Retrieving Social Flooding Images Based on Multimodal Information

Zhengyu Zhao, Martha Larson
Radboud University, Netherlands
z.zhao@cs.ru.nl,m.larson@cs.ru.nl

ABSTRACT

This paper presents the participation of the RU-DS team at the MediaEval 2017 Multimedia Satellite Task. We design a system for retrieving social images that show direct evidence of flooding events using a multimodal approach based on visual features from images and the corresponding metadata. Specifically, we implement preprocessing operations including image cropping and test-set pre-filtering based on image color complexity or textual metadata, as well as re-ranking for fusion. Tests on the YFCC100M-Dataset show that the fusion-based approach outperforms the methods based on only visual features or metadata.

1 INTRODUCTION

Recent advances in satellite imagery and popularity of social media are opening up a new interdisciplinary area for earth monitoring, especially on natural disasters. The objective of the MediaEval 2017 Multimedia Satellite Task is to enrich the satellite information with multimodal social media for a more comprehensive view of flooding events [2]. We participate in the Disaster Image Retrieval from Social Media subtask, which requires us to retrieve social images that show a direct evidence of flooding events. Previous work in [1, 4, 5] addresses a similar challenge by leveraging visual and textual content from Social Media to enrich remote-sensed events in satellite imagery. In this paper, we investigate the exploitation of visual features and textual metadata for image representation, as well as propose a fusion method based on test-set pre-filtering and list re-ranking.

2 PROPOSED APPROACH

Table 1 contains a description of the approaches used for our three runs, which involve three different parts: pre-processing, feature extraction and fusion strategy. For the first run, (`Visual`), we apply image cropping and test-set pre-filtering based on color complexity, and use the SVM classifier on visual features to rank the images in descending order by the output decision values. For the second run, (`Text`), we rank the images by searching for flood-related keywords in metadata without any preprocessing. Finally, for the third run, (`Fusion`), we develop a 3-step approach: first the Run 2 system for pre-filtering, then the Run 1 system for ranking, and finally the Run 2 system again for re-ranking.

2.1 Visual Features

We have investigated nine conventional visual descriptors provided by the task organizers on the dev-set using an SVM classifier, and found the CEDD feature, which incorporates color and texture information, achieved the best performance.

Table 1: Run Description

Run	Pre-processing	Features	Fusion
Run 1: `Visual`	Image cropping and pre-filtering	Visual	-
Run 2: `Text`	-	Text	-
Run 3: `Fusion`	Pre-filtering	Visual+text	Re-ranking

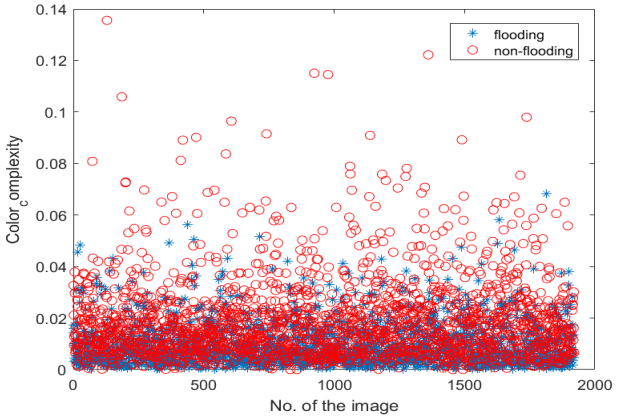

Figure 1: Color Complexity of 3960 Cropped Dev-set Images

The approach of our `Visual` run is based on the insight that the body of flood water parts of the image are more important than other parts for flood retrieval. Because the body of flood water is usually located in the lower part of a flooding image [3], we try to extract this part from each test image. Experiments on dev-set show that eliminating the top 60% of the image as well as 10% on each side could bring about an accuracy improvement of 4.5%. Moreover, using cropped images could save computation time of feature extraction and eliminate the interference from the sky region.

Another insight that we use is related to the observation that flood regions are visually homogeneous. We address this insight by computing color complexity of the cropped images. Color complexity here is defined by the equation: $Color\ complexity = \frac{N_h}{S}$, where N_h indicates the number of hues of an HSV image and S is the area of the image, i.e. the number of pixels. As shown in Fig. 1, the cropped non-flooding images tend to have higher color complexity than the flooding ones. We set an empirical threshold, $T = 0.05$ so as to remove a good number of non-flooding images, but few

MediaEval'17, 13-15 September 2017, Dublin, Ireland

Table 2: Keywords Distribution on the Dev-set Images

User_tags	Title	Description	Positives / all	Precision
1	1	0	421 / 478	88.07%
1	1	1	166 / 208	79.81%
1	0	0	780 / 1013	77.00%
1	0	1	174 / 256	67.97%
0	1	1	50 / 76	65.79%
0	1	0	152 / 266	57.14%
0	0	1	177 / 751	23.57%
0	0	0	0 / 2233	0
-	flooded	-	181 / 208	87.02%
1+water	-	-	398/ 467	85.22%

flooding ones. These removed images will be ranked in ascending order by color complexity in the end part of the final list.

2.2 Textual Metadata Features

We search for flood-related keywords including "flood(s)", "flooding" and "flooded" in the three main fields "User_tags", "Title" and "Description" of the accompanying metadata to rank the images. Table 2 shows the relationship between keyword occurrence and relevance reflected by the precision scores on the dev-set images, where 1 indicates that flood-related keywords are present in a specific field, and 0 means they are not. We use "x x x" in the first three columns to indicate eight general conditions and two special ones (one condition per row except for the header row). The latter two columns show the corresponding retrieval precision scores for each condition.

Overall, as shown, we find that keywords in the "User_tags" field are the most helpful, and keywords in the "Title" field are less reliable. Also, "Description" tends to give misleading information. Furthermore, because the ground truth defined images showing "unexpected high water levels in industrial, residential, commercial and agricultural areas" as positives [2], the conditions "1+water - -" and "- flooded -" (where the presence of water bodies are implied) are more likely to be positive.

In order to create the final result list for Run 2, we concatenate the sublists retrieved by each of above eight general conditions, in descending order by precision scores. Meanwhile, for each sublist, we will put the images that also meet the latter two conditions "1+water - -" or "- flooded -" as the top part.

2.3 Feature Fusion

In this section, we describe our fusion strategy based on pre-filtering and re-ranking using both visual and metadata information. First, we rank all the images that meet the conditions "0 0 0" and "0 0 1" using our metadata-based system to generate the sublist 2, which will be the end part of the final list because as shown in Table 2, these images are very unlikely to be positive. Then, the rest of the images are fed into our visual-based system and ranked in descending order by decision value to generate the sublist 1. Finally, we re-rank the images in sublist 1 whose decision values are non-positive using our metadata-based system again.

Table 3: Official Evaluation Results on the Test-set (Best results in bold)

Run	AP @ 480	mAP @ (50, 100, 250, 480)
Run 1	51.46	64.70
Run 2	63.70	75.74
Run 3	**73.16**	**85.43**

3 RESULTS AND DISCUSSION

Table 3 presents the official results for our three submitted runs on the test-set. We see that the third run achieves the best performance for both evaluation metrics. We can also observe the retrieval process benefits from fused visual and metadata information. Specifically, implementing test-set pre-filtering based on flood-related keywords in our metadata-based approach leads to considerable better performance that our visual-based approach based on color complexity. Further, the visual-based approach is verified to perform better than the metadata-based one in the conditions except for "0 0 0" and "0 0 1". The reason for this effect could be that some images mentioning flooding in the metadata are relevant to flooding, but do not visually depict any floodwater. Such images will be labeled as negatives in the ground truth.

4 CONCLUSION AND OUTLOOK

In this paper, we presented an approach for retrieving images showing evidence of flooding events based on visual and textual (metadata) information. Final results showed using both visual and textual features outperforms using either feature individually.

During the exploratory experiments that led to our Run 1 `Visual` approach, we tried to first divide the cropped image into blocks before the other steps, and then to compute the final score for an image based on the scores of each block. This approach did not achieve a better performance. It maybe because most blocks divided from a homogeneous region in non-flooding images are more likely to be regarded as a body of flood water without contribution from a global feature that contains information such as the white line in the road or the boats on the river.

In the future, we will try segmentation algorithms to extract the body of flood water more accurately and develop better visual descriptors to differentiate the body of flood water from other water bodies in non-flooding images in the large-scale dataset. We will also explore the word relations between user tags to avoid the mistaken decision when the flood depicted in the image did not consist of water. Finally, for the fusion strategy, methods of combining the feature vectors of different modalities will be explored.

ACKNOWLEDGMENTS

This research is partially supported by China Scholarship Council (201706250044).

REFERENCES

[1] Benjamin Bischke, Damian Borth, Christian Schulze, and Andreas Dengel. 2016. Contextual Enrichment of Remote-Sensed Events with Social Media Streams. In *ACM Multimedia Conference 2016*. ACM, 1077–1081.

[2] Benjamin Bischke, Patrick Helber, Christian Schulze, Srinivasan Venkat, Andreas Dengel, and Damian Borth. The Multimedia Satellite Task at MediaEval 2017: Emergence Response for Flooding Events. In *Proc. of the MediaEval 2017 Workshop* (Sept. 13-15, 2017). Dublin, Ireland.

[3] Paulo Vinicius Koerich Borges, Joceli Mayer, and Ebroul Izquierdo. 2008. A Probabilistic Model for Flood Detection in Video Sequences. In *IEEE International Conference on Image Processing 2008*. IEEE, 13–16.

[4] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, 851–860.

[5] Jie Yin, Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power. 2012. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems* 27, 6 (November 2012), 52–59.