Filtering Images Extracted from Social Media in the Response Phase of Emergency Events

Sara Barozzi
Politecnico di Milano, Milan, Italy,
sara.barozzi@mail.polimi.it

Amudha Ravi Shankar
University of Geneva, Geneva, Switzerland,
amudha.ravishankar@unige.ch

Jose Luis Fernandez-Marquez
University of Geneva, Geneva, Switzerland,
JoseLuis.Fernandez@unige.ch

Barbara Pernici
Politecnico di Milano, Milan, Italy,
barbara.pernici@polimi.it

ABSTRACT

The use of social media to support emergency operators in the first hours of the response phases can improve the quality of the information available and awareness on ongoing emergency events. Social media contain both textual and visual information, in the form of pictures and videos. The problem related to the use of social media posts as a source of information during emergencies lies in the difficulty of selecting the relevant information among a very large amount of irrelevant information. In particular, we focus on the extraction of images relevant to an event for rapid mapping purpose. In this paper, a set of possible filters is proposed and analyzed with the goal of selecting useful images from posts and of evaluating how precision and recall are impacted. Filtering techniques, which include both automated and crowdsourced steps, have the goal of providing better quality posts and easy manageable data volumes both to emergency responders and rapid mapping operators. The impact of the filters on precision and recall in extracting relevant images is discussed in the paper in two different case studies.

Keywords

Rapid mapping, floods, information extraction, filtering, crowdsourcing

INTRODUCTION

During the first phases of crisis response after an emergency event such as a flood, a storm, or an earthquake, there is a need for creating awareness about the ongoing event and for providing as much information as possible to the emergency operators. Social media are considered a potentially valuable source of information to support gathering information from people living near the event location or informed about it. Several initiatives based on social media have been proposed, as well as tools to manage automatically the large amounts of data produced by social media in those cases (Imran et al. 2015), that can reach hundreds of thousands or even millions of posts related to the event within hours. It is well known that among this large amount of data, only a small amount of it consist of valuable information. For instance, in an analysis of Dresden floods in 2013 (Fohringer et al. 2015), from 15 millions of tweets in the flood period, only 5 images were considered useful to estimate water levels in the affected area.

Very recently, in (Alam et al. 2018) the extraction of relevant images from social media has been discussed with the goal of providing only relevant images about the event, eliminating duplicates and using supervised machine learning techniques to assess the damage's intensity in the area of interest. In fact, the concept of relevance is heavily dependent on the task to be performed, be it an assessment of water levels as in (Fohringer et al. 2015), or an assessment of the intensity of damage as in (Alam et al. 2018).

In the present paper, we focus on the task of extracting images which are useful for rapid mapping purposes, for a fast provision (hours-days) of geospatial information to support emergency management activities immediately following an emergency event.

The goal of this paper is to discuss how information retrieved from social media posts extracted from Twitter, can provide a visual evidence about a disaster.

In particular, we discuss how different filtering techniques can support the selection of the images which are relevant to an event from the large number of incoming social media posts. The different filtering options, which include automatic filtering and crowdsourcing, are presented and analyzed in detail, focusing mainly on two parameters: precision and recall. Precision allows us to provide the operators with images which are really relevant to the event, while recall evaluates how much a filter could result in a loss of contents, since it filters our possibly useful images.

The paper is structured as follows. First, we discuss the state of the art, then we introduce the crisis maps created in the E2mC project and how visual information can be retrieved from social media. Then we discuss possible filters to reduce the amount of irrelevant information provided to the volunteers for crowdsourcing and to emergency rapid mapping operators. The evaluation of the effect of filters in terms of precision and recall is then performed in two cases studies, discussing possible threats to validity and future work.

RELATED WORK

Information systems for emergencies have been rapidly developing since the first successful experiences with Ushahidi[1] (Goldstein and Rotich 2008) and its use in the Haiti earthquake [2].

Several other initiatives focus also on real time collection of information, such as in SensePlace, focusing on near real-time access for big streaming data (MacEachren et al. 2011; Pezanowski et al. 2018) and TWINE (Nozza et al. 2017) for real-time analysis of posts.

Many studies have been carried out in the last years regarding the extraction of relevant information through the analysis of the textual content of the posts published during disasters. On the other side, the visual contents have not been studied that much. (Peters and Albuquerque 2015) have demonstrated that the images extracted can increase awareness on the situation and that they can be indicators of the relevance of the post in which they are contained; however some authors underlined that to obtain, manage and analyze them automatically is complex (Murthy et al. 2016). They reported an analysis of how the contents of the posted images change during the different phases of the disaster, by classifying them manually. (Gupta et al. 2013) tried to filter fake images extracted from social media during hurricane Sandy in 2012. Since they noted that traditional image analysis methods were difficult to apply, they used tweet-based features to create a Decision Tree classifier that reached the 97% of accuracy in detecting fake pictures. (Chen et al. 2013) built an automated classifier utilizing text, image and social context features to distinguish visually-relevant and visually-irrelevant tweets obtaining a macro F1 of 70.5%. Another recent work from (Alam et al. 2018) shows a combined use of human and machine computing to process social media images for damage assessment during a disaster. They propose a real-time social media image-processing pipeline where unwanted images are discarded through a deduplicating passage, based on perceptual hashing, and through a relevance filter, based on Convolutional Neural Networks architectures. The images are then labelled and classified by a crowd of volunteers and used to train the automatic machine classifier to assess the level of damages in area as severe, mild or none. Their results show that this approach can help to gain situational awareness during an ongoing event. They also show that filtering the images allows increasing the precision of the classification both of the human and of the machine system.

PRODUCING CRISIS MAPS IN E2MC

The E2mC project[3] has the goal of producing crisis maps for emergency events, to support rapid mapping activities performed by operators in Emergency Mapping Services (EMS) of Copernicus[4], the European initiative to provide timely and accurate information derived from satellite data and additional sources on the ground.

As described in (Havas et al. 2017), the E2mC platform is composed of a set of tools aimed to support the analysis of social media posts, extracted from Twitter, Flickr, and YouTube. The goal is to provide EMS rapid mapping operators with timely and relevant visual information extracted from posts. The processing of posts is performed in near real-time, as the event progresses. The starting point is a crawler that collects posts through the social media APIs. As most of the posts are not natively geolocated, in particular in Twitter and YouTube, crawling is followed by a geolocation tool, called CIME (Context-based Image extraction) (Francalanci et al. 2017), that localizes the posts as precisely as possible from the textual contents of each post using OpenStreetMap[5] as a gazeteer. Only the

[1] https://www.ushahidi.com/
[2] https://www.ushahidi.com/blog/2010/04/14/crisis-mapping-haiti-some-final-reflections
[3] https://www.e2mc-project.eu/
[4] https://www.copernicus.eu/services/emergency
[5] http://www.openstreetmap.org/

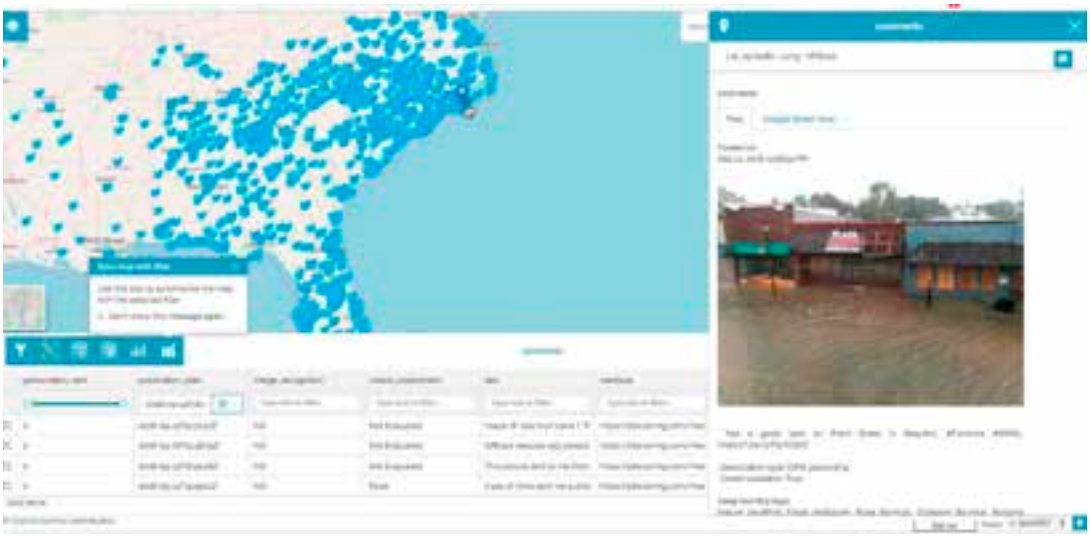

Figure 1. Witness user interface for the E2mC crisis maps

posts that contain an image or a video and are located at least at locality level (town or village) are retained for further analysis.

Figure 1 shows the webGIS interface developed in E2mC, called Witness. The Witness interface shown in the figure provides a crisis maps with located tweets (displayed in the center of their OSM location) and gives the possibility to show the contents of each tweet, accessing Twitter through its APIs. In addition, a set of possible manual filters is shown, to select localities at different levels of precision (from locality level down to precise points in the map), to examine only relevant tweets as evaluated from crowdsourcing, and to select only images detected as relevant to the specific type of event (e.g., floods).

The tweets are processed near-real time during the event, so that the classifications of the filters are incrementally populated.

In the present paper, we present different ways to reduce the number of tweets, and in particular using a set of filters which are discussed in detail in next section. Some of the filters illustrated in the following are part of the core E2mC platform, others, like the color and text filters, are proposed in the present paper and evaluated in the present work.

Two case studies are considered in the following parts of the paper, with datasets containing posts from Twitter which are geolocated at least at locality level and include an image: i) Southern England Floods in 2014 (corresponding to EMS activation number EMSR069, started on Feb. 10, 2014), consisting in 343 tweets, collected in the period February 10-14, 2014; ii) the Hurricane Florence over USA East coast in September 2018 (EMSR311), where the storm continually dumped heavy rain along coastal areas from September 13, 2018; in this case more than 28,500 posts have been crawled, of which we consider a data set of 2,940 posts collected in 24 hours on Sept. 13-14, 2018, from which we randomly sampled 301 posts for performing manual annotations for the analysis work presented in this paper. Both datasets are available on request from the authors.

FILTERS

During a disaster, the main problem is managing timeliness: for the operators it is necessary to have valuable information on the affected area in a short time. Images coming from social media can help in obtaining this information, but, on the other side, the volume of the data extracted may be too high to be managed within the short time frame they have. In addition, the data extracted is usually of low quality, since the images obtained from the crawling includes a high percentage of irrelevant information. Thus, it is necessary to send readily usable aggregated and cleaned data to the emergency rapid mapping operators and emergency responders. Consequently, many filters have been designed and analyzed to understand their potential in reducing the volume and improving the quality of the set of images available.

In the next subsections, we introduce a set of possible filters that can be applied to reduce the volume of posts with associated images to be provided to rapid mapping operators and we analyze their efficacy in improving the section of posts for rapid mapping. The following filters are considered in this work:

- Redundant information reduction
  - Duplicate reduction
  - Retweets elimination
- Image filtering
  - Colour-based filtering
  - Elimination of pre-event images
  - Face detection
  - Text detection
- Crowdsourcing
  - Crowd-sourced relevance tagging task

For each filter, a short description is provided in this section, illustrating the main characteristics, parameters, its effects, and time to be executed on a laptop HP Pavillon, with Intel Core i5 and 8GB RAM. In the following section, the filters are evaluated in the two case studies examined in this work.

Redundant information reduction

Duplicate reduction

During disasters, people use social media to post a huge amount of information, often sharing pictures they are seeing on pages of other users. Thus, the goal of this filter is to detect all the duplicates and near duplicates (that are pictures very similar to the original but with little modifications like addition of borders, colour enhancement and luminosity changes, addition of text, cropping), since they usually do not provide new information. However, it can happen that the same image is posted with different textual descriptions that may bring additional and useful information. The goal of the filter is to consider duplicate images only once in the filtered dataset.
The detection of duplicates can be implemented in different ways. The one adopted for the E2mC project provides the use of image hashing (in particular the perceptual hashing). Perceptual hashes are fingerprints of the pictures, created on the basis of the features and visual content of the images, so they are similar to each other if the images contain similar features. The perceptual hash is obtained through the Python library *phash* that is an open source software library. For all the pictures, the hash is computed and it is compared with the set of all the hashes found in the data set using the hamming distance, that is the number of positions at which the bits of the two strings of the hashes differ. If the distance between two hashes is lower than a certain threshold, then the pictures are considered duplicates. The threshold has been fixed to 15 since analyses on the data sets shown that this value guarantees the detection of a high number of duplicates while committing a low number of errors. This system has the advantage of being very fast (one second per image) since it just compares simple strings.
A modification that can be done to the previous algorithm consists in the addition of a feature detection system. A SIFT detector is used to identify the small characteristics of the pictures and compare them. These features are then compared with the feature of other images to assess if they match or not. With this improvement it is possible to detect more near-duplicates but the computation time increases 5 times.

Retweets elimination

A retweet is the re-post or the forward of the message posted by another user. During a disaster, the most impactful images or the funniest ones are retweeted several times. For the E2mC project it is important to understand if retweets bring new information or if they only increase the volume of the dataset. Retweets repost the same image and usually also the text is not changed, so they spread duplicated information. They can be identified by the initial "RT" at the beginning of the post's text, so they are easy to be recognized automatically. The goal of the analysis presented in this paper is to evaluate the impact of eliminating retweets on the resulting dataset, in terms of evaluating the loss of relevant images.

Figure 2. Examples of a white image, a black image and an image with peaks

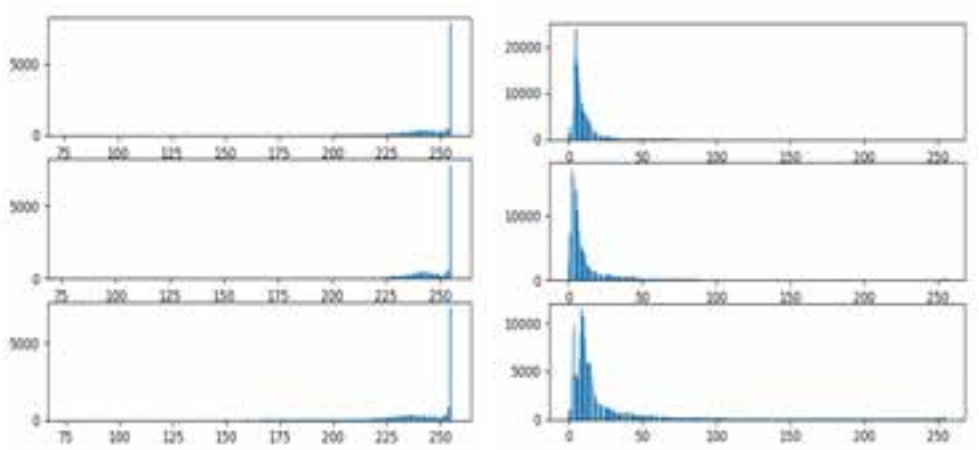

Figure 3. Histogram of a light image (left) and of a dark image (right)

Images filtering

Colour based filtering

The aim of this tool is the filtering of the images on the basis of their colours. During a disaster, pictures containing graphs, weather reports, jokes, texts are published, and they are not useful for disaster rapid mapping. This kind of images represent a huge source of noise during the analysis of the crawled posts. Therefore, three filters have been created to eliminate pictures that are too light, too dark or that contain graphs, drawings, not "taken photos". An example of these categories is reported in Figure 2.

The filter is based on an analysis of the image's intensity value histogram, that represents the distribution of the pixel values in the picture. Pictures that are very light have histograms shifted towards the higher values, while dark images have a histogram concentrated at the lower values. The difference is shown in Figure 3. There are three histograms because the levels of the RGB (red-green-blue) in the image are analyzed separately.
Pictures with graphs and drawings, that have a part of the picture of the same colour, present peaks of values in correspondence of these colour intensities. The difference between an image with peaks and a natural image can be seen in Figure 4.

For each filter a threshold is fixed to understand if the image should be discarded. The thresholds have been defined after an accurate analysis of the histogram of some test images. The thresholds for the "white" and "black" filters are based on the quartile (0,25,50,75,100 percentiles) subdivision of the histogram values. In fact, if the first quartile, that is 25, is high, it means that there is a concentration of pixel values towards the right side of the histogram, while if the third quartile, that is 75, is small, it means that values are concentrated in the left side of the histogram, meaning that a lot of pixels have a dark colour. In addition, the percentage of dark and light pixels in the image can be computed. Thus, if the first quartile is higher than 200 and the percentage of light pixels is higher than the 80% of the total values, the image is classified as 'white', while if the third quartile is lower than 35 and a percentage of dark pixels is higher than the 80% of the total values, the image is classified as 'black'. The images with peaks are, instead, identified as the images that have a total number of peaks higher than 3. A peak is found when the

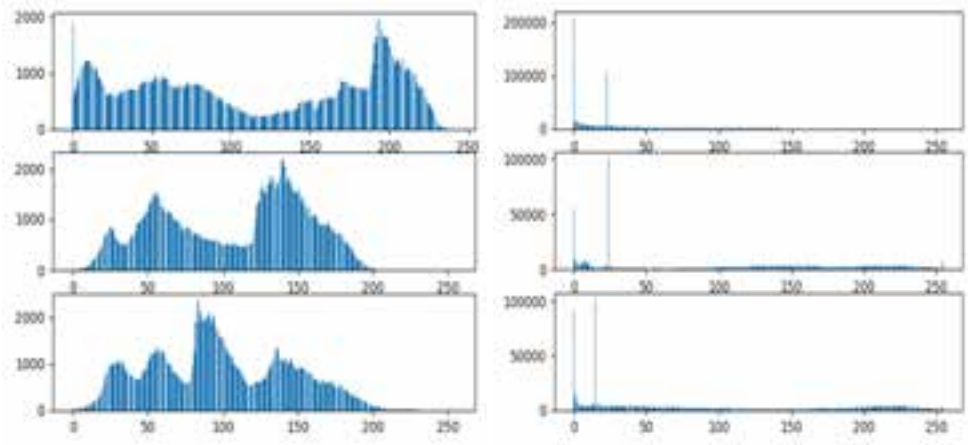

Figure 4. Example of a natural image histogram (left) and an image with peaks (right)

difference between the pixel intensity's before and after the analyzed pixel is high. This difference has been fixed to 3,500, but it can be lowered to create a filter that discards more images.
The classification of the images with these tools is fast and requires about 1 second for each picture.

Pre-event images

A problem that has been raised regarding the use of pictures posted in social media as a reliable source of information is the presence in datasets of pictures coming from previous events. These pictures can be mixed up with pictures of the current event, leading to the non-credibility of the whole dataset and to errors in the damage assessment of the affected area. The tool that has been used to face this issue is Google images[6].

This application offered by Google allows the user to search among all the pictures that have been posted on-line by keyword, by URL or uploading a picture to search all the similar ones present in the web. The parameters used to classify an image as being old are: (1) The presence in the Google results of older dates than the date of beginning of the disaster. (2) The presence in the results of pictures equal to the given one. (3) The presence of the searched image inside the website given as a result by the Google reverse image search.

Some problems limit the potentialities of this tool. There is a limitation on the call to the API that makes it necessary to add some waiting time in the process that increases the computational time of the algorithm. In addition, due to copyrights and recent policies on privacy, some pictures are no more freely available on the Internet. Because of this, some pictures coming from previous events may not be found.

Faces detection

Another kind of pictures that are present in the dataset during an emergency are the pictures with faces. These pictures often are not useful, such as jokes and selfies, but include also pictures extracted from television news reports, that in some cases could report useful information. Because of that, this filter should be used in case of overload of incoming pictures and when there is the need of reducing the dataset.
The tool, that exploits a built-in face classifier and detector, OpenCV[7], finds each face in the picture and draws around it a rectangle. Checking the number of faces found and the total area of the rectangle (or rectangles), it is possible to understand if the faces occupy the biggest part of the picture, thus meaning that it is not useful. A threshold value is needed to specify how many possible "face features" each candidate rectangle should have to consider it a face. In the current implementation, the threshold has been fixed to 8. This tool requires 0.5-1 seconds to evaluate each picture and it is very reliable. The limitation of this algorithm is related to the built-in cascade XML provided by OpenCV for the face recognition that is able to detect only frontal faces.

Text detection

Text detection can be used to reduce the volume of relevant images that arrive to geolocation crowdsourcing volunteers and to emergency rapid mapping operators, to provide a set of images easier to geolocate, an important characteristic when dealing with rapid mapping. In fact, the presence of words in an image can be related to road signs, license plates, shop signboards which are features that can help to find the place where the picture was taken.

[6]https://images.google.com/
[7]https://opencv.org/

Figure 5. Examples of relevant images with words detected in them

It is useful, in particular, in case of big datasets, since volunteers and emergency rapid mapping operators, that are overloaded with work, would prefer to see only the most relevant pictures.
The filter is a text detector created with the Python OpenCV's EAST text detector. The parameter that changes its performance is the probability that the area analyzed in the picture contains words. An area contains text if the probability is higher than a certain threshold, that has been fixed to 0.99 to ensure a reliable result in the finding of words. The time required to evaluate each picture is less than 2 seconds. Time can be reduced making the dimension of the picture under analysis smaller, but this operation would also decrease the precision of the tool.

Examples of the images extracted are reported in Figure 5. The EAST text detector has a limitation for the images with metal grills or meshes. In fact, it confounds tiny metal grills with words.

Crowdsourcing

Crowdsourcing is used in E2mC to both identify relevant posts and to validate and improve the precision of the geolocations generated by the CIME algorithm, with the help of volunteers. In the present paper we focus on crowdsourcing used for relevance tagging of the posts.

Crowdsourcing relevance tagging task

The selection of the pictures useful for creating disaster maps can be performed manually by a crowd of volunteers that evaluates the relevance of each image. Volunteers are asked to decide if a picture is relevant or not in a platform specifically created for the E2mC project, where the image and the related post are shown, as shown in Figure 6. The relevance value can be 1, when the image is considered useful and 0, when it is considered not useful. Each picture should be evaluated at least by 5 people to reduce the possibility of a wrong result by comparing the answers given by all the volunteers. In some cases of lack of time, the minimum number of results can be reduced, to process all the pictures in less time. The quality of the result given by the crowd, in this case, will decrease. Then the mean answer value is computed, and, if it is higher or equal to 0.8 (i.e. 4 out 5 volunteers considered the image relevant), the image is considered relevant. The time required to evaluate each picture is short, but the number of pictures that can be processed with this tool is not predictable, because it depends on the number of volunteers that participate in the project. From the case studies, each volunteer in the crowdsourcing can evaluate approximately 500 images per hour, however the effects of fatigue should also be considered, as discussed in (Purohit et al. 2018).

EVALUATION

In the present section we evaluate the illustrated filters, considering their effects on the goal of providing the rapid mapping operators with an informative set of images. Each filter can be considered a binary classifier that, depending on the value of a threshold, marks the images as 'negative' when they are classified as pictures to be eliminated because they are not relevant, or 'positive' if they should be kept, because they are useful for mapping purposes.

The result of the classification produced by each filter is then compared with the manual classification of relevance of each picture evaluated by two experts, to create confusion matrices that present the numbers of True Positive, True Negative, False Positive, False Negative. Thus, the filtering performance is evaluated in term of *precision*, i.e., the fraction of relevant posts among the retrieved posts, *recall*, i.e., the fraction of relevant posts that have been

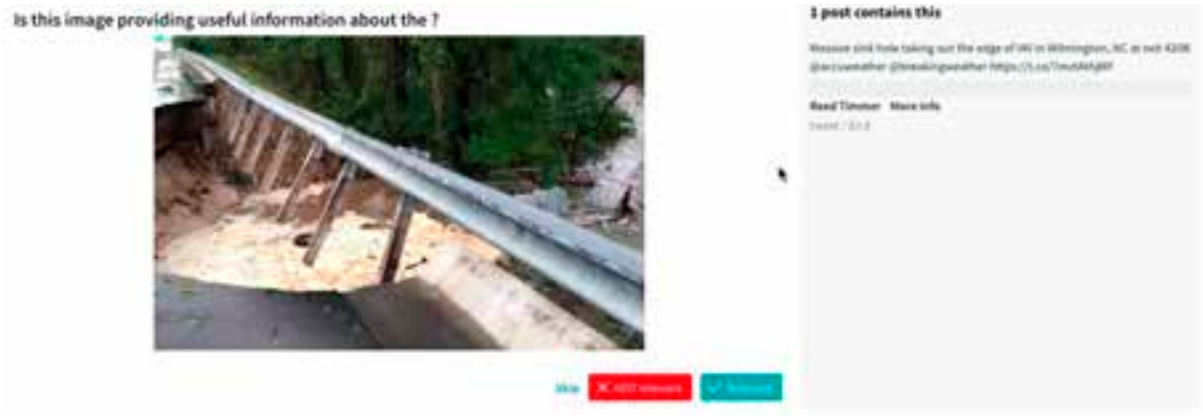

Figure 6. Crowd4EMS tool developed in E2mC

Table 1. Initial datasets expert evaluations

	UK flood		*Hurricane Florence*	
	Number of images	*Percentage*	*Number of images*	*Percentage*
Relevant	176	51.31%	38	12.62%
Not relevant	167	48.69%	263	87.38%

retrieved over the total amount of relevant posts, and the *percentages of relevant and non-relevant images discarded*. For each filter the mean percentage of reduction of the initial dataset and the precision gain after filtering is reported. The filters have been tested on two case studies, the floods that affected the Southern region of United Kingdom in 2014 and the Hurricane Florence in September 2018, which hit the North America East Coast. The first case study dataset is composed by 343 images that have been manually annotated, while for the second dataset a subset of 301 images has been used for the following analysis, manually annotating the relevance of the images by experts. For the analyses presented in this paper, random samples of the posts where manually annotated by two experts with the relevance of the images to the rapid mapping activity, as a basis for comparing the results obtained applying the different types of filters. Post selection is detailed in the following. The initial precision of the two datasets as evaluated by the experts and considered as ground truth are reported in Table 1.

Evaluation of redundant information reduction filters

Duplicates

Duplicated pictures have been detected in the two datasets with the *phash* technique. The algorithm shows limitations for both rotated and cropped images, which are not recognized as near duplicates. Duplicates are spread both between the relevant and non relevant images. The reduced precision of the final dataset in the set of hurricane Florence is due to the high percentage of duplicates between the relevant images. Table 2 summarizes the results.

Table 2. Duplicate filter results

	UK flood	*Hurricane Florence*
Precision	55.74%	9.70%
Recall	77%	60.53%
Relevant eliminated	23.00%	39.47%
Non relevant eliminated	35.33%	18.63%
Dataset reduction	28.86%	21.26%
Precision gain	+8.63%	-23.13%

Table 4. Colour filters results

	Filter white		*Filter black*		*Filter peaks*	
	UK flood	*Florence*	*UK flood*	*Florence*	*UK flood*	*Florence*
Precision	52.69%	13.38%	51.03%	12.50%	56.11%	14.81%
Recall	100%	100%	99.42%	97.37%	96.59%	84.21%
Relevant eliminated	0%	0%	0.58%	2.63%	3.41%	15.79%
Non relevant eliminated	5.39%	6.46%	1.78%	1.52%	20.36%	30.04%
Dataset reduction	2.62%	5.65%	1.17%	1.66%	11.66%	28.24%
Precision gain	+3.88%	+6.02%	+0.60%	-0.95%	+9.35%	+1.41%

Retweets

The filtering of the retweets has been applied on both datasets.The differences in the results on the two cases are due to the fact that hurricane Florence's dataset is characterized by a high percentage of retweeted posts, while for UK flood case they are less (see Table 3).

Table 3. Retweet filter results

	UK flood	*Hurricane Florence*
Precision	52.10%	6.86%
Recall	91%	32%
Relevant eliminated	9%	68%
Non relevant eliminated	11.38%	38.02%
Dataset reduction	9.91%	41.86%
Precision gain	+1.54%	-45.64%

Evaluation of image filters

Colour based

In Table 4 the results of the application of the three filters on the colour are reported. The low percentage of relevant images eliminated shows that they can be applied without discarding too many relevant pictures. The slightly higher value of relevant images eliminated by the "peak" filter on the hurricane Florence dataset is related to the scarce amount of relevant images in this dataset. Applying them one after the other can improve the quality of the final datasets and can eliminate high percentages of the non relevant images.

Images from previous events

The filter has been used on both the datasets. For the UK flood case study the algorithm was able to detect only one image considered relevant coming from a previous event. 11 not relevant images have been found that where posted from 2007 to the end of January 2014. In the case of hurricane Florence, no relevant images and 15 non-relevant images coming from previous events have been found.
This technique is not perfect and paid APIs would be needed to further improve it. Thus a faster and easier way to detect images from past events would be to create a database that collects the hashes of the viral and more repeated images of each event, to compare them with the new-coming images. This database could be implemented and tested when more case studies will be available.

Faces detection

The filter has been used on both the dataset showing that is useful to eliminate a part of the unwanted images without eliminating relevant ones (see Table 5.

Text detection

The filter has been applied to both the case studies' dataset to see how many images are discarded and if the images selected to be kept can be useful. In Table 6 the percentage of relevant and non-relevant images containing words

Table 5. Faces filter results

	UK flood	*Hurricane Florence*
Precision	53.66%	13.52%
Recall	100%	100%
Relevant eliminated	0%	0%
Non relevant eliminated	8.98%	7.60%
Dataset reduction	4.37%	9.63%
Precision gain	+4.58%	+7.13%

Table 6. Texts filter results

	UK flood	*Hurricane Florence*
Relevant with text	17.20%	9.97%
Non relevant With text	30.32%	62.46%
Reduction of the set of relevant	49.57%	21.05%

with respect to the whole dataset is reported to show that non-relevant images contain words more frequently than the relevant ones. Thus the last line of the table shows how much the set of relevant images would be reduced with this tool.

Evaluation of crowdsourcing activities

Crowdsourcing relevance tagging task

A crowd of volunteers has been used to evaluate both data sets. In the case of the Southern England floods dataset, the images have been evaluated five times by different members of the crowd, while in the case of hurricane Florence only once. This provokes a higher number of relevant images eliminated and lower number of non-relevant images discarded in this second case study (see Table 7). Anyhow the crowd filtering gives recall and precision values higher than any other filter for both case studies. The limitation of the tool is related to the human factor, since not always a sufficient number of persons are available to do these tasks when a disaster strikes.

Table 7. Crowdsourcing filter results

	UK flood	*Hurricane Florence*
Precision	96.93%	39.24%
Recall	88.76%	81.58%
Relevant eliminated	11.24%	18.42%
Non relevant eliminated	97.01%	81.75%
Dataset reduction	52.48%	73.75%
Precision gain	+88.91%	+210.94%

Evaluation of filters combinations

Some initial experimentation has been performed on the combination of the tools, in particular considering the configuration currently chosen in the E2mC project, which at the moment is adopting retweet filtering and crowdsourcing. The results have been evaluated using as ground truth an expert evaluation of the posts.

From the first experimental analyses, we gathered the following evidence:

- with the current E2mC configuration, retweets are filtered out and then a relevance task is performed by the crowd. We obtain an 80.11% recall and 2.99% False Positives in the UK flood case study and 69.92% recall and 9.09% False Positive rates in the case of Hurricane Florence, maintaining recall at a good level and with a limited number of false positives while reducing the manual effort. i.e., the resources needed to produce the images which are sent to the mapping operators for rapid mapping.

- Some filters appear to be highly correlated. For instance, black vs white filters are not correlated, while retweets and duplicates are highly correlated. Further investigation is needed to assess filter correlation in general.

THREATS TO VALIDITY

While for some of the filters, e.g., the colour filters, the recall and precision values have comparable values in the two case studies, in other cases they appear to be largely different, such as for instance for retweets. In future work, an estimation of these values could be derived case by case during different events, to study how to adapt thresholds and to estimate precision and recall accordingly.

Another aspect to be considered when combining different filters is that some of them are strongly correlated (e.g., the duplicates and the retweets filters). In those cases, the probability of joint occurrences for the two filters has also to be estimated, in order to be able to estimate the final precision and recall values for the combined filters. Ongoing work is analyzing the relationships between filters, to assess which ones are independent (such as for instance, it appears to be true in the pair white filter - duplicate filter) and the correlations patterns in correlated filters.

Finally, thresholds need to be fixed for each filter. They have been empirically estimated on the available case studies, therefore the values can not be considered valid for every kind of disaster. Also in this case further work is needed to be able to dynamically evaluate the appropriate thresholds for each filter given the type of emergency event and the number of available posts to be processed in real time.

CONCLUDING REMARKS

In this paper we have proposed a general approach for reducing the amount of irrelevant information extracted from social media during an emergency to support the activities of rapid mapping operators. As the quantity of posts can be variable in different events, due to their severity, their location, the availability of infrastructures after the event and the diffusion of social media, the aim is to support the choice of the best combination of filters depending on the characteristics of the on-going event. In future work, as the focus of this work has been on image analysis, we plan to analyze more in detail other types of filters, including also filters based on automatic image classification, being developed within the E2mC project. An adaptive tool combination and selection is also going to be studied in the future, to facilitate the rapid mapping operators activities, avoiding manual selection of filtering criteria, and aiming to generate a good combination of precision and recall in image selection. The aim of this work is to create an adaptive platform to select dynamically the filters and their thresholds, with an adaptive approach suited to the specific ongoing event.

ACKNOWLEDGMENT

This work has been funded by the European Commission H2020 project E^{2}mC "Evolution of Emergency Copernicus services". Grant Agreement No. 730082. This work expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this work.

REFERENCES

Alam, F., Ofli, F., and Imran, M. (2018). "Processing Social Media Images by Combining Human and Machine Computing during Crises". In: *International Journal of Human–Computer Interaction* 34.4, pp. 311–327.

Chen, T., Lu, D., Kan, M.-Y., and Cui, P. (2013). "Understanding and classifying image tweets". In: *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, pp. 781–784.

Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K. (2015). "Social media as an information source for rapid flood inundation mapping". In: *Natural Hazards and Earth System Sciences* 15.12, pp. 2725–2738.

Francalanci, C., Pernici, B., and Scalia, G. (2017). "Exploratory Spatio-Temporal Queries in Evolving Information". In: *Mobility Analytics for Spatio-Temporal and Social Data - First International Workshop, MATES 2017, Munich, Germany, September 1, 2017, Revised Selected Papers*, pp. 138–156.

Goldstein, J. and Rotich, J. (2008). "Digitally networked technology in Kenya's 2007–2008 post-election crisis". In: *Berkman Center Research Publication* 9, pp. 1–10.

Gupta, A., Lamba, H., Kumaraguru, P., and Joshi, A. (2013). "Faking Sandy: characterizing and identifying fake images on Twitter during hurricane sandy". In: *Proceedings of the 22nd international conference on World Wide Web*. ACM, pp. 729–736.

Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J. L., Achte, T. V., Zeug, G., Mondardini, M. R. (, Grandoni, D., et al. (2017). "E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time". In: *Sensors* 17.12, 2766, pp. 1–32.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. (2015). "Processing social media messages in mass emergency: A survey". In: *ACM Computing Surveys (CSUR)* 47.4, p. 67.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., and Blanford, J. (2011). "Senseplace2: Geotwitter analytics support for situational awareness". In: *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*. IEEE, pp. 181–190.

Murthy, D., Gross, A., and McGarry, M. (2016). "Visual Social Media and Big Data. Interpreting Instagram Images Posted on Twitter". In: *Digital Culture & Society* 2.2, pp. 113–134.

Nozza, D., Ristagno, F., Palmonari, M., Fersini, E., Manchanda, P., and Messina, E. (2017). "TWINE: A real-time system for TWeet analysis via INformation Extraction". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Software Demonstrations*, pp. 25–28.

Peters, R. and Albuquerque, J. P. de (2015). "Investigating images as indicators for relevant social media messages in disaster management". In: *12th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Krystiansand, Norway, May 24-27, 2015.*

Pezanowski, S., MacEachren, A. M., Savelyev, A., and Robinson, A. C. (2018). "SensePlace3: a geovisual framework to analyze place–time–attribute information in social media". In: *Cartography and Geographic Information Science* 45.5, pp. 420–437.

Purohit, H., Castillo, C., Imran, M., and Pandey, R. (2018). "Ranking of Social Media Alerts with Workload Bounds in Emergency Operation Centers". In: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, pp. 206–213.