TOWARDS THE VERIFICATION OF IMAGE INTEGRITY IN ONLINE NEWS

ABSTRACT

The widespread of social networking services allows users to share and quickly spread an enormous amount of digital contents. Currently, a low level of security and trustworthiness is applied to such information, whose reliability cannot be taken for granted due to the large availability of image editing software which allow any user to easily manipulate digital contents. This has a huge impact on the deception of users, whose opinion can be seriously influenced by altered media. In this work, we face the challenge of verifying news online by analyzing the images related to the particular news article. Our goal is to create an empirical system to verify the consistency of visually and semantically similar images used within different news articles on the same topic. The search provided by our system returns a set of images connected to the same topic, so to be easily compared and analyzed in order discover possible inconsistencies, which will help in classifying news as reliable or not.

***Index Terms*—** Media Verification, news

1. INTRODUCTION

The Internet is a system that enables immediate distribution of any type of digital media and it is more and more populated with user-generated contents. In particular, billions of images are nowadays publicly available thanks to the increasing diffusion of social networking services[1], thus leading to an overload of information to be parsed for any regular user. The social impact of such system is enormous and encompasses all aspects of our lives, including how we shape social relationships and how we form our opinions and share them with the rest of the world.

Moreover, the increasing diffusion of mobile phones have made citizen journalism easily accessible to people, as it happened, for instance, with the "Arab Spring", the "Occupy Wall Street movement" or the "2013 riots in Turkey", where large availability of technology allowed citizens to report breaking news even more quickly than traditional media reporters. As a matter of fact, the journalistic approach is relying more and more on user generated content and grass-root data [1].

In order to stay relevant, the media industry needs to adjust, adopt and provide new ways and new offers for an ever demanding public, while stressing the fundamental principles that make a content provider authoritative and trustworthy. Moreover, over the last decade, image editing has become a very common and easy practice for anyone, thanks to the wide spread of sophisticated graphic technologies, e.g., Picasa or Photoshop. Edited images are very often visually compelling and hard to be distinguished from original un-edited photos. An increasing trend towards the usage of edited images in every sector of our daily life, like news publishing, blogs or advertising, can be observed [2]. This often results in users deception [3] which may influence and manipulate the public opinion, from the self-esteem of teenagers and personal health choices[2] to public opinion in major political arenas[3]. Although manipulated images are often exposed, it may take weeks, and by that time it may have already influenced the opinion of millions of people. This may raise serious issues on the trustworthiness of digital multimedia on the web, since it casts increasing doubts on the face value of the information we daily access through the Internet [4]. This problem is becoming more and more serious, posing significative challenges to the worldwide media industry (see a recent example in Figure 1).

Fig. 1: The Charlie Hebdo march in Paris featuring state leaders from all over the world. An orthodox paper edited out the female leaders from the original image before publication.

The urgent need to add robustness to the information exchanged on modern communication highways through the diffusion of trustworthy media contents is evident. Solutions aimed at automatically verifying multimedia content published on the web (also with regard to the semantic information of the page) are then strongly needed and the scientific community recently started focusing on the main challenges in this direction[4] and studying related issues [5].

In this paper, we face the problem of analyzing online news by adding a new layer of resilience into the process of evaluating the validity of adopted imagery. Indeed, the proposed tool allows verifying the consistency of images used

[1] http://techcrunch.com/2013/01/17/facebook-photos-record/

[2] http://goo.gl/5Ba9wm

[3] http://www.globalissues.org/article/532/media-manipulation

[4] http://www.multimediaeval.org/mediaeval2015/verifyingmultimediause/

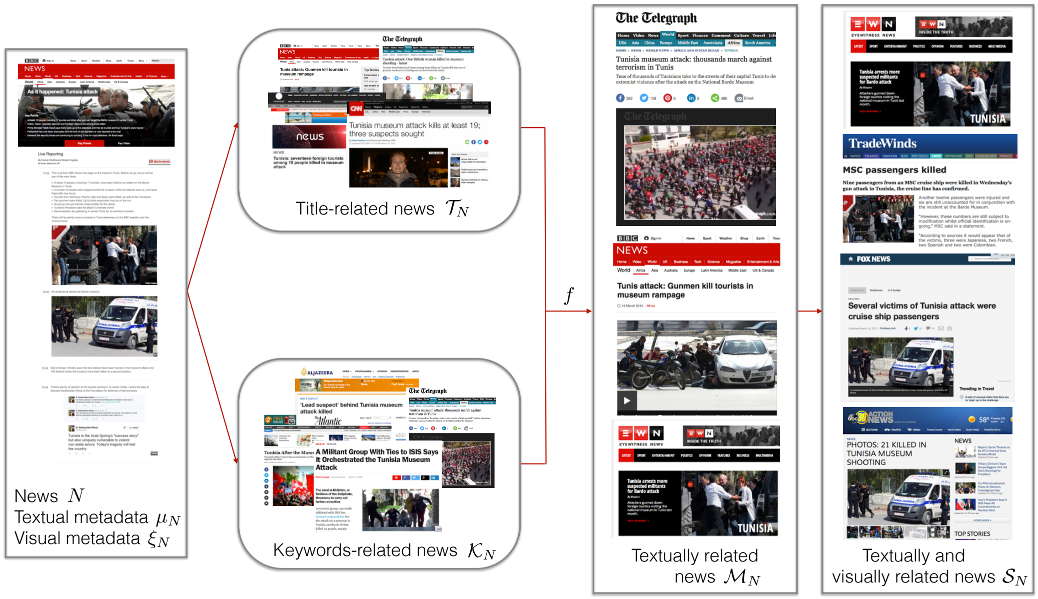

Fig. 2: Pipeline of the proposed model.

within a news article with other visually similar pictures related to the same topic. This can be done by visual inspection or exploiting multimedia forensics tools for detecting manipulation [6]. In particular, the search provided by the proposed approach would act as a support to the forensic analysis by providing a set of images connected with the same news, without requiring fully blind image evaluation.

All details about the proposed framework are provided in Section 2, while the implementation details of the proposed system are reported in Section 3. Section 4 provides the results of the experimental analysis, verifying the effectiveness of the proposed approach. Finally, in Section 5 we draw some concluding remarks, as well as a discussion on the open challenges in the field.

2. GENERAL MODEL

In the framework proposed, we suppose to have a web page reporting a news (a web news) and a number of attached images, and we want to verify the consistency of such images with other visually similar images related to the same topic. Precisely, the web news is submitted to a system which is able to provide as output a number of web pages, each of them concerning the same event *and* containing images that have some visual relationship with the ones of the original link. In this section we formulate the general model devised to carry out such analysis, while the actual implementation of each step is discussed in the next section.

As a first step, we assume that for each web news e N we can extract a structure μ_N containing *textual metadata* (the title, the body, the date and a list of keywords) and a set ξ_N of *visual metadata* (the images contained in the web page).

Our general goal is then to identify a set $\mathcal{S}_N$ of web news where for each $N' \in \mathcal{S}_N$ we have that both $\mu_N, \mu_{N'}$ and $\xi_N, \xi_{N'}$ are related by means of some metrics.

The system performs the following steps, which are summarized in the scheme reported in Fig. 2:

- **Metadata extraction:** given the input page N, μ_N and ξ_N are extracted

- **Selection on textual metadata:** a first set $\mathcal{M}_N$ of links is identified according to their similarity in terms of textual metadata. In particular, we suppose to determine a first set $\mathcal{T}_N$ of web news with a similarity in the title and a second set $\mathcal{K}_N$ with similarity in the keywords. The rationale behind this choice is the fact that title-based similarity will likely identify links that are strictly related to the very fact reported in the input news N and usually published the very same day. On the other hand, the keyword-based similarity contributes to broaden and diversify the results by identifying the topic of the news and providing also links published in a wider temporal range.

  Such sets of webpages are then used to obtain the set $\mathcal{M}_N$, i.e.,

  $$\mathcal{M}_N = f(\mathcal{T}_N, \mathcal{K}_N)$$

  where $f(\cdot,\cdot)$ expresses the strategy used to combine the two sets.

- **Selection on visual metadata:** a visual similarity analysis is performed between each image in ξ_N and each image in $\xi_{N'}$, $\forall N' \in \mathcal{M}_N$. This is done by means of a binary function $\mathrm{sImg}(\cdot,\cdot)$ such that two images I_1, I_2

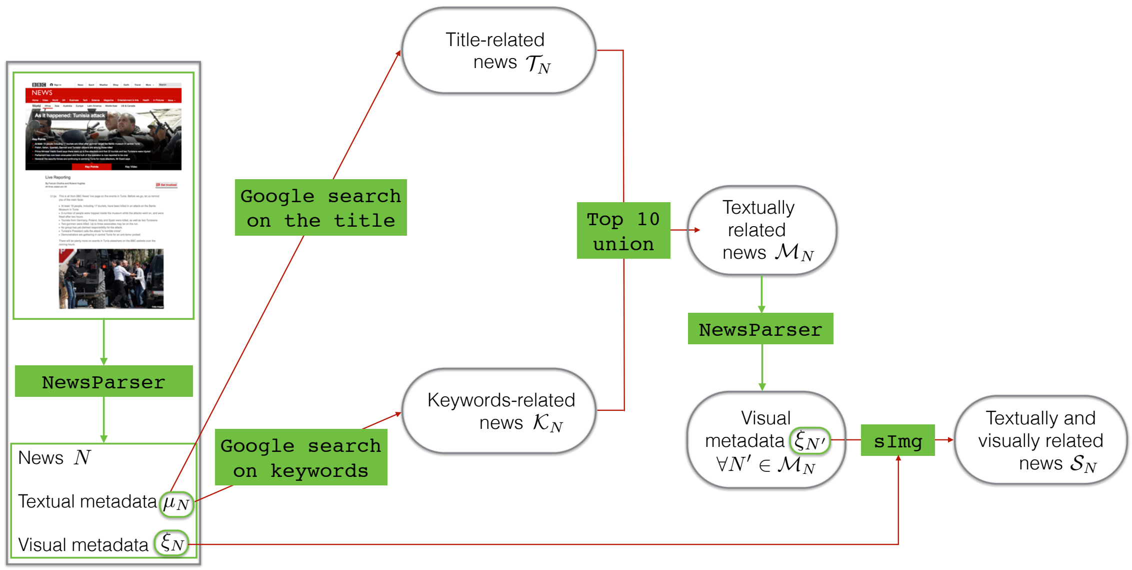

Fig. 3: Implementation scheme of the verification system.

are considered similar if $\mathrm{sImg}(I_1, I_2) = 1$ and not similar if $\mathrm{sImg}(I_1, I_2) = 0$ by means of suitable metrics described in the next section. Once the similarity check is performed on each pair of images, the final set $\mathcal{S}_N$ is defined as follows

$$\mathcal{S}_N = \{N' \in \mathcal{M}_N | \\ \exists I_N \in \xi_N, I_{N'} \in \xi_{N'} \text{ s.t. } \mathrm{sImg}(I_N, I_{N'}) = 1\}.$$

Thus, $\mathcal{S}_N$ identifies all the web news textually and visually related with the original one N, as depicted in Fig. 2

Finally, the user can visualize the set of original images and, for each of them, the corrisponding set of similar images detected among the textually related web news in $\mathcal{M}_N$. Under the hypothesis that an image $I \in \xi_N$ is the modified version of an original image appearing in some news of the same event (or viceversa), the latter should be detected by the system as element of $\xi_{N'}$ for some $N' \in \mathcal{S}_N$. At this stage, potential inconsistencies between the original and modified version of the same image (due, for instance, to an enhancement operation or a splicing) can be manually identified by the users, thus suggesting the use of multimedia forensics techniques for a comprehensive analysis aimed at determining which of the two has been tampered with and how. Moreover, we stress that the algorithm might also be applied recursively to each element of $\mathcal{S}_N$, in order to broaden the search and collect a higher number of image matches.

3. IMPLEMENTATION

Each of the phases described in Section 2 could be implemented in different ways according to a specific rationale and technical needs. In this section, we describe the tools and strategies we chose to use in this work in order to build an automatic system supporting the verification of online news, whose pipeline is summarized in Fig. 3.

As a first step, we create a Python module indicated as *NewsParser* which takes as input the URL of a web news N and provides as output both the textual and visual metadata μ_N and ξ_N. All the metadata are extracted via the *newspaper* Python module (available online[5]) with the exception of the keywords, which are obtained by means of a simple tagger function based on the *Natural Language Toolkit (NLTK)* open source platform[6].

The two sets $\mathcal{T}_N$ and $\mathcal{K}_N$ are considered as the set of links obtained by searching on Google the title and the list of keywords of N, respectively. This is done by means of *GoogleScraper*[7], a Python module available online which parses Google search engine results and allow users to extract all the links found. *NewsParser* is then applied to each URL to obtain textual metadata and the set of images contained in the webpage. All the data are stored and managed by means of a Python interface to SQLite3.

At this stage, it is necessary to define a selection strategy $f(\cdot,\cdot)$ to obtain $\mathcal{M}_N$ starting from $\mathcal{T}_N$ and $\mathcal{K}_N$. In our case, we decided to take the union of the first 10 links of $\mathcal{T}_N$ and the first 10 links of $\mathcal{K}_N$, i.e. the first 10 results of the two Google searches. Clearly, this is just one of the possible choices and, for instance, a different proportion between the number of elements taken from $\mathcal{T}_N$ and $\mathcal{K}_N$ would result in a higher or lower diversity in $\mathcal{M}_N$.

Once a set of textually related news is identified, we need to define a visual similarity metrics between two images. In this respect, many options are available in the literature but, among the others, we chose to employ SURF features mat-

[5]https://github.com/codelucas/newspaper
[6]http://www.nltk.org
[7]https://github.com/NikolaiT/GoogleScraper

ching and the correlation value, coupled with a face detector where MSER features matching is performed on the faces detected in the two images, if any. We are well aware of the fact that other measures could be employed, but we found these two approaches quite effective as a first attempt to cope with the problem, as showed in Section 4.1. We exploited the Matlab Computer Vision System Toolbox, which offers algorithms for both SURF computation (based on [7]), face detection (based on [8]), MSER features computation (based on [9]) and feature matching [10]. The output values of the built-in commands are properly combined and used to define the binary function $\mathrm{sImg}(\cdot,\cdot)$ determining the similarity of two images, that will be discussed in the next section.

4. TESTING AND RESULTS

In order to evaluate the effectiveness of our approach in revealing visual inconsistencies among semantically and visually related images, we performed several experiments according to different settings.

At first, we carried out a preliminary testing phase aimed at properly defining the criteria for the selection of the web news collected in $\mathcal{M}_N$, based on the similarity of their visual metadata with the one of the original input link ξ_N (Section 4.1). This step will lead to the identification of visually similar images among all the ones contained in each element of $\mathcal{M}_N$. Subsequently, several web news were submitted to the system and the automatic retrieval was performed recursively on the elements of $\mathcal{S}_N$, thus obtaining a more comprehensive set of images (Section 4.2). Finally, we considered some examples of modified images submitted to the proposed system Section 4.3), thus deeply investigating its capability of retrieving the original image or correlated image which may support and simplify the forensic analysis.

4.1. News selection based on visual metadata

In this preliminary phase, we focused on the selection of news in $\mathcal{M}_N$ based on their visual metadata by defining the binary visual similarity function $\mathrm{sImg}(\cdot,\cdot)$, so to derive the final set $\mathcal{S}_N$, In particular, we run the algorithm for some web news obtaining $\mathcal{M}_N$ and we manually classified the images in $\xi_{N'}$ for each $N' \in \mathcal{M}_N$ as visually similar or not to the images in ξ_N.

As significant examples, we considered in this step the following news:

Charlie	*Charlie Hebdo attack: Three days of terror* [8]
Pam	*Cyclone Pam: Dozens feared dead in Vanuatu in "one of worst storms in Pacific history"* [9]
Kathmandu	*Turkish Airlines Jet Skids Off Foggy Runway in Kathmandu, Nepal* [10]
Nemtsov	*Russia opposition politician Boris Nemtsov shot dead* [11]

For each news, a number of correlated links, together with their associated images, were retrieved by the system. Each pictures was then manually compared with the visual metadata of the starting web news and classified as visually similar or not. Shown in Table 1 are the number of retrieved images, as well as the matched ones, for all the four examples here reported.

News	N. of image pairs	N. of similar images
Charlie	5532	92
Pam	2320	51
Kathmandu	200	7
Nemtsov	184	4

Table 1: For all the 4 examples here considered, the number of images retrieved by the system, as well as the number of images matched with the original input one based on their similarity are reported.

The performed manual matching served to build a reliable baseline needed for the definition and verification of a valuable automatic metric for image similarity. In Fig. 4 the SURF matching (horizontal axis) and correlation values (vertical axis) of each pair of images are reported, for each one of the considered news. Please note that the horizontal range of the represented segment shows the minimum and maximum value of SURF matching. The red and blue segments correspond to image pair manually classified as similar and not similar, respectively.

It can be noticed that the correlation values for similar image pairs are well clustered and mainly located in the upper-right part of the plot, showing that SURF matching and correlation can provide a representative measure. Moreover, the faces present in the two images are detected and matched by means of MSER feature matching, in order to specifically deal with images that are composite of people, which is a quite common case in web news.

As such, the similarity function $\mathrm{sImg}(\cdot,\cdot)$ is a combination of these three factors (SURF matching, correlation value and face matching, if any), which are ruled by a set of thresholds empirically determined based on the preliminary tests on the selected four news. Despite the low number of news considered and the simple approach adopted, we verified that this was suitable for our preliminary analysis, although a deeper analysis in this respect would help in gaining accuracy and will be subject of future work. In Fig. 5 we report an example of images automatically selected as similar, based on the SURF matching and correlation strategy (left pair) and based on the face detection matching (right pair).

4.2. Recursive application

In order to refine and extend the set of retrieved contents $\mathcal{S}_N$, we applied a recursive procedure. In particular, we recursively submitted the found news in $\mathcal{S}_N$ to the system, thus retrieving more related news articles. In Fig. 6, we report an example of such recursive application for the news **Charlie**

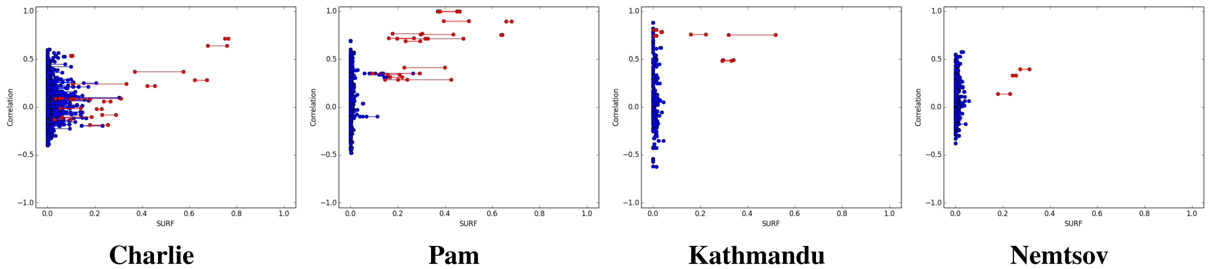

Fig. 4: Representation of correlation and SURF matching values of image pairs for each news.

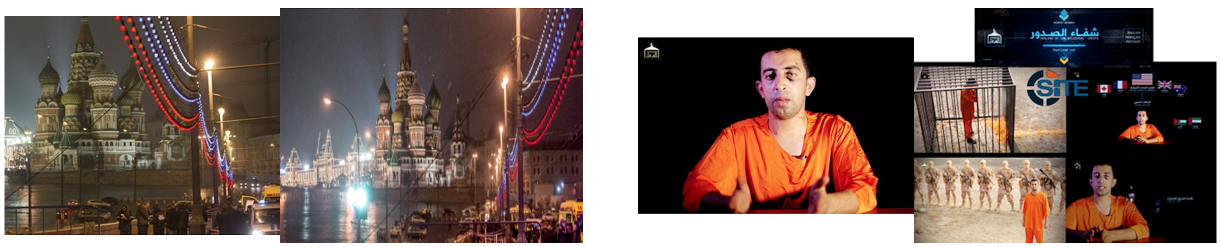

Fig. 5: Example of image pairs automatically classified as similar by means of correlation and SURF matching (on the left) and by means of face detection matching (on the right)

mentioned before. It becomes clear that the number of retrieved images significantly increased, as well as their diversity. When replicating the same procedure with other news, we achieved very similar results, as showed in Table 2 where the number of found images increased at each iteration of the recursive process.

Iteration	1	2	Iteration	1	2
Charlie	92	885	**Kathmandu**	7	10
Pam	51	825	**Nemtsov**	4	282

Table 2: Number of similar images retrieved at each iteration.

4.3. Simulations of manipulated images

Finally, we wanted to test the ability of our system in verifying the integrity of the starting images by retrieving images that are visually similar and related to the same topic. In other words, we wanted to observe the behavior of the system when a manipulated image is associated to a given news. Although there have been many examples of modified images published in online news, an *a posteriori* analysis by means of our system is quite hard to perform. This is due to the fact that, when exposed, fake images and their corresponding webpages are generally removed from the web. Moreover, even if the URL is still active the system will likely retrieve webpages regarding the fact itself that the pages was revealed as fake, thus generating noise in the results.

For these reasons, starting from web news of a certain topic we created a fake webpage with very similar textual metadata and containing manipulated versions of one or more attached images. By doing so, we simulated the real-world scenario in which a news is published on the web with an edited picture.

We then applied our algorithm (with two iterations) to the page we created and we observe that the system generally finds the original versions of the images. Two examples are reported in Fig. 7: the first depicts to the MotoGP champion Valentino Rossi during the Qatar race in 2015 (we deleted the the adhesive of a sponsor from his suit) and the second was taken during a parade organized after the Charlie Hebdo attack (we removed the big pencil held by the man in the foreground). We find the original images on the left, the edited ones in the center and the similar one retrieved by the system on the right.

This serves as a valuable example of demonstrating the effectiveness of our method in retrieving similar images connected to the same news. In this case, the visual comparison of the two images immediately leads to the detection of the modified part. However, in case of more sophisticated manipulations which are hardly visible just at visual glance, specific digital forensic techniques could be employed. Please note that in this work we did not make use of any, since this will be considered as future work given this paper to be a preliminary approach to an unexplored area of research. On the other hand, we had a case of edited image for which the original version was not found by the system. This is due to the fact that it was published in less popular websites which were not provided by the Google search. As future development, we plan to broaden the list of textually related links considered (that we currently limit because of the computational complexity in the image similarity computation) as a possible approach to overcome such limitation.

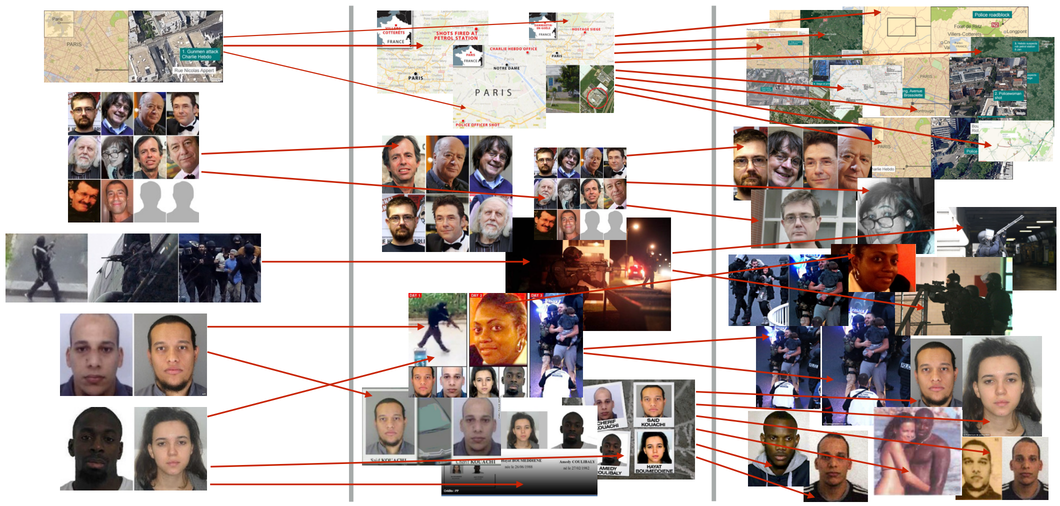

Fig. 6: Two iterations of the algorithm for **Charlie**.

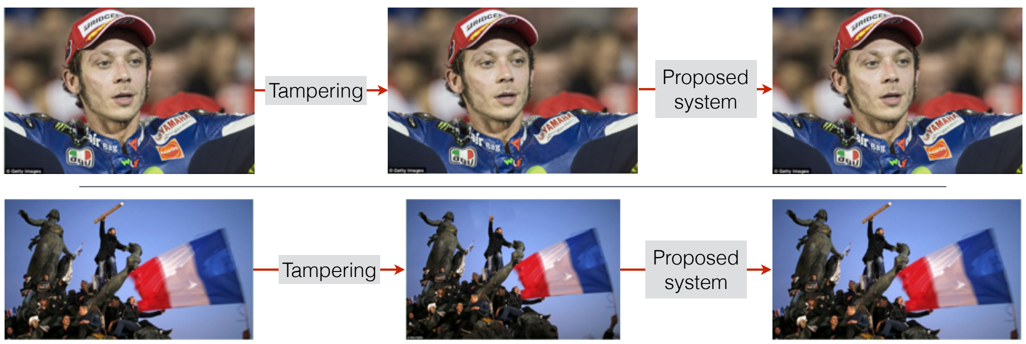

Fig. 7: Examples of edited images for which the system retrieved the original version.

5. DISCUSSIONS AND FUTURE DIRECTIONS

In this work we presented a general framework towards the identification of reliable web news through the analysis of the associated multimedia contents. In particular, we designed a novel system able to automatically retrieve news over the web related to the same topic. Such set of news is associated to some particular images, which we compare in order to verify their integrity. The proposed method is simple, yet very effective and we verified its effectiveness on a few number of cases. To the best of our knowledge, this is a first approach in the direction of image verification related to news online. As such, it represents only a starting point and much work can be done in the future, e.g., widening the image search, improving the visual features exploited for the similarity metric. We envision also the application of the proposed approach for detecting decontextualized images, i.e., images used in completely different contexts. This could be done by reversing the here proposed system, thus first retrieving similar images of a web news and then comparing their related textual information with respect to the original one.

6. REFERENCES

[1] N. Heise and J. Spangenberg, "News from the crowd: Grassroots and collaborative journalism in the digital age," in *ACM WWW Companion*, 2014, pp. 765–768.

[2] V. Conotter, D.T. Dang-Nguyen, M. Riegler, G. Boato, and M. Larson, "A crowdsourced data set of edited images online," in *ACM Workshop on Crowdsourcing for Multimedia*, 2014.

[3] V. Conotter, D.T. Dang-Nguyen, G. Boato, M. Menendez, and M. Larson, "Assessing the impact of image manipulation on users' perceptions of deception," in *SPIE Electronic Imaging*, 2014.

[4] D. L. M. Sacchi, F. Agnoli, and E. F. Loftus, "Changing history: Doctored photographs affect memory for past public events," *Appl. Cognit. Psychol*, , no. 21, pp. 1005–1022, 2007.

[5] C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schifferes, and N. Newman, "Challenges of computational verification in social multimedia," in *ACM WWW Companion*, 2014.

[6] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, pp. 1–22, 2013.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Surf:speeded up robust features," in *Computer Vision and Image Understanding (CVIU)*, 2008, vol. 110, pp. 346–359.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 511–518.

[9] J. Matas, O. Chum, M. Urba, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference*, 2002, pp. 384–396.

[10] M. Muja and D. G. Lowe, "Fast matching of binary features," in *Conference on Computer and Robot Vision (CRV)*, 2012.