Integration of Text and Image Analysis for Flood Event Image Recognition

Min Jing*[1], Bryan W. Scotney[2]
and Sonya A. Coleman[1]
[1]School of Computing and Intelligent Systems
[2]School of Computing and Information Engineering
Ulster University, United Kingdom
{m.jing;sa.coleman;bw.scotney}@ulster.ac.uk

Martin T. McGinnity
School of Science and Technology
Nottingham Trent University,United Kingdom
martin.mcginnity@ntu.ac.uk

Xiubo Zhang, Stephen Kelly
Khurshid Ahmad
School of Computer Science and Statistics
Trinity College Dublin, Ireland
kellys25@tcd.ie; {xizhang;khurshid.ahmad}@scss.tcd.ie

Antje Schlaf, Sabine Gründer-Fahrer and Gerhard Heyer
Department of Computer Science
University of Leipzig, Germany
{antje.schlaf;heyer}@informatik.uni-leipzig.de;
gruender@uni-leipzig.de

***Abstract*—Flood event monitoring plays an important role for emergency management. With the fast growth of social media, a large number of images and videos are uploaded and searched on the internet during disasters, which can be used as "sensors" for improving efficiency of emergency management. This work proposes a novel framework in which the rich information available from social media is incorporated with image analysis to enhance image retrieval for disaster management. The text associated with images of flooding events was used to extract prominent words associated with flooding. The image features are represented by a histogram of visual words obtained using the Bag-of-Words (BoW) model. The text and image analysis are integrated at the feature level, in which the text features are conjoined directly with image features. The proposed approach was evaluated based on two flood event corpuses obtained from the US Federal Emergency Management Agency media library and public Facebook pages and groups related to flood and flood aid (in German). The experimental results demonstrate the improved performance of image recognition after incorporating the text features, which suggests the potential to enhance the efficiency of emergency management.**

***Index Terms*—flood event image recognition; social media analysis; multimodal data fusion; emergency management.**

I. Introduction

The use of social media in disaster and crisis management is increasing rapidly within the EU. In recent research conducted in an EU-FP7 Project *Security Systems for Language and Image Analysis (Slandail)* [10], the end-user partners, An Garda Siochana (Irish Police), Police Service of Northern Ireland, Protezione Civile Veneto, and Bundeskommando in Leipzig Germany, have reported use of social media together with legacy media in natural disasters focusing on flooding events in Dublin, Belfast, Venice and Leipzig respectively. Existing web search platforms, such as Bing, Google and Yahoo, are based on searching contextual information, i.e., tags, time or location. Although text-based search is fast and convenient, the search results can be mismatched, of low relevance, or duplicated due to noise [13]. Techniques developed for visual content analysis are valuable for improving search quality and recognition capabilities of current emergency management systems. A recent study [5] has shown that whilst the current focus in disaster management system is on text analytics, visual content made available through social media will initially leverage text analytics and in the longer term image analytics will have a profound positive impact on disaster management.

Social media comprise contextual information such as tags, comments, geo-locations and metadata arising from the capture device, which are valuable for web-based applications. Content-based analysis such as image or video analysis can be used to enhance visual content filtering, selection, and interpretation, with the potential to improve the efficiency of an emergency management system. The use of texts found collateral to an image captions, titles of texts that comprise an image, or references to an image presented in a paper, have been used together with image features for categorizing images and for annotating images with keywords [2]. Machine learning and neural network systems have been used to train systems to automatically annotate images with keywords found in collateral texts [15]. Attention has been focused on fusing textual and visual aspects in various applications. For example [4] introduced a Content-aware Ranking model, in which textual and visual information are simultaneously leveraged in the rank learning process. The visual information is modelled into a regularization term and an efficient cutting plane algorithm is used to learn the model. In an application for web videos [13] proposed a framework that combines the contextual information and content analysis to achieve real-time near-duplicate video elimination. The video time duration, number of views and thumbnail images are used in an initial step to identify the near-duplicate web videos. The content analysis is then based on colour and local points to provide further validation of duplication. Attempts are also made to improve the tagging quality associated with images and videos. In [6] a

978-1-5090-3409-3/16/$31.00 ©2016 European Union

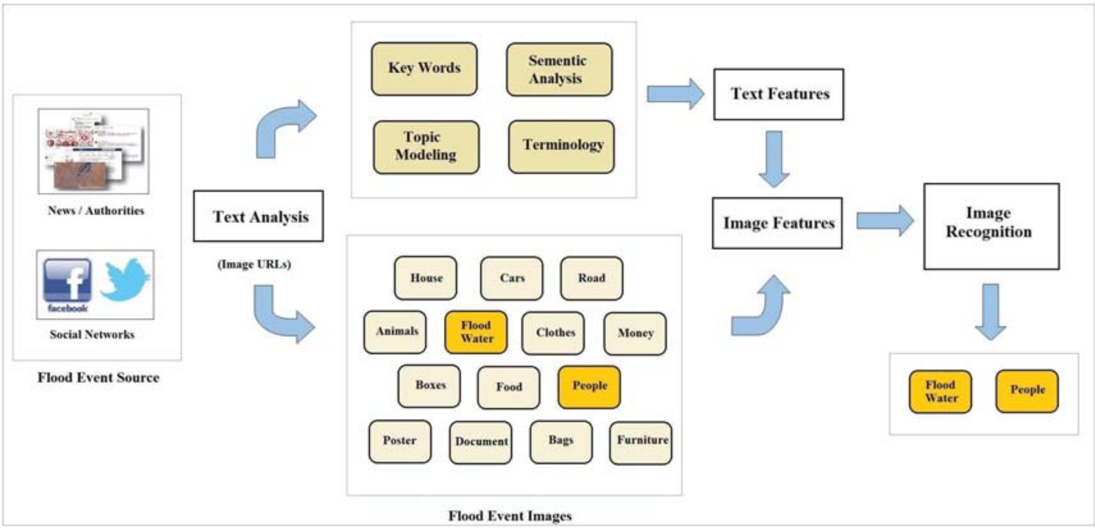

Fig. 1: Flood image recognition system including social media resources and integration of text and image analysis.

framework is proposed for social photo tagging by taking both user preference and geo-location information into account. The user preference is obtained from the user tagging history. The visual content of photos is represented by a bag-of-visual-words based on local image features.

The contextual information such as geo-tag is also important for the applications such as natural disaster management, in which the location of disaster and emergency support resource can be vital for the end users. In [7] landslides are detected based on integrating data from physical sensors (seismometers for earthquakes and weather satellites for rainstorms) and social media sources (Twitter and YouTube). The noise is filtered by using keywords, geo-tags and URLs. The disaster-related social media can be displayed using a geographic map. Many existing emergency management platforms directly share or display the visual content provided by simple text searches [7] [9], in which the social media images are used only for information sharing without incorporation of image analysis. In this work we aim to develop a framework to maximize the utility of rich information available from social media by fusion of the text and image analysis.

The rest of paper is structured as follows. The framework for the flood event image recognition system that integrates text analysis is described in Section II. The details of fusing image and text analysis at feature level are also explained. In Section III, an evaluation of the recognition performance based on the flood event corpus from the US FEMA web site and Facebook (in German) is provided, followed by discussion of the results and conclusions in Section IV.

II. Method

A. The Proposed Framework

A block diagram of the proposed flood event image recognition framework is presented in Figure 1, which includes the web image resources, together with integration of text and image analysis. Firstly, text analysis is performed on the flood event corpus that is obtained from a range of resources such as news feeds, government agency web sites and social networking sites. The corpus may include information on flood event location, time, related articles and posts, plus the images' titles, descriptions and URLs; thus each image is associated with corresponding text. The image URLs are used to extract the flood event images, which may contain flood water, people, roads, cars, and other entities. The image features are extracted and the feature representation is obtained based on the Bag-of-Words (BoW) model [8] as explained in Section II.B. The text-based features are developed (as in Section II.C) before being conjoined with the image features. The recognition system is based on the new features by combining text and image features to identify the target event images, such as images containing flood water and people. The output from the image recognition process can be saved in a common data format (such as XML Metadata Interchange) to facilitate further information exchange and interoperability between the image and text analysis systems, which is helpful when building an efficient emergency management platform.

B. Image Recognition Model

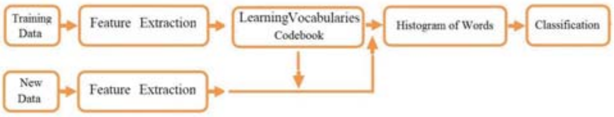

Fig. 2: The recognition system based on the BoW model.

The image recognition process is based on the BoW model [8] as shown in Figure 2. For each image the "Speeded-up Robust Features" (SURF) features [3] are extracted first. These local features are then mapped to a codebook created by the k-means clustering method. The feature presentation used for classification is obtained by calculating the histogram of the visual words for each image. As the BoW model does not use spatial relationships between the local features, learning

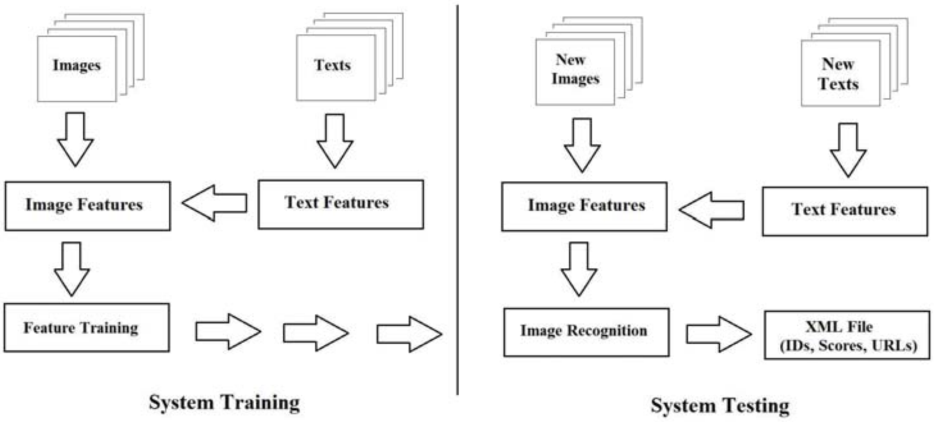

Fig. 3: Flood image recognition system via integration of text and image analysis at the feature level.

is computationally efficient. It should be noted that, for the image recognition system, the "word" refers to the "visual word", which is represented by a set of feature "centres" resulting from the clustering method. Classification is based on a Support Vector Machine (SVM). The output can be saved in a text format that may include the image IDs, recognition scores and identified class, which can be used easily for further text and image analysis integration.

C. Integration of Image and Text Features

Integration of image and text analysis can be done in different ways. A common approach is to apply the available text such as tags or keywords for pre-selection of images or videos [7], [13]. For the text analysis, some approaches have been proposed in which information from different domain is fused at the learning function level. For example [11] proposed a novel classification model called Constrained Weight Space SVM (CW-SVM). In CW-SVM the domain knowledge is represented by ranked labelled features, which are incorporated as the constrained weight by directly encoding expert feature knowledge through the definition of weight constraint sets. More recent work also show interests in multiple kernel learning (MKL) [12] in which different features are learned by the separate kernel functions within a classification system.

In this work, we propose a novel approach in which the text and image analyses are integrated at the feature level as shown in Fig. 3. As explained in the image recognition model (Fig. 2), the image feature representation is based on the BoW model, by which each image can be represented as a single vector whose length is equal to the number of visual words. The text features are obtained based on the occurrence of the keywords or predefined terms from the text associated with each image, such as by counting the presence of the keywords linked to each image. Hence a new feature can be formed by directly linking each image feature with a text feature. The extension of the text features can be developed by applying an advanced text analysis system, such as use of the disaster terminology developed in the Slandail project [10], which will be explored in the future work.

III. Experimental Results

The flood event corpuses were collected from two sources, the US FEMA media library and public Facebook pages and groups (in German) related to flood and flood aid which represent the resources of a government agency and a social networking site respectively. Each source has images with different levels of quality in terms of image size and resolution. The focus is to distinguish the flood water or person images from the background images. The images were extracted from the web sites by use of a web scraping tool. The images were selected and categorized manually into three groups: flood water, person, and background. The background images contain neither flood water nor people. Images of people may contain single or multiple persons. Attention has been given to ensure that the flood water images do not contain people and vice versa. In the BoW model, the number of visual words was 500 (as evaluated in [5]) and the recognition performance was evaluated based on mean Average Precision (mAP) obtained from 5-fold classification.

A. Facebook Flood Data

1) Data Details: As one of the most popular social networking sites, Facebook contains a large number of images related to flood events. In our experiments the flood event corpus was collected from public Facebook pages and groups related to Hochwasser (*flood*) and Fluthilfe (*flood aid*) posted during and after major floods in the Leipzig (Germany) area in May/June 2013. Note that the two words chosen by three German native speakers (co-authors from Germany) and are used as terms in specialist literature. The image URLs were obtained by identifying and searching German public Facebook accounts (public sites or public groups), account names containing two German words: Hochwasser (literally high water, and can be used to denote flood/flooding, high tide) and Fluthilfe (literally flood aid). Hochwasser is used as a term in hydrology, hydraulics, water engineering and geology, Fluthilfe is a term used in seeking or giving aid to flood victims. From these accounts, the public messages or posts with the type "photo" having a "link" and a "picture" (since both contain URL) were selected and their URLs were saved. The corpus includes 9,087 Facebook posts with type "photo", and approximately 5,000 Facebook images were extracted by the web scrapper. A total of 2,000 images were selected for this experiment, which include 1,000 images for each of the flood water and background groups. To maintain the connection between the images and corresponding text, each image was named using its unique Facebook post ID.

2) Flood Event Terms and General Language Words: There are 14 flood water related words in German used in searching, some of these words are used in German as specialist terms. The details of each term in German and the corresponding translation to English are shown in Table I, which is a glossary of prominent German words associated with the images. Some of the words are used as specialist terms in German (*flut, hochwasser, sandsack, pegel, pegelstand, wasserstand* and *deich*) whilst others are words of everyday use in German (*aktuell, gehfar, wasser,meter and strasse*). The occurrence of each of the selected terms or words (Table I) associated with each image was identified in three possible locations in the page: text, caption and description. The text includes the content presented in the Facebook posts and comments. The caption and description belong to the part of metadata which may not be publicly visible (depending on the page setup). The presence or absence of a term was indicated by "true" or "false", respectively. As seen from Table I, only flood water related terms were available, and so the evaluation for Facebook data was based only on flood water recognition.

TABLE I: Flood Related Terms and General Language Words in German and English

Germany	English
hochwasser	flood caused by water
wasser	water
sandsack	sandbag
sandsäcke	sandbags
pegel	water level
deich	dike
pegelstand	water level
meter	meter
flut	flood
straße	road, street
aktuell	current, present
wasserstand	water level
hochwasserschutz	flood protection
gefahr	danger

3) Terms vs Locations: For the Facebook data we first examined the occurrence of each term in three locations: text, caption or description. Since the focus was to identify flood water images from background images, we selected six flood water related terms. The six selected terms which include general language words are: hochwasser (flood), wasser (water), pegel (level or gauge), pegelstand or wasserstand (water level), flut (flood). We used 740 flood water images and for each image the occurrence of term (indicated by "True") in each location was counted. The average of occurrence in each location over 740 images was calculated. The results are shown in Fig. 4, which suggests that the words appeared most frequently in the text compared with image captions and descriptions. As the overall occurrence in each case is small, in the remaining experiments for Facebook data we combined the occurrence in all three locations to build the text based features.

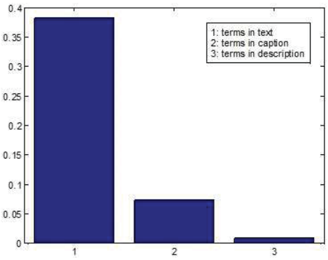

Fig. 4: Average occurrence of the water-related terms in three locations: text, description and caption.

4) Terms vs Groups: In our experiments we compared the occurrence of each term within the groups of flood water and background images. For each group we selected 615 images, 14 terms were used, and the average frequency of occurrence for each term across all three locations was calculated. The comparative results of average occurrences of all 14 terms within the two groups are presented in Fig. 5. It can be seen that some tokens, such as wasser (water), pegel (water level), pegelstand (water level) and wasserstand (water level), show are more frequent in flood water images than in background images. Two terms, hochwasser (flood) and flut (flood), appear less frequently in the flood water images than in background images. This may be due to the fact that words are used in everyday language are likely to be present in the text associated with both flood water and background images. Other words that are frequent in flood water images, include strae (street) and aktuell (current/present), which suggests that they may help to enhance the retrieval of images associated with flood events. For each image the occurrence of each term was used to build a text based feature resulting in a vector with a length equal to the number of terms. The text features were then linked to the image features for image recognition in the next stage of the experiment.

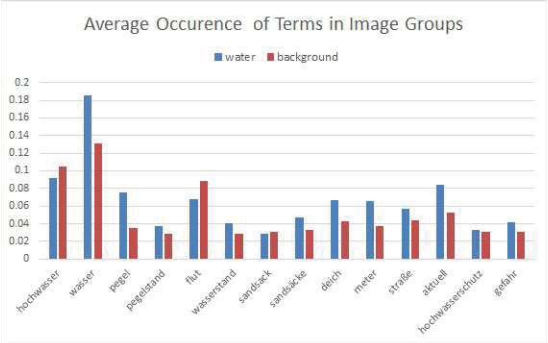

Fig. 5: Average occurrence of 14 terms in groups of flood water and background images.

5) Flood Image Recognition: For evaluation we compared the recognition performance with and without integration of text features. For each image the text based feature has dimension of 1×14, and the image feature based on histogram of visual words has a dimension of 1×500 (because the number of visual words was set 500 as evaluated in [5]). A new feature built by directly combining the text and image features gives a feature dimension of 514. We used 1,000 images from each group, 5-fold cross-validation was carried out, and mAP calculated. The results based on image features and combined features are presented in Fig. 6 and show that the recognition performance is improved by integrating the text and image features. For further evaluation we removed two terms, hochwasser (flood caused by water) and flut (flood), which show higher occurrences in background images than in flood water images, and the results based on the remaining 12 terms are shown in Fig. 7. The results show that using 14 terms is slightly better than using 12 terms. In both cases the results demonstrate an improved performance by fusing of text and image features.

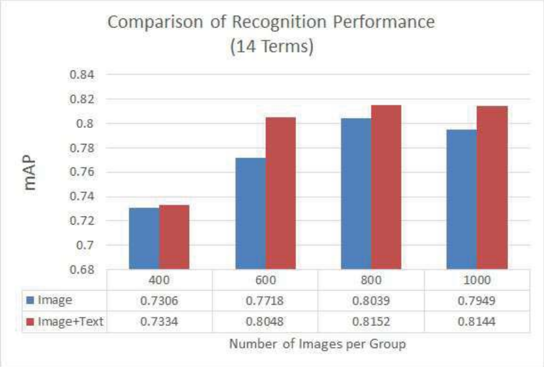

Fig. 6: Comparison of performance for recognition of flood water with and without integration of text features (based on 14 terms).

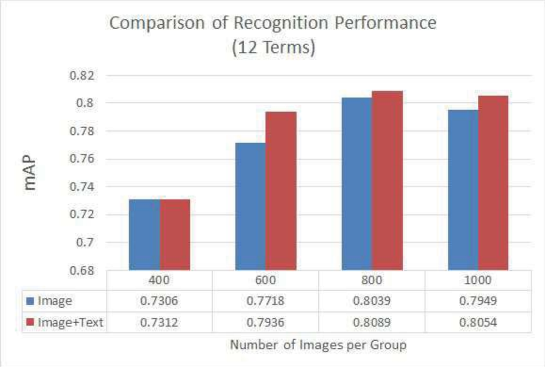

Fig. 7: Comparison of performance for recognition of flood water with and without integration of text features (based on 12 terms).

B. FEMA Flood Data

1) Data Details: As an emergency management authority, the FEMA web site provides images with high image resolution. The original FEMA images were collected from the FEMA media library based on text-based searching for the disaster type "flooding". A total of 1,800 FEMA images were selected for the experiments, including 600 images for each of the three groups: flood water, people, and background, respectively.

2) FEMA Keywords: To link images with text, each FEMA image was named using a unique ID associated with its URL. The web page content related to the image was analysed and the keywords were extracted using the CiCui System [14]. The occurrence of each keyword was measured by its term frequency inverse document frequency *(tfidf)* and *weirdness* [1]. The term frequency indicates the number of times the specific word occurred in the text. The document frequency is the number of documents in the corpus containing this word. A high *tfidf* indicates that a word is used frequently within a small subset of documents, suggesting the importance of the word. The *weirdness* is the ratio between a word's relative frequency in a domain-specific corpus and that in the general language. The relative frequency of a word in a corpus is the number of times the word occurred in that corpus divided by the total number of words in the corpus. A word with a high *weirdness* score indicates the importance of the concept for the domain. In total 17 FEMA keywords were selected by setting a threshold for their *weirdness* score. Based on the score from high to low, the candidate terms are: flood, flooding, sandbag, levee, survivor, resident, neighbourhood, disaster, floodwater, volunteer, outreach, homeowner, mitigation, storm, assistance, center and tornado. Note that some of the candidate terms identified were used to build the text features, which were then integrated with the image features.

3) Keywords vs Groups: The average occurrence of all keywords within three image groups is shown in Fig. 8. It can be seen that for the flood water images, the words "flood" and "flood water" appear more frequently than those in the background images. For images containing a person, "survivor" and "volunteer" have a higher frequency of occurrence than in the background images. For each image the occurrence of each keyword was used to form a text based feature, yielding a vector of length 17. Since there is a limited number of words related to persons, we did not build separate text features for flood water and person, but instead we used all keywords to build the text feature and examined the performance of recognition.

4) FEMA Flood Image Recognition: For each image the text based feature vector has a dimension of 1×17, and the image feature vector has a dimension of 1×500 (as explained above the number of visual words was set as 500). A new feature vector built after combining the text and image features has a dimension of 517. We used 600 images from each group, 5-fold cross-validation was conducted, and mAP calculated. Comparison of classification performance for flood water and person images based on image features only and on combined text and image features is presented in Fig. 9 and Fig. 10, respectively. It can be seen that for both cases the performance is improved after integration of text and image features. The results are also accordance with those obtained using Facebook

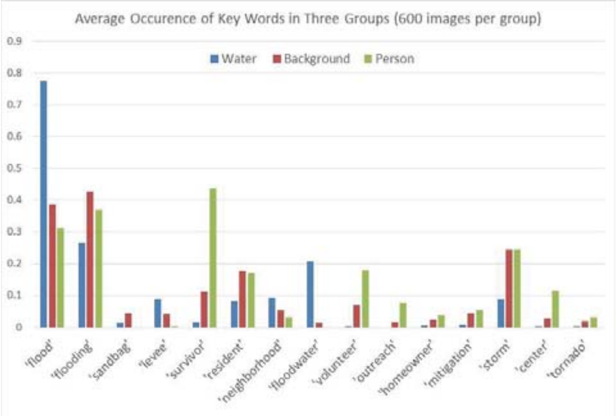

Fig. 8: Average occurrence of FEMA keywords in groups of flood water and background images.

data.

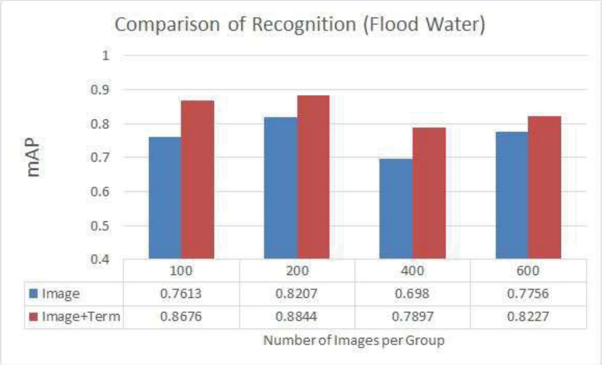

Fig. 9: Comparison of recognition of flood water images before and after integration of text-based features.

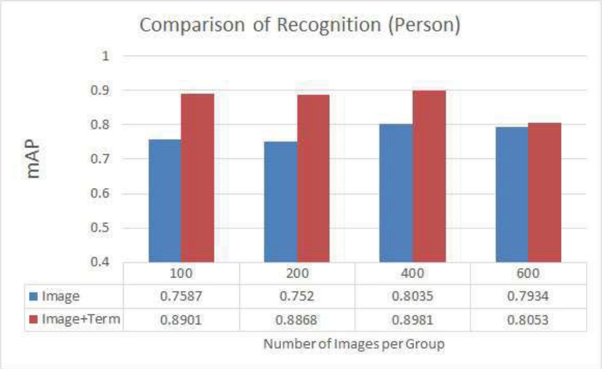

Fig. 10: Comparison of recognition of person images before and after integration of text-based feature.

IV. Conclusion

We have presented a novel framework for integration of text analysis with flood event image retrieval. Linguistic evidence was studied and flood related words and candidate terms were extracted based on text analysis, which were then used to build text-based features before linking with image features. The proposed approach was evaluated using the flood event corpus collected from the US FEMA web site and a public Facebook site from Germany. The results demonstrate the improved performance for specific flood related image retrieval when text analysis is combined with image analysis, suggesting potential for improving the efficiency of disaster management systems. Future work will explore the integration of text-based and image-based features on learning at the function level and possible extension of the set of text features based on the Slandail disaster terminology.

Acknowledgment

The research leading to these results has received funding from the European community's Seventh Framework Programme under grant agreement No. 607691, SLANDAIL (Security System for Language and Image Analysis).

References

[1] K. Ahmad and M. A. Rogers, "Corpus Linguistics and Terminology Extraction," In (Eds.) Sue Ellen Wright and Gerhard Budin, Handbook of Terminology Management (Volume 2), Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 725-760, 2001.

[2] K. Ahmad, M. Tariq, B. Vrusias, and C. Handy, "Corpus-Based Thesaurus Construction for Image Retrieval in Specialist Domains," In (Ed). Fabrizio Sebastiani. Proc 25th European Conf on Inf. Retrieval Research (ECIR-03) LNCS-2633, Heidelberg:Springer Verlag, pp. 502-510, 2003.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features", In Proc. ECCV, vol. 1, pp. 404-417, 2006.

[4] B. Geng, Y. Yang, C. Xu and X. S. Hua,"Content-aware Ranking for visual search Computer Vision and Pattern Recognition," In Proc. CVPR, pp. 3400-3407, 2010.

[5] M. Jing, B. W. Scotney and S. A. Coleman et. al, "Flood Event Image Recognition via Social Media Image and Text Analysis," IARIA conference COGNITIVE, 2016.

[6] J. Liu, Z. Li, J. Tang, Y. Jiang and H. Lu, "Personalized Geo-Specific Tag Recommendation for Photos on Social Websites," IEEE TRANSACTIONS ON MULTIMEDIA, vol. 16, no. 3. pp. 588-600, 2014.

[7] A. Musaev, D. Wang, and C. Pu, "LITMUS: Landslide Detection by Integrating Multiple Sources," In Proc. ISCRAM2014 (Information Systems for Crisis Response and Management), pp. 677-686, 2014.

[8] J. C. Niebles, H, Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," In Proc. BMVC, vol. 3, pp. 1249-1258, 2006.

[9] D. Pohl, A. Bouchachia, and H. Hellwagner, "Supporting Crisis Management via Sub-event Detection in Social Networks," IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), pp. 373-378, 2012.

[10] FP7 Project Slandail web site: www.slandail.eu.

[11] K. Small, B. C. Wallace, C. E. Brodley and T. A. Trikalinos "The Constrained Weight Space SVM: Learning with Ranked Features", In Proc. International Conference on Machine Learning, 2011.

[12] Y. Shynkevich1, T. M. McGinnity, S. Coleman and A. Belatreche,"Predicting Stock Price Movements Based on Different Categories of News Articles," In Proc. IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (IEEE CIFEr), 2015.

[13] X. Wu, C. W. Ngo, A. Hauptmann, and H. K. Tan, "Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context," IEEE Trans. Multimedia, vol. 11(2), pp. 196-207, 2009.

[14] X. Zhang, K. Ahmad, "Ontology and Terminology of Disaster Management", DIMPLE: Disaster Management and Principled Large-scale information Extraction Workshop Programme, 2014.

[15] C. Zheng, A. Long, Y. Volkov, A. Davies and K. Ahmad,"A Cross-Modal System for Cell Migration Image Annotation and Retrieval", 20th IJCNN:International Joint Conference on Neural Networks, pp. 1738-1743, 2007.