The Workshops of the Tenth International AAAI Conference on Web and Social Media
Social Media in the Newroom: Technical Report WS-16-19

# Web and Social Media Image Forensics for News Professionals

**Markos Zampoglou, Symeon Papadopoulos, Yiannis Kompatsiaris**
Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

**Ruben Bouwmeester, Jochen Spangenberg**
Deutsche Welle, Bonn/Berlin, Germany

## Abstract

User-generated content – commonly referred to as "eyewitness media" – has become an essential component in journalism and news reporting. Increasingly more news providers, such as news agencies, broadcasters and Web-only players have set up teams of dedicated investigators or are in the process of training parts of their journalistic workforce to gather and evaluate material from social networks and the Web. If verified, such content can be invaluable in delivering a news story. However, while source checking and verification is as old as journalism itself, the verification of digital material is a relatively young field, with protocols and assisting tools still being developed. In this work, we present our efforts towards a Web-based image verification platform. The platform, currently in its alpha stage, features image tampering detection using a number of state-of-the-art algorithms and image metadata visualization. We discuss the current strengths and limitations of the platform and the implemented state-of-the-art with respect to the specific requirements of the task, resulting from its Web-based nature and its intended use by news investigators with limited expertise in the domain of image forensics.

## Introduction

One of the core duties of journalists is checking and assessing the credibility and accuracy of information before it is spread or used for situation assessment. This is nothing new. What is new, however, is the availability and accessibility of large quantities of digital material residing on social networks, which also brings the need to filter and verify the accuracy of information shared online. Consequently, journalistic reporting has changed to some extent, requiring new skills as well as new tools.

Reporting about or covering events as they unfold is no longer in the hands of a select few (e.g. professional journalists working for established players such as news agencies, equipped with professional reporting equipment). Instead, all that is required to report about events – be it a demonstration, a war/conflict, a natural disaster and the like – is a digital device to capture information, an Internet connection, and upload facilities to a platform such as Facebook, Instagram, or Twitter. All this is combined in a smartphone – the reporting device of the 21st century. Having a smartphone at hand, news nowadays gets "reported" as it happens. Hardly an event goes by without somebody reporting[1] about it in one way or another. This is what the term "eyewitness media" entails.

However, not all content that is posted to social networks is what it pretends to be. There are numerous reasons why people post and share inaccurate information. These could include a) personal motives (the famous "five minutes of fame"; "just for the fun of it", to get attention, vanity, etc.), or b) intentions to influence opinion or bring across particular viewpoints (e.g. PR, spin, marketing, propaganda, etc.).

As it is so easy to spread information these days, while, on the other hand, an increasing number of journalists and media outlets rely on eyewitness media or user-generated content for their reporting, a new skill and requirement is being added to the job profile of professional journalism: the skill to verify (or debunk) social media content. This is important as, without using eyewitness media and knowing how to appropriately deal with it, news outlets:

- would often miss out on the opportunity of "being first", or breaking a news story;
- would not be able to include valuable information and imagery in many cases;
- would have difficulty involving numerous (directly involved) sources;
- would have fewer opportunities to interact with contributors or sources;
- would miss out on improving the user experience.

This said, the verification of images is of paramount importance, especially in the news business, as news is a rather visual business. There is truth in the saying that an image says more than a thousand words: One such notable case is the image of the US Airways plane after its emergency landing on the Hudson river in New York in 2009, taken by ferry commuter passenger Janis Krums[2].

In turn, manipulated images frequently make the rounds – sometimes even headline news – on numerous occasions



[1] Reporting, in this context, refers to an event being covered by a non-professional journalist.

[2] http://twitpic.com/135xa

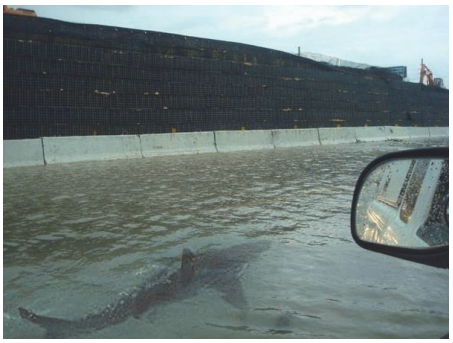

Figure 1: Forged photograph of a shark supposedly swimming in the flooded streets of New Jersey.

before they are detected and debunked. Examples exist in abundance, such as sharks apparently swimming through the flooded streets of New Jersey after Hurricane Sandy (Figure 1), or when the Daily Mail published a photo showing apparent female suicide bomber Hasna Aitboulahcen, killed during a police raid in Paris after the November 2015 attacks. In reality, the image was of a Moroccan woman called Nabila who had nothing to do with the events[3].

For journalists and media organizations, it is of utmost importance to get things right in their reporting, and not fall for deliberately misleading or accidentally inaccurate images. The reason is simple: some of the core assets of (serious) media organizations are at stake: these are trust, reputation and credibility, as well as brand maintenance and protection. So, while eyewitness media and user-generated content can greatly add to news reporting, it is important that verification becomes an integral part of the information assessment and distribution process.

At present, a variety of existing services such as TinEye[4] or Google reverse Image Search[5] can aid in the process of social media content verification. However, there is still a lack of tools that are specifically made for journalism. To this end, we propose a Web-based image verification service with the goal of bringing the power of image forensics to the hands of news professionals. The proposed service is designed following a number of requirements coming directly from the operational setting and practices of journalists, and integrates a number of state-of-the-art image forensics algorithms that are leveraged through an intuitive and highly visual user interface.

## Related work

### Journalistic requirements

From a journalistic perspective, there are several requirements when it comes to tools that could aid the process of image verification. An ideal image verification tool should:

- be easy to use; require little expertise and training;
- be easy to integrate in established verification workflows;
- allow for visually analyzing details, e.g. by allowing high-quality zoom-in;
- provide quick and accurate results, namely:
  - an exhaustive list of metadata, also showing what metadata were not available, along with an easy-to-digest summary of the most important metadata;
  - a set of discovered manipulations (shows what has been manipulated and how);
  - clear indications of whether images have not been tampered or manipulated (providing supportive evidence);
  - clear indications of whether an image pretends to be something else (by also analyzing the context, e.g. an unmanipulated image could pretend to be showing a hurricane one year earlier);
  - indications of image use in other online sites, in order to check the context in which the image is being used.
- offer an intuitive and simple interface, regardless of how new or experimental the implemented technologies are;
- allow researchers to save, export, and share their results and conclusions to document and archive their decisions;
- clearly explain the mechanics of the algorithms used under the hood –investigators should not be expected to trust "black-box" methods;
- be easy to integrate to existing IT systems and corporate processes (workflows, APIs, browser plugins, etc.);
- be up to date with the market and state of technology (and easy to upgrade/enhance);
- provide comparable or better results than existing solutions (some of which are free of charge today). This is challenging given that existing solutions include powerful tools such as Google reverse Image Search and TinEye.

Currently, the process of verifying images is a rather laborious and time-consuming process; an indicative example is the Visual Verification Guide for Photos provided by First Draft News[6]. Standard procedures involve a first visual check and assessment, asking questions such as whether an image is "too good to be true" and checking for obvious visual clues that reveal information (e.g. "Are there license plates, road signs, other writings in the image?" "Are striking buildings, scenery and the like revealed?" "How does all this relate to what an image supposedly portrays?").

Other steps in the image verification process are to run the respective image through a reverse image search in order to find out if and when it was previously published online, or to find similar images that can corroborate the event and deliver more information – and possibly different pespectives. If further clues about the location of the image are needed, the image can be visually compared with other images that can be found in tools such as Panoramio[7] or Geofeedia[8].

[3] observers.france24.com/en/20151125-photo-female-terrorist-tub-fake-hasna-aitboulahc en

[4] https://www.tineye.com/

[5] https://images.google.com/

[6] http://firstdraftnews.com/resource/visual-verification-guide-photos/

[7] http://www.panoramio.com/

[8] https://geofeedia.com/

Exif metadata checks are another step on the way to assessing the credibility of an image, but in order to get a complete metadata report, often multiple tools have to be used. All in all, there are many tools and processes that are useful and can be used for the verification of images (Silverman 2014). The current problem is that much of what has been outlined above can only be accomplished using a variety of tools. Ideally, all processes would be combined in a single tool that supports the journalistic process of image verification and allows for forensics results to be archived. Journalists involved in the verification of user-generated images would be very keen to have such a tool at their disposal.

### State-of-the-art in image verification

Some of the requirements regarding image verification and forensic analysis described above are pretty straightforward from an engineering point of view. Metadata extraction, for example, is a process that can be achieved in nearly any programming language using third-party libraries, and visualization of geolocation information based on GPS metadata is a simple process that can be implemented using the publicly available APIs or open map resources. It should be noted, however, that the presence of metadata in an image is by no means guaranteed. In fact, it has been observed that most social media platforms tend to remove metadata[9] to protect their users' privacy.

Other features that are often listed as journalistic requirements are currently rather difficult to implement without substantial financial investment – reverse image search at Web scale is the most notable one. Thankfully, certain online services offer public access for free, and thus it is possible to incorporate them in a platform without having to re-implement the entire functionality from scratch. On the other hand, other aspects of image verification still constitute open research problems. The detection and localization of image tampering operations with no information outside the image itself (such as possession of the capturing device claimed to have taken the image) is a challenging task, and an active research field has emerged around it.

Typical ways in which images can be manipulated include image splicing and copy-moving. The former refers to the practice of copying a part of one image and inserting it into another, so as to give the impression that an additional element was present in a scene. The latter means taking a part of an image and duplicating in another location within the same image. This can both be used to falsely add more information (e.g., make a crowd seem larger) or remove it (e.g., copy-moving the background over items). The distinction between the two practices is important as different algorithms can be used to identify each one. The detection of copy-move forgeries is usually based on finding internal replications of blocks (Cozzolino, Poggi, and Verdoliva 2015) or keypoints (Ardizzone, Bruno, and Mazzola 2015) within the image. On the other hand, splicing localization algorithms assume that the spliced region differs from the rest of the image in some significant aspect. Isolating this information can provide an indication of whether the image originates from a single source or not. Various types of information can be used to this end: Color Filter Array (CFA) interpolation patterns (Ferrara et al. 2012), noise patterns (Mahdian and Saic 2009), JPEG blocking artifacts (Li, Yuan, and Yu 2009), JPEG Double Quantization patterns (Lin et al. 2009), and JPEG Ghosts (Farid 2009). Algorithms in this category extract the value of some feature for each location in the image (pixel or block) and return an output map that can be used to investigate local discrepancies.

## The REVEAL image verification service

### Platform overview

The REVEAL project[10] aims to bring together partners from industry and academia in order to advance the necessary technologies to analyze information disseminated in the Web and social media with respect to higher level modalities, such as reputation, influence, or credibility of information. Within its scope, one major use case is the verification of eyewitness media for journalism and news reporting. The REVEAL image verification service[11] that we present here constitutes part of our research towards the development of novel tools that correspond to the needs of news professionals for eyewitness media verification. It is currently in its alpha stage, with features still being added or under consideration, and ongoing improvements in terms of stability and speed. In designing and developing the platform, we are maintaining a close collaboration between professionals from the fields of computer engineering and journalism, and the layout and features are set up and revised based on this collaboration. While at this stage we have not yet established a formal evaluation framework by professionals, it is certainly a consideration for the future, as it could help us adjust the service to the actual needs of the field. Other attempts at creating journalistic verification frameworks have also actively implicated news professionals, with encouraging results (Diakopoulos, Choudhury, and Naaman 2012; Brehmer et al. 2014; Park et al. 2016).

The platform currently offers three classes of functionality: metadata extraction and visualization, tampering localization analysis, and reverse search integration. The first supports full listing of metadata in a selected image, display of any embedded thumbnails, and depiction of the image location on a map, if GPS metadata are available. The second consists of six tampering localization maps produced by different forensics algorithms, aiming to capture different types of potential tampering traces. The third is currently implemented by linking to Google Image Search and presenting results in a new browser tab.

Metadata are extracted from images, organized in categories (e.g. Exif, IPTC) and presented to the user. Investigators can then evaluate the information provided, to consider the possibility of the image having been tampered. In parallel, if GPS longitude and latitude values are present in the image metadata, they are extracted and used to pinpoint the

[9] http://www.embeddedmetadata.org/social-media-test-results.php

[10] http://revealproject.eu/

[11] http://reveal-mklab.iti.gr/


Figure 2: Analysis of a real-world forgery using the REVEAL image verification service.

location on OpenStreetMap[12]. This can help identify possible discrepancies between the claimed image location and the location where the image was actually captured. Finally, if the image metadata contain any embedded thumbnails, these are also shown to the investigator. The rationale is that, in some cases, the tampering process may have changed the image content but neglected to replace the thumbnails with new ones, thus discrepancies may be uncovered.

The integrated tampering localization algorithms were chosen to represent the state-of-the-art with respect to the detection of the most common types of tampering traces that can be detected in an image. Thus, the service features the following algorithms: Double JPEG Quantization (Lin et al. 2009), JPEG Ghosts (Farid 2009), JPEG Blocking Artifact Inconsistencies (Li, Yuan, and Yu 2009), Median Filtering Noise Residue, Discrete Wavelet High Frequency Noise Variance (Mahdian and Saic 2009), and Error Level Analysis (Krawetz 2007). Out of these six algorithms, four are well-established methods in the research bibliography, one (ELA) is the dominant image tampering localization method currently used by practitioners, and one (Median Filtering Noise Residue) is a method which, although not systematically evaluated as yet, is featured in one of the major online image forensics platforms and was thus included in the REVEAL service for the sake of completeness. The service also provides technical descriptions of each algorithm and guidelines on interpreting the algorithm output, targeted at users with limited expertise in image processing. Examples of successful detections and non-detections are also provided, so that users can have a guide of what they should be expecting to see. We consider such features to be necessary in a platform that is targeted at investigators outside the image processing research community.

[12]https://www.openstreetmap.org

Table 1: Comparison between the currently publicly available image forensics services.

| Feature | FotoForensics | Forensically | Ghiro | REVEAL |
|---|---|---|---|---|
| Double Quantization | | | | ✓ |
| JPEG Ghost | | | | ✓ |
| Block Artifact | | | | ✓ |
| ELA | ✓ | ✓ | ✓ | ✓ |
| Median Noise | | ✓ | | ✓ |
| Wavelet Noise | | | | ✓ |
| Copy-Move | | ✓ | | |
| Thumbnail | | ✓ | ✓ | ✓ |
| Metadata | ✓ | ✓ | ✓ | ✓ |
| Geotagging | | ✓ | ✓ | ✓ |

Currently, there exist a few similar services, offering tools for image verification including tampering localization analysis. Specifically, two online services (FotoForensics[13] and Forensically[14]) provide a number of features similar to the REVEAL service, while Ghiro[15] is an open-source image forensics tool with some overlapping functionalities to those offered by the REVEAL image verification service. Table 1 compares the features offered by these three services and the proposed one. With the exception of copy-move forgery detection, the REVEAL service covers all features currently offered by similar tools, and in addition offers a number of state-of-the-art algorithms that are not offered elsewhere. Furthermore, the implementation of the REVEAL service has been open sourced and the Java code is freely distributed[16].

## Current challenges

There are multiple issues to consider while developing such a service. Even problems that are to a large extent addressed by the current state-of-the-art, such as metadata extraction, can pose interesting issues: for instance, out of the (occasionally overwhelming) list of metadata, which fields are relevant to an investigator and should be prominently displayed? Should there be automatic metadata checks?

However, the area where we face the most challenges is image tampering localization. With the exception of copy-move detection, our service currently offers all tampering localization algorithms offered by other platforms, and an additional number of state-of-the-art algorithms. The reason we have not yet proceeded with copy-move localization is that such methods can become computationally expensive and may thus make the service less responsive – the JavaScript implementation provided by Forensically is an extremely lightweight algorithm with limited capabilities, and, to our knowledge, no other effort has been made towards a Web-based copy-move localization algorithm. In any case, even with the multitude of forensic features provided, we are still a long way from offering a service that can live up to the ideal expectations of an investigator, as described in the requirements. Such a service should be able

[13]http://fotoforensics.com/
[14]https://29a.ch/photo-forensics/
[15]https://www.getghiro.org/
[16]https://github.com/MKLab-ITI/image-forensics

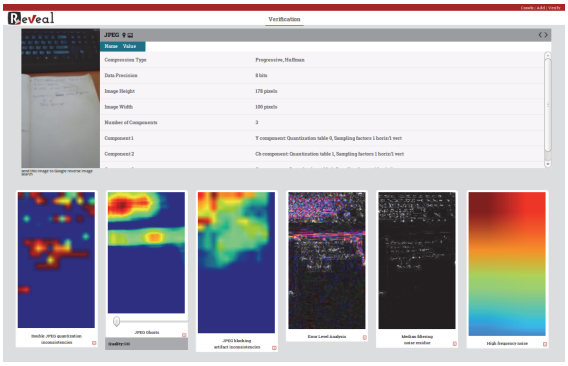


Figure 3: Analysis of an untampered image.

to produce a correct and unambiguous forensic report for a majority of the images found on the Web, while maintaining efficiency and scalability.

One major consideration is the ease with which an investigator can examine the algorithm results and come to an unambiguous conclusion, especially when they have no specialized background in image forensics. While a few algorithms produce probabilistic output maps indicating the probability that each region has been tampered, others produce values in arbitrary ranges; in such cases, interpretation may require at least a rudimentary understanding of the inner workings of the algorithms. Figure 2 offers a glimpse of the problem with respect to the analysis of tampered images: two algorithms (JPEG Ghost and ELA) produce output that is easily interpretable with only little training, and Double Quantization fails to detect anything and clearly expresses this in its output; yet, the remaining three algorithms produce extremely noisy output which is rather difficult to interpret. In this case, where the ground truth of the forgery is known (the forged region practically matches the Ghost output), it is easy to see that the three remaining algorithms do not produce any meaningful results. In real-world cases, where we are not certain what to expect, it is often hard for investigators to come to a conclusion by looking at such output, especially if they are not familiar with the algorithm operation. In the current state-of-the-art, things get even more challenging when dealing with untampered images. Figure 3 demonstrates the analysis output for an untampered image. It is far from straightforward for an investigator to conclude that the output maps are dominated by noise; ideally, we would like the analysis maps of untampered images to be uniformly blank as no tampered regions appear in the image, yet they are often dominated by image content and noise, producing outputs such as those in Figure 3.

Another issue pertains to the speed of processing. As investigators come across hundreds of images each day, we would ideally expect the framework to be able to produce forensic analyses in real-time, i.e. sub-second times. While certain algorithms have low computational cost, others have very high complexity. Although the issue of speed is related to the actual implementation choices as much as the theoretical computational cost of each algorithm, it should be noted that we are still far from the goal of real-time analysis, with some algorithms taking several minutes to complete, for high-resolution images. Finally, a major issue is the actual ability of image forensics algorithms to correctly identify forged images from untampered ones. In our previous work (Zampoglou, Papadopoulos, and Kompatsiaris 2015), we highlighted significant discrepancies between localization performance in datasets of artificially forged images and the performance on actual forgeries encountered on the Web. Investigating the extent of this discrepancy is extremely important for our work, as it implies that, the algorithms we choose to implement in our service based on their performance in the lab, may not show comparable effectiveness "in the wild". So, in the process of evaluating our platform, we decided to test the performance of the implemented algorithms in various contexts, effectively extending our previous evaluations.

## Evaluations

Since each algorithm aims at detecting different tampering traces, it is understandable that not all algorithms should be expected to work on all images. For example, most Double Quantization and Blocking Artifact-based algorithms work best with JPEG images of medium quality, while many noise-based methods fail after a low-quality JPEG compression as the resulting information loss erases the tampering traces.

We know that in the real world there exist many forged images that cannot be detected at all by algorithms. This can partly be explained by the degradation of detectable traces in images as they are transformed, resaved and re-posted on the Web and on social media outlets (Figure 4). However, trace degradation is not the only cause of tampering localization failure. Since localization algorithms have specific requirements concerning the tampering processes they can detect, there exist many cases in the real world where no algorithm can provide a successful detection. To assist evaluations with respect to the former issue, we published the Wild Web Tampered Image dataset (Zampoglou, Papadopoulos, and Kompatsiaris 2015) and made it publicly available[17]. In evaluating the second issue, we now present another dataset, which we call the Deutsche Welle (DW) Image Forensics Dataset[18]. The DW Image Forensics Dataset contains a small set of images that have undergone a number of tampering operations. There are six original images in the dataset, originating from three different sources: a smartphone camera, a semi-professional DSLR camera, and Flickr, in which two images taken with the DSLR camera were posted and then re-downloaded. The number of images is admittedly small compared to other tampering benchmark datasets. However, what distinguishes the dataset from others is the high resolution of the contained images, and the detailed documentation of the various tampering processes that were applied on the images. For our evaluations in this work, we will focus on one of the operations that is a combi-

[17] http://mklab.iti.gr/project/wild-web-tampered-image-dataset

[18] http://revealproject.eu/the-deutsche-welle-image-forensics-dataset/

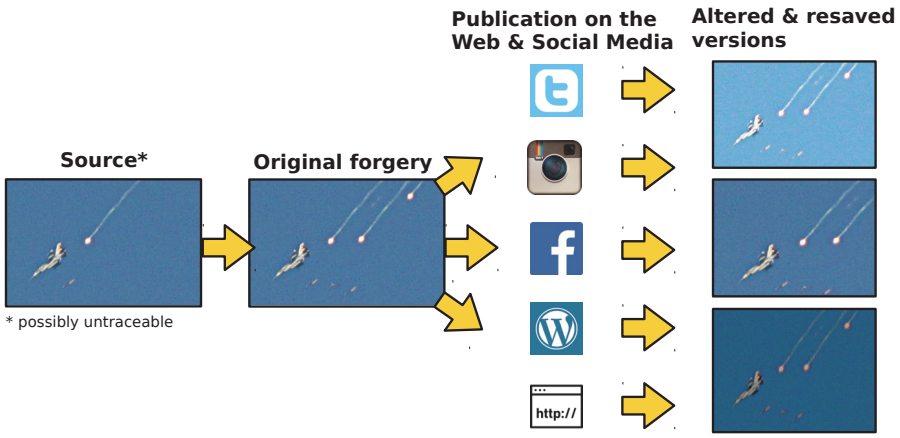


Figure 4: A forgery's lifecycle on the Web, from its creation to the analysis of some version of it by an investigator.

nation of copy-moving and in-painting in the images. Both these operations can leave traces that can be caught by many splicing localization algorithms. Indeed, such is the case for both smartphone images in the DW dataset (Figure 5). However, the same does not apply to the high quality images from the semi-professional camera. In fact, no algorithm in our possession can detect the forgeries in these images, despite the fact that they have not undergone further degradation following the forgery. One explanation for this is that, while the JPEG compression was not strong enough to leave detectable quantization or blocking traces, it was enough to remove CFA and noise patterns and make the forgery undetectable. In any case, the fact that an entire arsenal of splicing localization algorithms failed to detect the tampering on a simple, high-quality image that was resaved only once is indicative of the large distance that yet remains to be covered before we can claim to have solved the problem of tampering localization.

In order to have a more complete perspective on the performance of the implemented algorithms both on real-world and research forgeries, we proceeded to run a series of evaluations. Tests were executed on standard benchmark datasets, which allow more systematic evaluations, and real-world cases, which offer a greater challenge. We evaluated the six algorithms that have been integrated in the REVEAL image verification service. Throughout the evaluations, the algorithms are referred to using the following abbreviations: Double JPEG Quantization (DQ), JPEG Ghosts (GHO), JPEG Blocking Artifact Inconsistencies (BLK), Error Level Analysis (ELA), Median Filtering Noise Residue (MED), and Discrete Wavelet High Frequency Noise Variance (DWHF). These algorithms were tested on three well-established datasets: the Columbia Uncompressed Image Splicing Detection Evaluation Dataset (COLUMB) (Hsu and Chang 2006), the "realistic" dataset from (Fontani et al. 2013) (FON_REAL), and the training set used for the 1st IEEE-IFS challenge on Image Forensics[19] (CHAL). These three datasets contain images that were spliced, alongside ground-truth masks indicating the region where the splicing took place. Thus, they can be used to evaluate the discriminative capabilities of the localization algorithms' output.

For the evaluation of the algorithm performance we needed to examine whether the output map values in the tampered region displayed visible differences to the rest of the image surface. The Kolmogorov-Smirnov (K-S) statistic has served as a measure of this difference in the past (Farid 2009). We calculate the K-S statistic between the tampered region and the rest of the image. For untampered images we took an arbitrary square region in the centre of the image and performed a similar comparison. We then used a shifting threshold on the K-S statistic –images where the difference between the marked region and the remaining area was larger than the threshold were classified as detections. Figure 6 shows the True Positive-False Positive curves for the three datasets. A first observation is the different behaviour of the same algorithms for different datasets. In COLUMB, where the splices originate from different camera devices and the images are not JPEG-compressed, the noise-based algorithms (DWHF and MED) perform significantly better than anywhere else, while most JPEG-based algorithms do not exhibit strong performance. On the contrary, in FON_REAL where images have been JPEG-compressed at least once, some JPEG-based algorithms give very reliable results, while noise-based analysis does not seem particularly helpful. However, the absolute test is CHAL, where images are more realistic –there, practically all algorithms demonstrate limited success. To further examine the effect of recompression on the images, we ran evaluations on the same datasets, following JPEG recompression at various qualities. Thus, images were recompressed at qualities 65, 75, 85, 95 and 100. Figure 7 shows the effect of recompression on the three datasets. Each graph shows the True Positives rate at a threshold value that generates 5% False Positives. The sharpest effect is shown in the COLUMB dataset, where recompressions below a certain quality level completely destroy the discrimination capabilities of most algorithms. On the other hand, CHAL remains relatively unaffected, mainly because the algorithm performance is al-

[19]http://ifc.recod.ic.unicamp.br/fc.website/index.py

Figure 5: A successful detection on an image from the DW dataset. Left: the original image. Centre: the tampered image. Right: the output of the Double Quantization algorithm from our service.

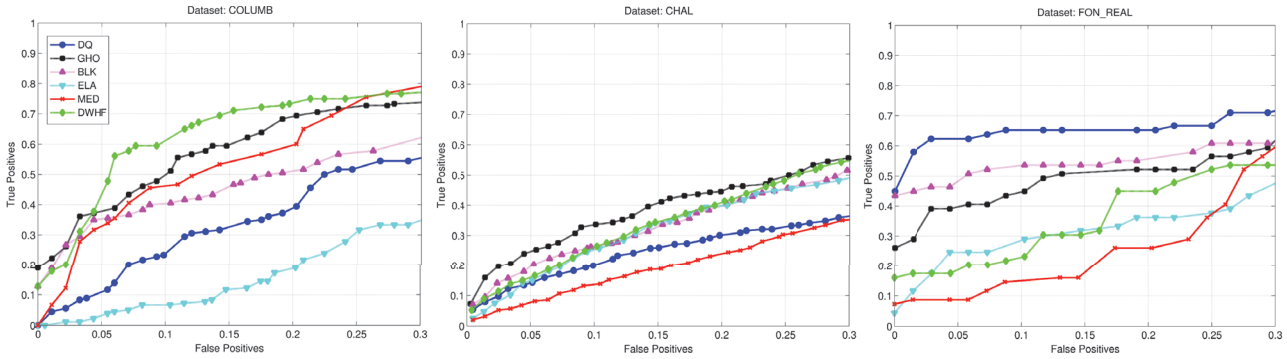


Figure 6: Performance of the six compared algorithms on three experimental datasets.

ready very low, due to past recompressions and transformations. A final, interesting observation is that, in some cases, algorithm performance goes up following a recompression at high quality. This can be attributed to the fact that a JPEG compression –at a quality that does not destroy too much information– can make certain features more prominent, which can then be detected.

While experimental datasets are very helpful in evaluating the performance of existing methods, it is still likely that they do not exhibit all the features of images encountered in the real world. This is the reason that we decided to also evaluate the selected algorithms on the Wild Web dataset (Zampoglou, Papadopoulos, and Kompatsiaris 2015), which consists entirely of real-world forgeries from the Web. When used for algorithm evaluations, the dataset suffers from one major disadvantage, in that it does not contain untampered images, and thus it does not allow the creation of TP-FP curves. What we opted to do instead, similar to (Zampoglou, Papadopoulos, and Kompatsiaris 2015), is threshold each output image at multiple values and apply different morphological operations on each version, thus creating multiple binary outputs. We then keep the binary output that best resembles the ground truth mask, and evaluate their similarity based on a surface-matching metric. If it is above a certain threshold (0.7 in this case), we consider the detection to be successful. Table 2 shows the results for the 80 forgeries of the Wild Web dataset. Overall, only 15 forgeries were detected by at least one algorithm, and in fact, as our approach may be overestimating the performance of the tested algorithms, it is likely that the actual algorithm performance is even lower than reported. In parallel, we used this large-scale evaluation to also measure the speed of our implementations. The second row of Table 2 gives the average time, in seconds, that each algorithm takes to process a single image. Overall, without underestimating the degree in which the implemented algorithms can provide valuable assistance in verifying eyewitness media, it is clear that, to a large extent, significant improvements in the state-of-the-art are still required to provide fully reliable tools for investigators.

Table 2: Performance on the Wild Web dataset.

| Algorithm | DQ | GHO | BLK | ELA | DWHF | MED |
|---|---|---|---|---|---|---|
| Detections | 3 | 12 | 3 | 2 | 3 | 1 |
| Time (sec) | 0.27 | 6.12 | 13.40 | 1.29 | 188.47 | 0.54 |

## Next steps

In this paper we presented a Web-based image verification service, currently in its alpha stage, offering many features that differentiate it from other free similar services. We are dedicating significant efforts in identifying the requirements from a practitioner's point of view, and attempt to implement, adapt and advance the state-of-the-art in image forensics in order to best address them. The close-knit collaboration between experts from the engineering and journalistic domains is guiding a process of development that aims to close the gap between practitioners' requirements and the

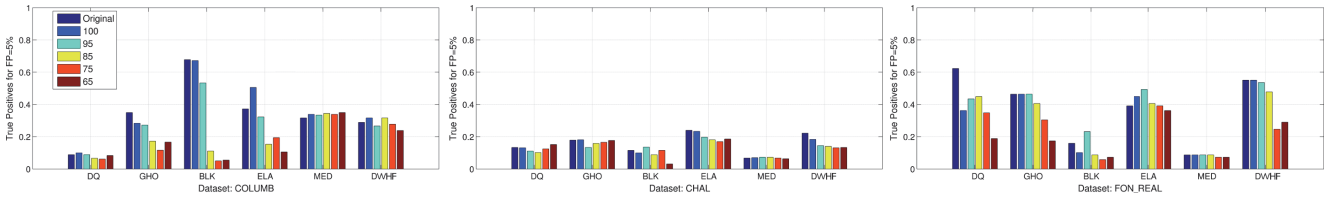


Figure 7: The effect of JPEG recompression on the six algorithms in our platform.

current state-of-the-art. The comparisons and evaluations we have conducted thus far point to how our service does constitute a step forward from the current state of the art, by offering a more complete image forensics toolset to support eyewitness media verification for journalistic investigations. However, our evaluations also bring forward several limitations of existing approaches, the most prominent being the low detection performance of splicing localization algorithms in real-world cases. It appears that, overall, existing methods and tools are a long way from offering a fast, reliable and easily interpretable solution to image verification. A necessary step would be the development of novel approaches and the improvement of existing ones, with the aim of increasing their detection rates. There is in fact significant hope in this direction, as image tampering localization is an active and productive research field. In parallel, however, there is work to be done in making algorithm results accessible and readable by non-experts. Algorithms that provide clear localizations of areas that are likely to have been tampered, as well as clear indications of areas that have not been touched (for example, by producing probabilistic output) are preferable and more useful than algorithms producing feature maps with little room for interpretation. In this sense, research should be encouraged to turn to methods targeted at non-expert users. Finally, we are still working to improve the response times of the service. Ideally, we are aiming at scaling up to the increased demand by news providers for professional on-the-fly eyewitness media verification.

## Acknowledgements

This work was supported by the REVEAL project, which has been partially funded by the European Commission (contract no. FP7-610928).

## References

Ardizzone, E.; Bruno, A.; and Mazzola, G. 2015. Copy-move forgery detection by matching triangles of keypoints. *IEEE Transactions on Information Forensics and Security* 10(10):2084–2094.

Brehmer, M.; Ingram, S.; Stray, J.; and Munzner, T. 2014. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics* 20(12):2271–2280.

Cozzolino, D.; Poggi, G.; and Verdoliva, L. 2015. Efficient dense-field copy-move forgery detection. *IEEE Transactions on Information Forensics and Security* 10(11):2284–2297.

Diakopoulos, N.; Choudhury, M. D.; and Naaman, M. 2012. Finding and assessing social media information sources in the context of journalism. In Konstan, J. A.; Chi, E. H.; and Höök, K., eds., *Conference on Human Factors in Computing Systems, CHI '12*, 2451–2460. ACM.

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

Ferrara, P.; Bianchi, T.; Rosa, A. D.; and Piva, A. 2012. Image forgery localization via fine-grained analysis of CFA artifacts. *IEEE Transactions on Information Forensics and Security* 7(5):1566–1577.

Fontani, M.; Bianchi, T.; Rosa, A. D.; Piva, A.; and Barni, M. 2013. A framework for decision fusion in image forensics based on dempster-shafer theory of evidence. *IEEE Transactions on Information Forensics and Security* 8(4):593–607.

Hsu, Y.-F., and Chang, S.-F. 2006. Detecting image splicing using geometry invariants and camera characteristics consistency. In *ICME*, 549–552. IEEE Computer Society.

Krawetz, N. 2007. A pictures worth... digital image analysis and forensics. Online article on: http://www.hackerfactor.com/papers/bh-usa-07-krawetz-wp.pdf. Accessed: 2016-02-26.

Li, W.; Yuan, Y.; and Yu, N. 2009. Passive detection of doctored JPEG image via block artifact grid extraction. *Signal Processing* 89(9):1821–1829.

Lin, Z.; He, J.; Tang, X.; and Tang, C.-K. 2009. Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognition* 42(11):2492–2501.

Mahdian, B., and Saic, S. 2009. Using noise inconsistencies for blind image forensics. *Image and Vision Computing* 27(10):1497–1503.

Park, D. G.; Singh, S.; Diakopoulos, N.; and Elmqvist, N. 2016. Supporting comment moderators in identifying high quality online news comments. In *Conference on Human Factors in Computing Systems, CHI '16*. ACM.

Silverman, C. 2014. *Verification Handbook*. European Journalism Centre. http://verificationhandbook.com/.

Zampoglou, M.; Papadopoulos, S.; and Kompatsiaris, Y. 2015. Detecting image splicing in the wild (WEB). In *ICME Workshops*, 1–6. IEEE.