah Palin holding a rifle (Jackson 2008) right after being nominated a Vice-President candidate for the Republican Party in 2008; President George W. Bush holding a children's book upside down (Jaffe 2002); the Pope endorsing Donald Trump for President (Christensen 2016); and the 2011 Benetton's online Unhate ad campaign (Pownall 2015). Still, the list goes on and on and this is just the tip of the iceberg.

Notwithstanding the fact that the recognition of exactly duplicated images is straightforward, the identification of semantically-similar images (when transformations are applied) and their compositions is a challenging task. Moreover, identifying the source image and the relationship among all images under investigation are paramount to support digital forensic analysis. To aid in this complex analysis, research groups have been developing several methods (Dias et al. 2013b, Melloni et al.


(a) (b)

(c)

**Figure 8 -** "The Situation Room" photo. (**a**) Source image taken by the White House photographer Pete Souza; (**b**) A version produced by the Brooklyn-based Hasidic newspaper removing Secretary of State Hillary Clinton and another woman from the photo. Source: (HUFFPOST 2011); and (**c**) A *meme* depicting the politicians as super heroes. Source: (ENews 2012).

2014, Oikawa et al. 2016, Costa et al. 2016) for multimedia phylogeny and provenance integrity over the years.

Inspired by the biological process of characteristics inheritance, *multimedia phylogeny* aims at identifying the relationship between a set of near-duplicate or semantically-similar images. Relationships are mapped as an image phylogeny tree (IPT) or forest (IPF), enabling the identification of the temporal sequence of modifications based on an image ancestral lineage and descendants. The IPTs/IPFs are presented by directed acyclic graphs, where arcs are created and weighted using a dissimilarity function.

In the method proposed by Dias et al. (2013a), the first step is to identify the images relationship pixel-by-pixel. Relevant points are matched between two images (ancestral/descendant) using SURF (Bay et al. 2008) and RANSAC (Fischler and Bolles 1987) methods. Normalization and compressing techniques are applied, and a pixel-wise comparison is carried out based on an homography matrix and used as a minimum squared error metric. Figure 9 shows an example of this process, which results into the dissimilarity matrix of $n$ near-duplicate images. Based on this matrix, a modified minimum spanning tree algorithm (Dias et al. 2012) constructs the IPT as a second step. The process starts with a $n$-root forest and sorts

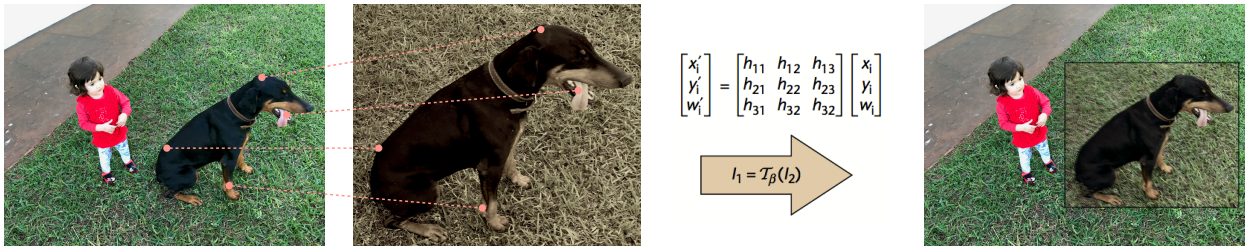



**Figure 9 -** Image mapping process: to calculate the dissimilarity between images $I_1$ and $I_2$, robust points of interest are computed. Then, an homography matrix is calculated to enable pixel-by-pixel comparison and, therefore, the dissimilarity matrix. Adapted from (Dias et al. 2013a).

the dissimilarity matrix elements based on their computed dissimilarity values. Different trees are joined according to their sorted order.

In real-world setups, the complete set of near-duplicates is often not available, forcing the technique to deal with missing nodes. In addition, multiple source images can exist, particularly in splicing/composition cases. In this scenario, a forest of IPTs needs to be constructed.

The multiple trees approach is also applicable for analyzing evidence involving image montages, blending, or a combination of different camera viewpoints. Such extensions are called *multiple parenting phylogeny* (Oikawa et al. 2016). Going back to "The Situation Room" example, Figure 10 shows how the proposed IPT helps in understanding the relationship between the related images collected from the Internet and their process of evolution from the very original photograph taken by White House photographer Pete Souza.

## WHAT COMES NEXT IN TERMS OF UNDERSTADING REAL-WORLD EVENTS

Criminal activities evolve and adapt quickly, being fake news particularly on the spotlight, evincing the necessity of effective tools to help us answer the four most important aspects of an event: "who", "in what circumstances", "why", and "how". The covered techniques in this article are pinnacle examples of research aiming at answering one or more of these questions. However, a much richer "bird's eye view" of an event is pivotal to fully understand the nuances and details of an event.

There are a few researches currently focusing on representing and understanding real-world events as a whole, from media content in which is immersed a sea of fake news. Two projects — DéjàVu (Rocha 2017) and Forensic Architecture (Weizman et al. 2014) — can be highlighted, due to their heterogeneity regarding the representation and tools considered to tackle the problem.

The recently launched DéjàVu project (Rocha 2017) focuses on synchronizing, in space and time, all multimedia information collected from a target event, enabling fact-checking and mining persons, objects, and contents of interest. This process of synchronization is referred to as *X-coherence*. Such multimedia information may come from varying heterogeneous sources, such as social media, the Internet, and surveillance cameras. The synchronization allows us to better understand an event by virtually reconstructing it — the before, during, and aftermath. Once we can move through the reconstructed event, we have a higher chance of answering the important forensic questions mentioned before, likely providing irrefutable evidence to what really happened.

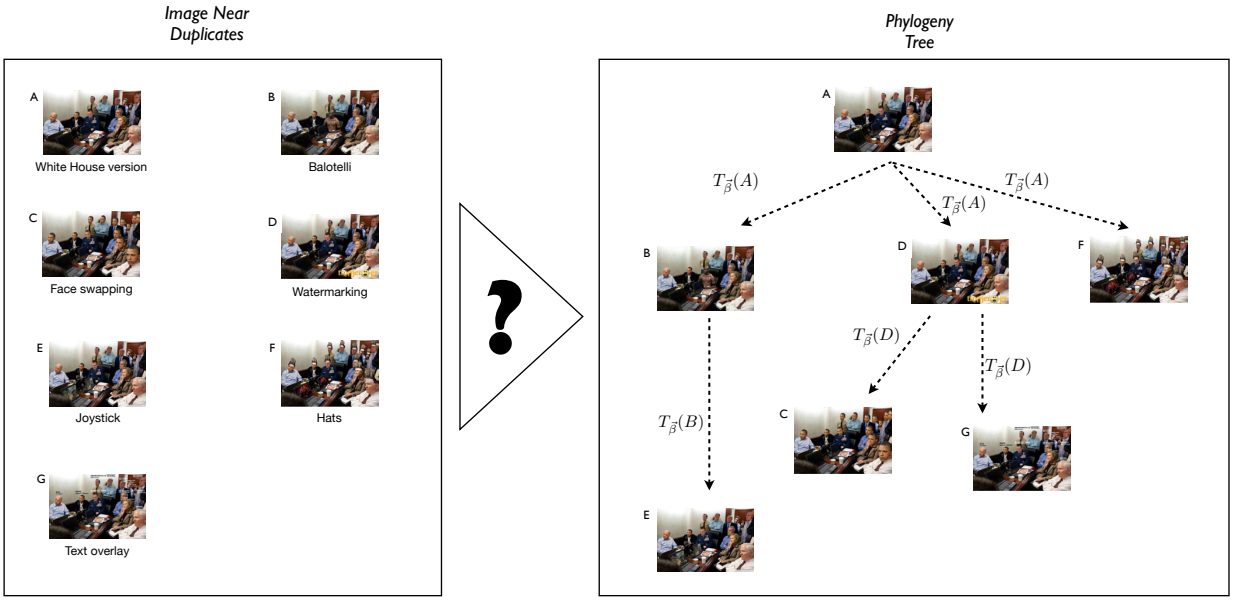



**Figure 10 -** IPT construction example using images collected from the Internet about "The Situation Room" episode. Adapted from (Oikawa et al. 2016).

With the *X-coherence*, which refers to feature-space-time coherence, it is possible to find physical (e.g., where something happened), temporal (e.g., when it took place), and feature (e.g., creating a transformed and unified feature space so as to allow content discovery and pattern understanding) relations. It can be seen as a natural evolution of the traditional multimedia phylogeny solutions, in particular of the multiple parenting phylogeny and the ultimate integration of all forensic analysis pieces.

In order to hint at the *X-coherence* strength, Lameri et al. (2014) analyzed a pool of videos related to specific events, first focusing on the reconstruction of longer parent sequences describing the event itself. By mainly focusing on the Boston Marathon Bombing event (NEW YORK TIMES 2013), where two bombs went off near the marathon finish line, the proposed technique was capable of reconstructing longer sequences of videos (at times complementing the smaller ones) associated with the event. Considering the social impact of this event and, therefore, the flood of data produced by social and mainstream media, it was possible to provide the right chronological sequence, joining assorted multimedia materials in order to support, among other aspects, suspect identification and event understanding, which is one of the goals of the *X-coherence* synchronization.[3]

Forensic Architecture (Weizman et al. 2014), on the other hand, is a project and multidisciplinary research group which considers architectural techniques to investigate cases of human rights violation around the world. It aims at producing and presenting architectural evidence in contemporary conflicts. By analyzing shared media, they are able to model dynamic events as they unfold in space and time, by creating navigable 3D models and interactive cartographies of sites of conflicts. These techniques allows the presentation of the events

[3] A video demonstrating the technique is available at http://tinyurl.com/q6fslbj and more information about the project can be found at http://dejavu.ic.unicamp.br/.

in an accountable manner, also generating new insights.

An example of their work is the Grenfell Tower fire examination,[4] which aims at facilitating the investigation of an unprecedented fire that destroyed the Grenfell Tower in London. The event was captured live by thousands of cameras and smartphones, which together can provide evidence and unique information about the event. The final goal is to construct a continuous "3D video" of the fire, mapped onto an architectural model of the Grenfell Tower.

Another remarkable example of their work on X-coherence is now known as the Black Friday Reconstruction, which is a collaboration between Forensic Architecture and Amnesty International aiming to provide a detailed reconstruction of Israel bombing and attacks in Rafah, Gaza, from 1 through 4 August 2014, based primarily on material found on social media. As the investigation team did not have ground access to Gaza, they have developed a number of techniques aimed to reconstruct the events from hundreds of images and videos recorded by professional and citizen journalists. The images were thereafter located in a 3D model of Rafah. This resulted in the *Image Complex*, a solution capable of allowing the exploration of spatial and temporal connections of different sources and reconstruct events as they happened.[5]

People all over the world use their mobile devices to capture and share all sorts of events they are witnessing, and, at the same time, a profusion of manipulated versions of this data are propagated through the same channels. Some research projects, such as DéjàVu and Forensic Architecture, are taking advantage of this scenario in order to facilitate the solution of real and virtual crimes, in the pursuit of accountability. It enables the aggregation of significant forensic solutions and the design and development of novel methods to analyze interactions between heterogeneous sources, targeting the prevention and investigation of crimes, while also fighting back fake news proliferation. This is certainly a significant step forward in the process of understanding the world around us, taking full advantage of a myriad of sources registering what is happening around the world.

[4]http://www.forensic-architecture.org/case/grenfell-tower-fire/

[5]https://www.forensic-architecture.org/case/rafah-black-friday/

**ACKNOWLEDGMENTS**

The research for this paper was financially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), DéjàVu grant #2017/12646-3 and grant #2017/12631-6; by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), DeepEyes grant; and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grants #304472/2015-8 and #423797/2016-6.

**REFERENCES**

ALLCOTT H AND GENTZKOW M. 2017. Social Media and Fake News in the 2016 Election. J Econ Perspect 31(2): 211-236.

AVERBUCH-ELOR H, COHEN-OR D, KOPF J AND COHEN MF. 2017. Bringing portraits to life. ACM T Graph 36(6): 196.

BAY H, ESS A, TUYTELAARS T AND VAN GOOL L. 2008. Speeded-up robust features (SURF). Comput Vis Image Underst 110(3): 346-359.

BAY H, TUYTELAARS T AND VAN GOOL L. 2006. SURF: Speeded Up Robust Features. In: Leonardis A, Bischof H and Pinz A (Eds), Computer Vision – ECCV 2006, p. 404-417. Berlin, Heidelberg: Springer Berlin Heidelberg.

BERTRAND R, GOMEZ-KRÄMER P, TERRADES OR, FRANCO P AND OGIER JM. 2013. A system based on intrinsic features for fraudulent document detection. In: International Conference on Document Analysis and Recognition (ICDAR), p. 106-110.

BESTAGINI P, MILANI S, TAGLIASACCHI M AND TUBARO S. 2013. Local tampering detection in video sequences. In: International Workshop on Multimedia Signal Processing (MMSP), p. 488-493.

BHARGAVA M, MEHNDIRATTA P AND ASAWA K. 2013. Stylometric analysis for authorship attribution on Twitter. In: International Conference on Big Data Analytics, p. 37-47.

BILLER D. 2018. Fake News Risks Plaguing Brazil Elections, Top Fact-Checkers Say. https://www.bloomberg.com/news/articles/2018-01-09/fake-news-risks-plaguing-brazil-elections-top-fact-checkers-say. Accessed: 2018-01-29.

BREIMAN L. 1996. Bagging Predictors. Machine Learning 24(2): 123-140.

CARVALHO T, FARIA FA, PEDRINI H, TORRES RDS AND ROCHA A. 2016. Illuminant-Based Transformed Spaces for Image Forensics. IEEE Trans Inf Forensic Secur 11(4): 720-733.

CARVALHO T, FARID H AND KEE E. 2015. Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis, In SPIE Symposium on Electronic Imaging.

CHEN A. 2015. The Agency. The New York Times Magazine. URL https://www.nytimes.com/2015/06/07/magazine/the-agency.html. Accessed: 2018-02-05.

CHRISTENSEN BM. 2016. No, the Pope Has NOT Endorsed Donald Trump For President. Hoax-Slayer, http://www.hoax-slayer.net/no-the-pope-has-not-endorsed-donald-trump-for-president/. Accessed: 2018-07-11.

CONFESSORE N, DANCE GJ, HARRIS R AND HANSEN M. 2018. The Follower Factory. The New York Times, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html. Accessed: 2018-07-11.

COSTA FDO, LAMERI S, BESTAGINI P, DIAS Z, TUBARO S AND ROCHA A. 2016. Hash-based frame selection for video phylogeny. In: IEEE International Workshop on Information Forensics and Security (WIFS), p. 1-6.

DIAS Z, GOLDENSTEIN S AND ROCHA A. 2013a. Large-scale image phylogeny: Tracing image ancestral relationships. IEEE Multimedia 20(3): 58-70.

DIAS Z, GOLDENSTEIN S AND ROCHA A. 2013b. Toward image phylogeny forests: Automatically recovering semantically similar image relationships. Forensic Sci Int 231(1-3): 178-189.

DIAS Z, ROCHA A AND GOLDENSTEIN S. 2012. Image phylogeny by minimal spanning trees. IEEE Transactions on Information Forensics and Security 7(2): 774-788.

ENEWS. 2012. 2012 Election: Best Political Memes – The situation room. E News, http://www.eonline.com/photos/6210/2012-election-best-political-memes/218210. Accessed: 2018-02-12.

FAROKHMANESH M. 2018. Deepfakes are disappearing from parts of the web, but they're not going away. The Verge, https://www.theverge.com/2018/2/9/16986602/deepfakes-banned-reddit-ai-faceswap-porn. Accessed: 2018-02-11.

FERREIRA A, BONDI L, BAROFFIO L, BESTAGINI P, HUANG J, SANTOS JAD, TUBARO S AND ROCHA A. 2017. Data-Driven Feature Characterization Techniques for Laser Printer Attribution. IEEE Trans Inf Forensic Secur 12(8): 1860-1873.

FERREIRA A, NAVARRO LC, PINHEIRO G, SANTOS JAD AND ROCHA A. 2015. Laser printer attribution: Exploring new features and beyond. Forensic Sci Int 247: 105-125.

FISCHLER MA AND BOLLES RC. 1987. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: Readings in Computer Vision, p. 726-740.

Folha de SPaulo. 2009. Autenticidade de ficha de Dilma não é provada. Folha de S.Paulo, http://www1.folha.uol.com.br/folha/brasil/ult96u556855.shtml. Accessed: 2018-07-11.

FRANTZESKOU G, STAMATATOS E, GRITZALIS S, CHASKI C AND STEPHEN HOWALD B. 2007. Identifying Authorship by Byte-Level N-Grams: The Source Code Author Profile (SCAP) Method. IJDE 6(1).

GIRONI A, FONTANI M, BIANCHI T, PIVA A AND BARNI M. 2014. A video forensic technique for detecting frame deletion and insertion. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p. 6226-6230.

HE K, ZHANG X, REN S AND SUN J. 2016. Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 770-778.

HUFFPOST. 2011. Hillary Clinton Removed From Situation Room Photo By Der Tzitung, Hasidic Newspaper (PHOTOS). HuffPost, https://www.huffingtonpost.com/2011/05/09/hillary-clinton-der-tzitung-removed-situation-room_n_859254.html. Accessed: 2018-02-12.

HUH M, LIU A, OWENS A AND EFROS AA. 2018. Fighting Fake News: Image Splice Detection via Learned Self-Consistency. arXiv:1805.04096v3 [cs.CV].

JACKSON B. 2008. Picture of Palin Is a Fake. FactCheck, http://www.factcheck.org/2008/09/picture-of-palin-is-a-fake/. Accessed: 2018-07-11.

JAFFE J. 2002. Dubya, willya turn the book over? Wired News, https://www.wired.com/2002/11/dubya-willya-turn-the-book-over/. Accessed: 2018-07-11.

JOSHI S AND KHANNA N. 2018. Single classifier-based passive system for source printer classification using local texture features. IEEE Transactions on Information Forensics and Security 13(7): 1603-1614.

KEE E AND FARID H. 2010. Exposing Digital Forgeries from 3-D Lighting Environments, In IEEE International Workshop on Information Forensics and Security.

LAMERI S, BESTAGINI P, MELLON A, MILANI S, ROCHA A, TAGLIASACCHI M AND TUBARO S. 2014. Who is my parent? Reconstructing video sequences from partially matching shots. In: IEEE International Conference on

Image Processing (ICIP), p. 5342-5346.

LAYTON R, WATTERS P AND DAZELEY R. 2010. Authorship attribution for Twitter in 140 characters or less. In: Cybercrime and Trustworthy Computing Workshop (CTC), p. 1-8.

MAHMOOD T, MEHMOOD Z, SHAH M AND SABA T. 2018. A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform. J Vis Commun Image 53: 202-214.

MELLONI A, BESTAGINI P, MILANI S, TAGLIASACCHI M, ROCHA A AND TUBARO S. 2014. Image phylogeny through dissimilarity metrics fusion. In: European Workshop on Visual Information Processing (EUVIP), p. 1-6.

MIKOLAJCZYK K AND SCHMID C. 2005. A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 27(10): 1615-1630.

MOREIRA D, AVILA S, PEREZ M, MORAES D, TESTONI V, VALLE E, GOLDENSTEIN S AND ROCHA A. 2016. Pornography classification: The hidden clues in video space–time. Forensic Sci Int 268: 46-61.

MORRIS I. 2018. Revenge 'Porn' Gets Even More Horrifying With Deepfakes. Forbes, https://www.forbes.com/sites/ianmorris/2018/02/05/fakeapp-allows-anyone-to-make-deepfake-porn-of-anyone/. Accessed: 2018-02-07.

MULLAN P, COZZOLINO D, VERDOLIVA L AND RIESS C. 2017. Residual-based forensic comparison of video sequences. In: IEEE International Conference on Image Processing (ICIP), p. 1507-1511.

NEW YORK TIMES. 2013. Boston Marathon Bombings. The New York Times, https://www.nytimes.com/topic/subject/boston-marathon-bombings.

NIGHTINGALE SJ, WADE KA AND WATSON DG. 2017. Can people identify original and manipulated photos of real-world scenes? Cognitive Research: Principles and Implications 2(30): 1-21.

NIZZA M AND LYONS PJ. 2008. In an Iranian image, a missile too many. The Lede, The New York Times News, https://thelede.blogs.nytimes.com/2008/07/10/in-an-iranian-image-a-missile-too-many/. Accessed: 2018-07-11.

NOVAK M. 2017a. 29 Viral Photos and GIFs From 2017 That Were Totally Fake. Gizmodo, https://gizmodo.com/29-viral-photos-and-gifs-from-2017-that-were-totally-fa-1821440079. Accessed: 2018-02-02.

NOVAK M. 2017b. That Viral Photo of Putin and Trump is Totally Fake. Gizmodo, https://gizmodo.com/that-viral-photo-of-putin-is-totally-fake-1796767457. Accessed: 2018-02-03.

OBAR JA AND WILDMAN S. 2015. Social media definition and the governance challenge: An introduction to the special issue. Telecomm Policy 39(9): 745-750.

OIKAWA MA, DIAS Z, DE REZENDE ROCHA A AND GOLDENSTEIN S. 2016. Manifold learning and spectral clustering for image phylogeny forests. IEEE Trans Inf Forensic Secur 11(1): 5-18.

OVERDORF R AND GREENSTADT R. 2016. Blogs, Twitter feeds, and Reddit comments: cross-domain authorship attribution. Proceedings on Privacy Enhancing Technologies 2016(3): 155-171.

PEREZ M, AVILA S, MOREIRA D, MORAES D, TESTONI V, VALLE E, GOLDENSTEIN S AND ROCHA A. 2017. Video pornography detection through deep learning techniques and motion information. Neurocomputing 230: 279-293.

POWNALL C. 2015. The Backfiring Campaign. In: Managing Online Reputation, p. 132-140.

REN S, HE K, GIRSHICK R AND SUN J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 39(6): 1137-1149.

RIESS C AND ANGELOPOULOU E. 2010. Scene Illumination as an Indicator of Image Manipulation. In: Information Hiding Workshop. Volume 6387, p. 66-80.

ROCHA A. 2017. DéjàVu: Feature-Space-Time Coherence from Heterogeneous Data for Media Integrity Analytics and Interpretation of Events. http://dejavu.ic.unicamp.br. Accessed: 2018-07-11.

ROCHA A, SCHEIRER W, BOULT T AND GOLDENSTEIN S. 2011. Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics. ACM Comput Surv 43(4): 26:1-26:42.

ROCHA A, SCHEIRER WJ, FORSTALL CW, CAVALCANTE T, THEOPHILO A, SHEN B, CARVALHO ARB AND STAMATATOS E. 2017. Authorship Attribution for Social Media Forensics. IEEE Trans Inf Forensic Secur 12(1): 5-33.

RÖSSLER A, COZZOLINO D, VERDOLIVA L, RIESS C, THIES J AND NIESSNER M. 2018. FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces. arXiv preprint arXiv:1803.09179.

SALTON G AND MCGILL MJ. 1986. Introduction to Modern Information Retrieval. New York, NY, USA: McGraw-Hill, Inc.

SANDERSON C AND GUENTER S. 2006. Short text authorship attribution via sequence kernels, Markov chains and author unmasking: An investigation. In: Conference on Empirical Methods in Natural Language Processing, p. 482-491.

SCHEIRER WJ, JAIN LP AND BOULT TE. 2014. Probability Models for Open Set Recognition. IEEE Trans Pattern Anal Mach Intell 36(11): 2317-2324.

SCHETINGER V, OLIVEIRA MM, SILVA RD AND CARVALHO TJ. 2017. Humans are easily fooled by digital images. Computers & Graphics 68: 142-151.

SCHWARTZ R, TSUR O, RAPPOPORT A AND KOPPEL M. 2013. Authorship attribution of micro-messages. In: Conference on Empirical Methods in Natural Language Processing, p. 1880-1891.

SEUFFERT J, STAMMINGER M AND RIESS C. 2018. Towards Forensic Exploitation of 3-D Lighting Environments in Practice. In: Sicherheit 2018, p. 159-169.

SHACHTMAN N. 2008. Iran Missile Photo Faked. Wired. URL https://www.wired.com/2008/07/iran-missile-ph/. Accessed: 2018-07-11.

SHANG S, MEMON N AND KONG X. 2014. Detecting documents forged by printing and copying. EURASIP J Adv Sig Pr 2014(1): 140.

SILVA E, CARVALHO T, FERREIRA A AND ROCHA A. 2015. Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes. J Vis Commun Image 29: 16-32.

SMITH E, BASHARAT A, ANTHONY HOOGS C AND OTHERS. 2017. A C3D-based Convolutional Neural Network for Frame Dropping Detection in a Single Video Shot. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, p. 86-94.

STAMATATOS E. 2009. A survey of modern authorship attribution methods. J Am Soc Inf Sci Technol 60(3): 538-556.

SUWAJANAKORN S, SEITZ SM AND KEMELMACHER-SHLIZERMAN I. 2017. Synthesizing Obama: learning lip sync from audio. ACM Trans Graph 36(4): 95.

TAN R, NISHINO K AND IKEUCHI K. 2004. Color Constancy Through Inverse-Intensity Chromaticity Space. J Opt Soc Am A 21: 321-334.

TEAHAN WJ AND HARPER DJ. 2003. Using Compression-Based Language Models for Text Categorization. p. 141-165.

THIES J, ZOLLHOFER M, STAMMINGER M, THEOBALT C AND NIESSNER M. 2016. Face2Face: Real-time face capture and reenactment of RGB videos. In: IEEE Conference on Computer Vision and Pattern Recognition, p. 2387-2395.

TSAI MJ AND LIU J. 2013. Digital forensics for printed source identification. In: IEEE International Symposium on Circuits and Systems (ISCAS), p. 2347-2350.

WEIJER JVD, GEVERS T AND GIJSENIJ A. 2007. Edge-Based Color Constancy. IEEE Trans Image Process 16(9): 2207-2214.

VITORINO P, AVILA S, PEREZ M AND ROCHA A. 2018. Leveraging deep neural networks to fight child pornography in the age of social media. J Vis Commun Image 50: 303-313.

WANG XY, JIAO LX, WANG XB, YANG HY AND NIU PP. 2018. A new keypoint-based copy-move forgery detection for color image. Appl Intell 48(10): 3630-3652.

WEIZMAN E AND OTHERS. 2014. Forensis: The architecture of public truth. Sternberg.

WU J. 2012. Power mean SVM for large scale visual classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 2344-2351.

ZHOU P, HAN X, MORARIU VI AND DAVIS LS. 2018. Learning Rich Features for Image Manipulation Detection, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).