

## Flood Event Image Recognition via Social Media Image and Text Analysis

Min Jing<sup>\*1</sup>, Bryan W. Scotney<sup>2</sup> and Sonya A. Coleman<sup>1</sup>

<sup>1</sup>School of Computing and Intelligent Systems

<sup>2</sup>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.mcginity@ntu.ac.uk

Stephen Kelly, Xiubo Zhang

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**—The emergence of social media has led to a new era of information communication, in which vast amounts of information are available that is potentially valuable for emergency management. This supplements and enhances the data available through government bodies, emergency response agencies, and broadcasters. Techniques developed for visual content analysis can be useful tools to improve current emergency management systems. We present a new flood event scene recognition system based on social media visual content and text analysis. The concept of ontology is introduced that enables the text and image analysis to be linked at an atomic or hierarchal level. We accelerate web image analysis by using a new framework that incorporates a novel “Squirrel” (square spiral) Image Processing addressing scheme with the state-of-art “Speeded-up Robust Features”. The focus of recognition was to identify the water or person images from the background images. Image URLs were obtained based on text analysis using English and German languages. We demonstrate the efficiency of the new image features and accuracy of recognition of flood water and persons within images, and hence the potential to enhance emergency management systems. The system for the atomic level recognition was evaluated using flood event related image data available from the US Federal Emergency Management Agency media library and public German Facebook pages and groups related to flood and flood aid. This evaluation was performed for and on behalf of an EU-FP7 Project *Security Systems for Language and Image Analysis* (Slandail), a system for managing disasters specifically with the help of digital media including social and legacy media. The system is intended to be incorporated by the project technology partners CID GmbH and DataPiano SA.

**Keywords**—flood event recognition; fast image processing; 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 and will catch up with similar use of social media in the USA. The end-user partners in the Slandail Project (An Garda Síochána the Irish Police, Police Service of Northern Ireland, Protezione Civile Veneto, and Bundeskommando Leipzig, Germany) have reported use of social media together with legacy media in natural disasters focusing on flooding events in Belfast, Dublin, Leipzig and Venice. The specification of the end-user partners is being used to develop the Slandail system and will be made publicly available in 2017 [15]. Our research has shown that whilst the current focus in disaster management system is on text

analytics, still and moving images made available through social media will initially leverage text analytics, in the longer term image analytics will have a profound positive impact on disaster management. The advantages of rapid information sharing between the victims and the disaster managers, facilitated by social media, is offset to some extent by the fear of incorrect or misleading information being spread through social media. For most existing web search platforms, such as Bing, Google and Yahoo, searches are based on contextual information, i.e., tags, time or location. Text-based search is fast and convenient, though search results can be mismatched, of low relevance, or duplicated due to noise [16]. There are off-line techniques for identifying fake images have been proposed [5] and some online (real-time) techniques for “debunking” fake images on social media reported in [8]. Techniques developed for visual content analysis are valuable for improving search quality and recognition capabilities of current emergency management systems. In this work, we focus on scene recognition to enhance the information available within emergency management systems, with particular emphasis on flood event recognition.

Although image analytics have been applied widely in many areas, social media image content analysis has not been exploited fully within emergency management systems. For example during the flood in Germany in 2013, many Facebook pages and groups were created (mainly by private persons) and used in order to exchange information and coordinate the help of volunteers, in which images posted on social media may be used as “sensors” for detecting or monitoring possible flooding events. Many existing emergency management platforms directly share or display the visual content provided by simple text search [13] [11], in which the social media images are used only for information sharing without incorporation of image analysis. Social media are equipped with rich contextual information such as tags, comments, geo-locations and capture device metadata, which are valuable for web-based applications. Not only are the images and videos described by meta-data fields (e.g., title, descriptions, or tags), but content analysis can be used to enhance visual content filtering, selection, and interpretation, with the potential to improve the efficiency of an emergency management system. This work aims to develop a novel and efficient emergency event recognition framework, in which text and image analysis

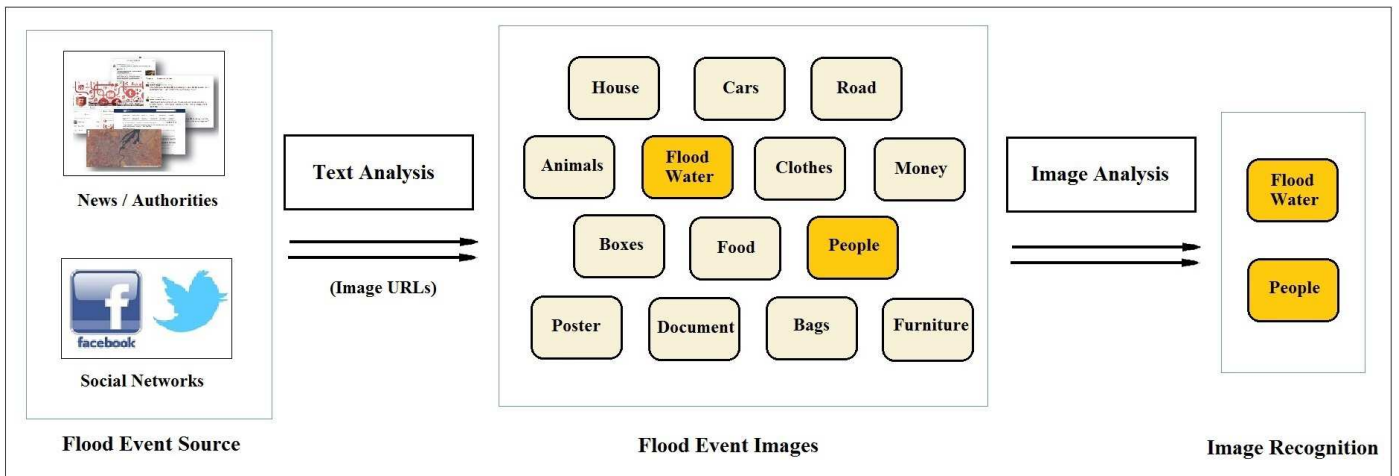


Figure 1. Flood event recognition system including image resources together with text and image analysis.

are deployed to identify flood event images from news feeds and popular social network web sites.

One key requirement for the wide-spread adaptation of image analytics is the ability of disaster management systems to react in real time: Here our contribution through the proposed “Squirrel” (square-spiral) Image Processing (SIP) framework will be significant. Different approaches have been proposed for fast image processing. Some studies have attempted to reduce the image size, such as in a study for mobile image search [10], the image is compressed first then learned by a 3D model developed for landmark recognition. The rich contextual information available from the web can be used to filter the visual content and therefore reduce processing time, such as using the features from YouTube thumbnail images for near-duplicate video elimination [16]. Some studies have also considered biologically motivated feature extraction [14] for fast feature extraction on hexagonal pixel based images. In recent work, we proposed a novel SIP framework [6] which develops a spiral addressing scheme for standard square pixel-based images. A SIP-based convolution technique is developed based on simulating the eye tremor phenomenon of the human visual system [14] [2], to accelerate the computation required for feature extraction. In this work, we incorporate the SIP addressing scheme within the Speeded-up Robust Features (SURF) [1] algorithm to improve the efficiency of web image recognition.

The development of the flood event image recognition algorithm and the overall recognition system that combines image and text analysis are described in Section II. The framework for fast image processing, essential for real time image and video analysis, is also outlined and an approach to link SURF with the SIP framework is presented. An evaluation of the recognition system performance and feature detection is also provided in Section III, followed by discussion of the results and conclusions in Section IV.

## II. METHODS

### A. Proposed Framework

A block diagram of the proposed flood event image recognition framework is presented in Figure 1. The system includes the web image resources, together with text and

image analysis. Firstly, text analysis is performed and the flood event related corpus is obtained from a range of resources such as news feeds, government agency web sites and social networking sites. The corpus includes information on event location, time, article titles, descriptions, and URLs for images. The URLs are used to extract the flood event images that may contain flood water, people, roads, cars, and other entities. The images collected are used in training the recognition system, which includes image feature extraction, learning of visual words and construction of feature representation based on the Bag-of-Words (BoW) model [12]. The details of feature extraction method is given in Section II.E. After training, the system is able to identify the target event images, such as images containing flood water and people. Output from the recognition process is saved in a text file using a common data format (such as XML Metadata Interchange) to facilitate information exchange and interoperability between the image and text analysis systems.

### B. Concept of Ontology

To facilitate the link between image and text analysis, we introduce the concept of an ontology as the basis of event recognition for selected applications within the scope of natural disasters. In general, an ontology can be defined as the formal specification of a vocabulary of concepts and the relationships between them. In the context of computer and information science, ontology defines a set of primitives, such as classes, attributes or properties and relationships between the class members [4]. The concept of ontology has been applied increasingly in automated recognition tasks such as recognition of objects [3], characters [4], and emotion [17]. In this work, we introduce the concept of ontology to image-based flood event recognition. An example of a simple ontology, representing the flood event image and the relationships between related event images, is shown in Figure 2. This example illustrates that a flood event image may contain both flood water and people. (In the following part of this paper, “water” refers to “flood water”.) This work was focused on single event recognition (atomic level). A more complex ontology structure can be constructed based on hierarchies and inheritance rules, which will be linked to text analysis in future development.

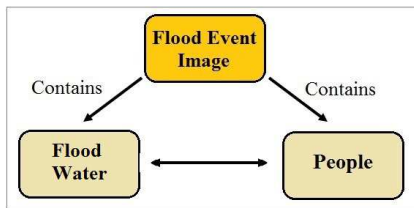


Figure 2. An example of a simple ontology representing flood event images.

C. Recognition Model

The image recognition is based on the BoW model [12]. In BoW the local features are first mapped to a codebook created by a clustering method such as k-means and then represented by a histogram of the visual words that is used for classification. As the BoW model does not rely on the spatial information of local features, learning is efficient (though loss of spatial information due to the histogram representation may affect accuracy). A system based on the BoW model is shown in Figure 3. Note that, for the image recognition system, the

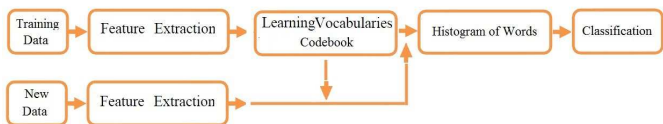


Figure 3. The recognition system based on the BoW model.

“word” refers to the “visual word”, which is represented by a set of feature centres resulting from the clustering method. The classification is based on a Support Vector Machine (SVM). The output can be saved in a text format for further text and image analysis integration. To accelerate recognition performance, in the feature extraction stage we have introduced a new SIP framework to link with SURF. The details of SIP addressing and the development of the feature are explained in sub-sections D and E.

D. “Squiral” (Square-Spiral) Image Processing (SIP)

Fast image processing is a key element in achieving real-time image and video analysis. Real-time data processing is a challenging task, particularly when handling large-scale image and video data from social media. Recently we have developed a novel SIP framework that introduces a spiral addressing scheme for standard square pixel-based images [6]. The SIP-based approach enables the image pixel values to be stored in a 1D vector, facilitating fast access and accelerating the execution of subsequent image processing algorithms by mimicking aspects of the eye tremor phenomenon in the human visual system. Layer-1 of the SIP addressing scheme comprises 9 pixels in a spiral pattern as shown at the centre of Figure 4. Subsequent layers of the SIP addressing scheme are built recursively: a complete layer-2 SIP addressing scheme is shown in Figure 4. The SIP structure facilitates the use of base 9 numbering to address each pixel within the image. For example, the pixels in layer-1 are labelled from 0 to 8, indexed in a clockwise direction. The base 9 indexing continues into each layer, e.g., layer-2 starts from 10, 11, 12, ... and finishes at 88. Subsequent layers are structured recursively. The converted SIP image is stored in a one-dimensional vector according to the

spiral addresses. Conversion of standard two-dimensional pixel indices to the 1D SIP addressing scheme can be achieved easily using an existing lattice with a Cartesian coordinate system. Furthermore, the approach can be used for efficient convolution of existing image processing operators designed for standard rectangular pixel-based images, and so the approach does not require any new operators to be developed.

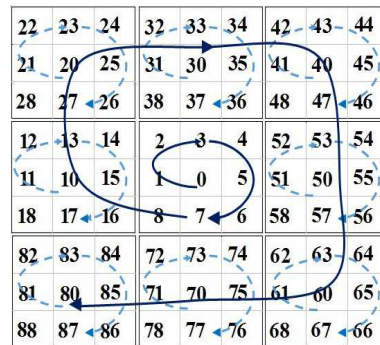


Figure 4. The spiral addressing scheme for layer-2 SIP.

E. SIP-based Features (SIPF)

We incorporate the SIP addressing scheme with the image feature SURF [1] to improve the efficiency of web image analysis. We refer to the resulting feature as SIP-based Features (SIPF). SURF has been used widely in image analysis and has shown advantages over SIFT [9]. It has been demonstrated in [6] [7] that SIP-based convolution produces exactly the same results as standard convolution, and hence in our current implementation we use the interest points detected by SURF but rearrange the SURF features according to the SIP addressing scheme. As shown in Figure 5 (a), the SURF feature is constructed based on a square region centred on the detected SURF interest point. The region is divided into smaller 4 × 4 sub-regions, and within each sub-region the wavelet responses are computed. The responses include the sums of  $dx$ ,  $|dx|$ ,  $dy$ , and  $|dy|$ , computed relative to the orientation of the grid, where  $dx$  and  $dy$  are the Haar wavelet responses in the horizontal and vertical direction respectively;  $|dx|$  and  $|dy|$  are the sums of the absolute values of the responses, respectively. Hence each sub-region has a four-dimensional descriptor vector  $[dx, dy, |dx|, |dy|]$ . Concatenating these for all 4 × 4 sub-regions results in a SURF descriptor vector of length 64. To

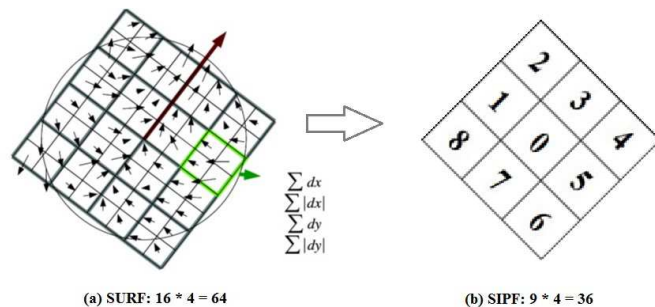


Figure 5. (a) SURF feature construction [1]; (b) SIPF feature based on layer-1 SIP addressing scheme.

construct the equivalent with the SIP framework, we apply the layer-1 SIP addressing scheme to rearrange the SURF feature obtained from each interest point. In order to match the layer-1 SIP structure, the 4 × 4 sub-regions are resized to 3 × 3

sub-regions using bicubic interpolation method (in which the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood), and then the corresponding response values are rearranged according to the layer-1 SIP addressing scheme as shown in Figure 5 (b). This results in a descriptor of length  $9 \times 4 = 36$ . Note that the current implementation does not involve full SIP image conversion and SIP convolution, but it yields the same outcome and may be considered as an initial stage from which future development of a full SIP image feature detection algorithm will be completed. Because the SIPF feature vector length is shorter than that for SURF (36 values rather than 64), we expect additional efficiency gains for computation as well as the benefits of the 1D addressing system. In our computational experiments the performance in terms of recognition and efficiency based on SURF and SIFP are evaluated and compared.

### III. EXPERIMENTAL RESULTS

#### A. Data

The flood event-related image data were collected from two sources: the US Federal Emergency Management Agency (FEMA) media library and public German Facebook pages and groups related to flood and flood aid. These choices represent the resources of a government agency and a social networking site respectively. A collection of images from official sources such as FEMA was compiled to act as a benchmark for comparison with potentially lower quality images published on social media platforms. As an emergency management authority, FEMA's web site provides high quality images with high image resolution. The original FEMA images (typically of maximum dimension 2000-4000 pixels) were collected from the FEMA media library using a web scraper based on text-based searching for the disaster type "flooding". A total of 6000 FEMA images were collected, in which 1200 images were selected and used in the experiments, including 400 images for each of three groups: flood water, people, and background, respectively. The background images contain neither flood water nor people. Images of people may contain single or multiple persons. The permission of publicly displaying the FEMA images were obtained from FEMA news desk. Ideally the flood water image does not contain person and vice versa, however this does not affect single event recognition which is the focus of this work.

As one of the most popular social networking sites, Facebook contains a large number of images related to flood events. Flood related images were collected from Facebook by using a keyword search, and the images collected have a maximum height of 720 pixels. The German Facebook image URLs were obtained by identifying and searching German public Facebook accounts (public sites or public groups), account names containing the word "Hochwasser" (flood) or "Fluthilfe" (flood aid or help in case of flood). From these accounts, the public messages or posts with the type "photo" having a "link" and a "picture" (since both contain an URL) were selected and their URLs were saved. A total of 5000 Facebook images were collected from German Facebook in which 1200 images were selected, which include 400 containing flood water, 400 containing a person (or persons), and 400 background images.

#### B. Comparison of Image Features

Comparison of performance based on image features SURF and SIPF was conducted using the original FEMA image data,

which include 50 flood water images and 150 background images. A two-fold cross validation was performed on the different image sizes, such as 0.2, 0.4, 0.6 and 0.8 of the original size. The number of words in the BoW model was 500. The system performance evaluation is based on the average precision (AP), which can be obtained based on the area under the precision-recall curve.

As high resolution images are expensive in terms of memory storage and processing time, we compared the computational efficiency using recognition run-time with different image scales using three feature extractors: SIFT, SURF, and SIPF, which have feature dimensions of 128, 64, and 36, respectively. The run-time includes the time for feature point detection, feature extraction, calculating the feature histogram, and SVM classification. The run-time results for water image recognition are shown in Figure 6. It can be seen that the computation time increases with the image size. The SIFT detector (dimension 128) is more time-consuming than SURF and SIPF. Both SURF and SIPF are similar in run-times, but SIPF is slightly faster (when the time for SIP conversion is excluded). We also compared the recognition performance based on SURF and SIPF features using different image sizes. The mean of AP (mAP) values are shown in Figure 7 and SIPF has a better recognition rate than SURF using different image sizes. Since the primary aim of this work is to develop a framework for flood event recognition, the evaluation was based only on flood event related images.

#### C. Evaluation of Event Recognition

To test the performance of flood event recognition, we used FEMA images containing flood water and persons. The images without water or persons are used as background images. The original FEMA images are resized to the standard FEMA web version size (dimension 1024 x 680). Using web-sized images suits the reality of end-user needs, as images presented on the FEMA web site are already resized and compressed.

1) *Test of Parameter Settings:* The number of words in the BoW model can affect the system's efficiency, such as a smaller number of words may help to reduce the processing time. We investigated how different parameter settings may affect the recognition performance based on different number of words and the total number of training data. For each group 200 images were used for testing, 200 for training. Half training data contains water or person and another half are background images, i.e., 400 training data include 200 water or person and 200 background images. The results are shown in Figure 8 and Figure 9. It can be seen that for water images, using 500 words results in better performance than using 1000 words; for person recognition, using 1000 words results in better performance. In terms of training data, the overall performance improves as the number of data examples is increased.

2) *Comparison of FEMA and Facebook Image Data:* The performance based on FEMA and Facebook image data set was compared. For each data set 800 images were used (each class has 400 images plus 400 background images). The number of words used was 500, 5-fold cross validation was performed and the mAP calculated. The results are shown in Figure 10 and Figure 11. The performance using FEMA and Facebook images appears to be similar, with the recognition system performing well for both. Furthermore, in terms of feature

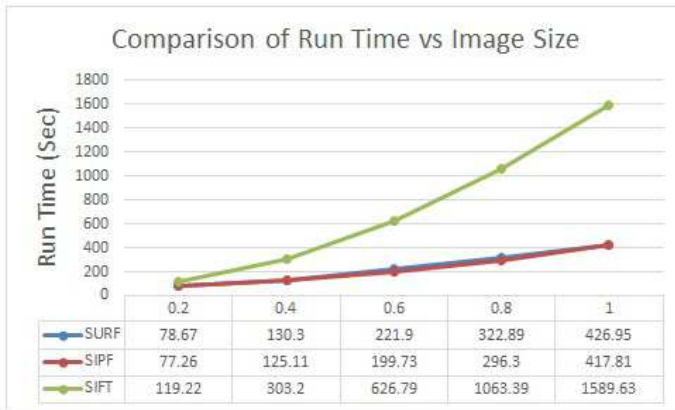


Figure 6. Comparison of run time using features SIFT, SURF and SIFP.

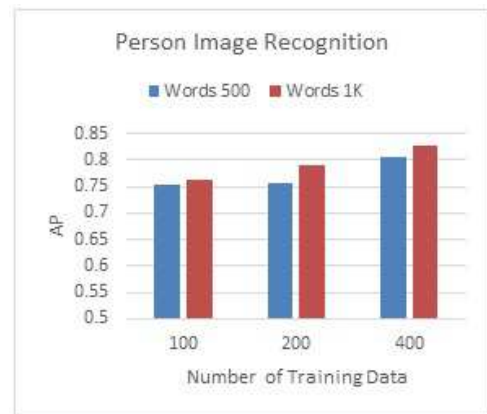


Figure 9. Performance using different number of words for person images.



Figure 7. Comparison of recognition rate based on SURF and SIFP.

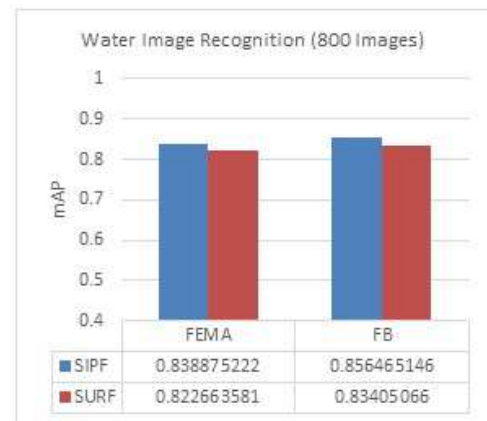


Figure 10. Comparison of performance based on water images from FEMA and Facebook (FB).

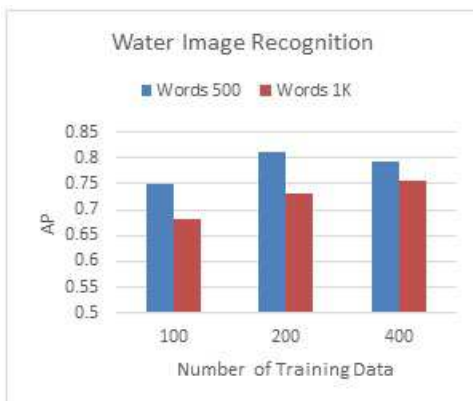


Figure 8. Performance using different number of words for water images.



Figure 11. Comparison of performance based on person images from FEMA and Facebook (FB).

performance, SIFP appears to be slightly better than SURF, as shown in both Figure 10 and Figure 11, supporting the use of the more compact representation of the SIP based features.

3) *Test of Event Recognition:* The atomic level recognition system is built based on a binary classification, which is designed to identify a single event, such as whether the image contains flood water. For a future development, a more complex recognition system will be built to incorporate multi-class classification. Examples of FEMA images recognised as containing water and as containing persons, respectively, are

shown in Figure 12 (a) and Figure 12 (b). The target images are identified and ranked by the recognition score provided by the SVM. For further integration of the image analysis with text analysis, the outputs of image recognition were saved in a text file, including the top N ranked images, scores and image IDs.



Figure 12. Examples of flood event image recognition: (a) water images (AP = 89.50%) and (b) person images (AP = 85.44%).

IV. CONCLUSION

In this work we propose a novel framework that introduces the SIP addressing scheme to facilitate fast web visual content analysis in the context of enabling linkage of visual content analysis and text analysis. The framework is developed with close linkage to text analysis, in which the images are obtained based on a corpus from text analysis. The outcomes of event recognition can be stored using a common data format to facilitate further system integration. The overall purpose is to enable more efficient information exchange in emergency management systems. Hence, an image-based event recognition system has been developed based specifically on flood events, in which images containing flood water and persons were used as examples of using concept of ontology. The system developed can be extended for a more complex ontology structure and higher level scenario recognition in future work.

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] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features”, In Proc. ECCV, 2006, vol. 1, pp. 404-417.  
 [2] S. A. Coleman, B. W. Scotney, and B. Gardiner, “A Biologically Inspired Approach for Fast Image Processing”, In IAPR Proc. Machine Vision Applications, 2013, pp. 129-132.  
 [3] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, and P. Gancarski, “Ontology-based Object Recognition for Remote Sensing Image Interpretation,” IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2007, pp. 472-479.  
 [4] A. Eutamene, H. Belhadef, and M. K. Kholadi, “New Process Ontology-Based Character Recognition. Metadata and Semantic Research Communications in Computer and Information Science,” 2011, vol. 240, pp. 137-144.  
 [5] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, “Tweetcred: Real-time credibility assessment of content on twitter,” In Social Informatics, Springer International Publishing, 2014, pp. 228-243.

[6] M. Jing, B. W. Scotney, S. A. Coleman, and T. M. McGinnity, “Biologically Inspired Spiral Image Processing for Square Images”, In Proc. IAPR MVA, 2015, pp. 102-105.  
 [7] M. Jing, B. W. Scotney, S. A. Coleman, and T. M. McGinnity, “Multiscale “Squirrel” (Square-Spiral) Image Processing,” In Proc. IMVIP, 2015, pp. 1-8.  
 [8] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, “Real-time Rumor Debunking on Twitter,” In Proc. ACM International on Conference on Information and Knowledge Management, 2015, pp. 1867-1870.  
 [9] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, 2004, vol. 60(2), pp. 91-110.  
 [10] W. Min, C. Xu, M. Xu, X. Xian, and B. K. Bao, “Mobile Landmark Search with 3D Models,” IEEE Transactions on Multimedia, 2014, 16(3), pp. 623-636.  
 [11] A. Musaev, D. Wang, and C. Pu, “LITMUS: Landslide Detection by Integrating Multiple Sources,” In Proc. ISCRAM2014 (Information Systems for Crisis Response and Management), 2014, pp. 677-686.  
 [12] J. C. Niebles, H. Wang, and L. Fei-Fei, “Unsupervised learning of human action categories using spatial-temporal words,” In Proc. BMVC, 2006, vol. 3, pp. 1249-1258.  
 [13] D. Pohl, A. Bouchachia, and H. Hellwagner, “Supporting Crisis Management via Sub-event Detection in Social Networks,” IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012, pp. 373-378.  
 [14] B. W. Scotney, S. A. Coleman, and B. Gardiner, “Biologically Motivated Feature Extraction Using the Spiral Architecture”, In Proc. IEEE ICIP, 2011, pp. 221-224.  
 [15] FP7 Project Slandail web site: www.slandail.eu.  
 [16] 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, 2009, vol. 11(2), pp. 196-207.  
 [17] X. Zhang, B. Hu, J. Chen, and P. Moore, “Ontology-based context modeling for emotion recognition in an intelligent web,” World Wide Web, 2013, vol.16, pp. 497-513.