Contextual Enrichment of Remote-Sensed Events with Social Media Streams

Benjamin Bischke Damian Borth Christian Schulze Andreas Dengel
German Research Center for Artificial Intelligence (DFKI)
D-67663 Kaiserslautern, Germany
{benjamin.bischke, damian.borth, christian.schulze, andreas.dengel}@dfki.de

ABSTRACT

The availability of satellite images for academic or commercial purpose is increasing rapidly due to efforts made by governmental agencies (NASA, ESA) to publish such data openly or commercial startups (PlanetLabs) to provide real-time satellite data. Beyond many commercial application, satellite data is helpful to create situation awareness in disaster recovery and emergency situations such as wildfires, earthquakes, or flooding. To fully utilize such data sources, we present a scalable system for the contextual enrichment of satellite images by crawling and analyzing multimedia content from social media. This information stream can provide vital information from the ground and help to complement remote sensing in situations. We use Twitter as main data source and analyze its textual, visual, temporal, geographical and social dimensions. Visualizations show different aspects of the event allowing high-level comprehension and provide deeper insights into the event as complemented by social media.

Keywords

Social Multimedia;Remote Sensing;Satellite Images

1. INTRODUCTION

We present a solution for the ACM Multimedia 2016 Grand Challenge on the task 'Sky and the Social Eye: Enriching Remote-Sensed Events'. The goal of this challenge is to augment events that can be detected in satellite images with multimedia content from social media streams.

Our system consists of (1) a data crawler, (2) an analysis component and (3) a browser with interactive visualizations. In order to contextualize an event with social media content, users specify a set of keywords and a date range. The crawler requests Twitter's Historical-Powertrack (HPT)API and retrieves all tweets with the corresponding keywords and date range. The system performs an automatic multimedia content analysis and provides visualizations showing the aggregated contextual facets of the event. We establish the link from satellite images to event summarization by considering the temporal and spatial dimension. The low temporal resolution of satellite images is augmented with a continuous timeseries of user activities in social media streams. Thereby, big changes in satellite images can be directly mapped to spikes in user activities and online discourse of the event. Extracted locations from social media are used for collecting satellite images and for linking them to user generated content such as images of the event.

Our solution addresses the evaluation criteria *Interestingness* and *quality of demonstration* by considering multimodality of multimedia content in combination with visualizations that emphasize the contextual facets of an event. The focus of our contextual augmentation is determined by textual, visual, temporal, geographical, social dimensions. Our goal is to tell the broader story of an event by providing a summarization with multiple views of the event.

Scalability and *Technical contribution* are addressed by presenting an architecture that is capable of analyzing tweets at large-scale. Our analysis component relies on a distributed task queue, which allows to scale horizontally, as tasks can be analyzed in parallel on multiple compute nodes. We further suggest a solution for Near-Duplicate Detecion of images in social media streams using features from Convolutional Neural Networks (CNN's).

The remainder of this paper is structured as follows. Section 2 covers related work and in Section 3 details of our system are presented. Section 4 describes the user interface and results while Section 5 concludes the paper.

2. RELATED WORK

Social Media Streams have emerged as a popular medium for collaborative information sharing thus representing valuable information sources for new applications. Researchers leveraged social media content for public opinion mining [18], [11], [6], for monitoring and prediction of infectious disease outbreaks [15], [16], for coordination of rescue efforts after disasters [7], [13], [10] as well as for analyzing the social reflection of the society [9]. Closest to our work is the investigation in the interaction of social media with events caused by natural disasters for situational awareness. The goal of such systems [24] [14] is to provide users a big picture of an event in a way that rescuer can asses the situation and act accordingly. [24] uses a data capturing component for accessing Twitter APIs and applies burst detection, online clustering for topic discovery and classification for impact assessment on these tweets. [14] monitors tweets and to detect a target event and find the center of the event loca-

MM '16, October 15-19, 2016, Amsterdam, Netherlands

DOI: http://dx.doi.org/10.1145/2964284.2984063

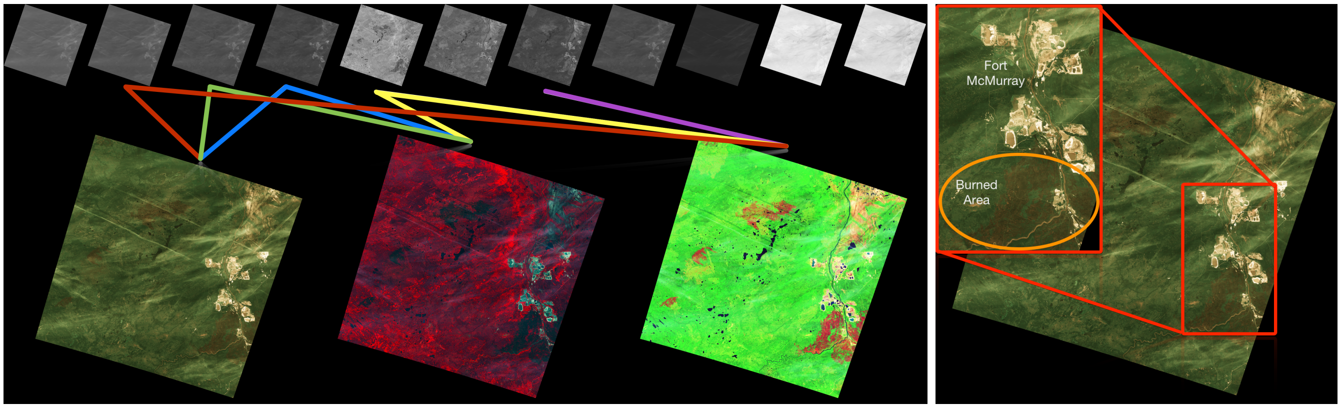

Figure 1: Left: This figure shows the eleven original bands of the Landsat 8 satellite and three generated images after combining bands. Left bottom shows an RGB-Image using Bands 4-3-2. Middle bottom: False-Color Image emphasizing the vegetation in red color using Bands 5-4-3. Right bottom: False-Color Image showing fire burn scars in red color using Bands 7-5-2. Right: This visualization shows an computed RGB-Satellite-Image of Fort McMurray. The brownish color illustrates burned forest scars after the Wildfire in May 2016.

tion. Tweets are classified based on keywords and passed to a probabilistic spatiotemporal model which estimates the location of target event with an particle filter. [1] considers only planned events (such as concerts, sports or political events) and suggests strategies to query and identify related content across social media sites.

While these systems put a strong focus on event detection using methods of burst detection, anomaly detection and unsupervised clustering using text or images, our solution can be understood as a targeted search approach for social media. Given a set of keywords for an event, our system crawls multimedia content and enriches it with multiple contextual aspects. Our motivation is to provide the big picture of an event going beyond a textual or visual event summarization utilizing independent signals in a multimodal fashion [23].

3. SATELLITE IMAGE PROCESSING

NASA's Landsat 8[1] is an American Earth observation satellite which senses the entire earth in 16 days capturing scenes with an approximate size of 170 km north-south by 183 km east-west. The Landsat 8 satellite scans the earth's surface with multiple remote sensors providing 11 bands of spectral images. To construct images from these spectral bands, the following steps are executed:

3.1 Satellite Image Collection

In a first step our system fetches the freely available Landsat 8 images from Amazon S3. In order to download images of a specific location, it is necessary to determine the correct scene's path and row of the satellite. We used the *Landsat8 Metadata API*[2] which provides the mapping from geolocation to scene-id. Our system collects image scenes in two ways: (1) by manually specifying the location and (2) by using extracted locations from social media. Potential positions of an event can be automatically identified by leveraging frequently mentioned location names in text messages and incorporating GPS-annotations of tweets.

[1]http://landsat.usgs.gov/landsat8.php
[2]https://github.com/developmentseed/landsat-api

3.2 Satellite Image Composition

Landsat 8 images consist of nine spectral bands (with different wavelength such as blue, green, red, near infrared(NIR) and short-wave infrared (SWIR)) as well as two thermal infrared bands. Insights of an scene can be extracted by combining different bands into one image. Within this work we computed an RBG color image and two False-Color Images. False-Color images incorporate not the usual Red, Green, Blue colors, but further information (such as NIR or SWIR) into the RGB-channels allowing to focus on different aspects of the scene. We computed two False-Color Images emphasizing the vegetation health and wildfire scars of the scene.

4. SOCIAL MEDIA PROCESSING

The proposed social media processing pipeline can be split into two major processing steps:

1. The social media data crawler gathers all tweets for an related event by requesting public and historical search API of Twitter. The downloaded tweets are bundled into packages and passed to the analysis component.
2. The social media analysis component is divided into multiple tasks, which are queued by the asynchronous task queue Celery. Celery is based on the distributed message passing system RabbitMQ and allows to distribute and analyze the tasks among multiple compute nodes. Analysis results are stored into a MongoDB.

4.1 Social Media Data-Crawler

We use Twitter as a primary data source for analyzing multimedia content. [11] reported that user activities such as online posts and comments are often triggered by specific events, making Twitter an important medium for information spreading. In addition to text messages, tweets contain rich metadata, such as timestamps, geolocations, hyper-links to webpages, news articles and images. This metadata is valuable for a multimodal analysis and for bootstrapping further multimedia content from social media.

Within the scope of this paper, our crawler gathered tweets by requesting Twitter's Historical PowerTrack-API. HPT allows to collect on historical tweets from the past, by specifying filter-rules such as keywords, date-ranges or geospatial searches. We use Twitter's Public-API for realtime events, as the returned tweets are limited to the last seven days.

4.2 Social Media Analysis

After downloading all tweets for one event the analysis is applied in four steps: (1) Metadata-Analysis, (2) Text-Analysis, (3) Image-Analysis and (4) Aggregation.

Metadata Analysis.

The goal of this processing step is to collect all metadata associated with the tweet including the resolution of URLs within the tweet and its linked media content (i.e. images, news). Since locally or globally discussed events are recognized by leveraging GPS-annotations of tweets this information provides a fundamental link to the geo-location for satellite imagery. Geo-locations are mapped to countries and binned into a projected hexagonal grid of the world. Our system creates a timeline by extracting tweet creation timestamps. For scalability reasons, we reduce the range of possible timestamps by means of discretization of the timestamps to the nearest four hours of every day. Influential users in social media applications are determined by multiple aspects, such as the Klout-Score. A Klout-Score represents the social online influence of users based on social media analysis. The score considers the user's network, the generated content and interactions of other users with that content. In addition, we rank users according to the retweet and favourite count of their published tweets. A high retweet or favorite count indicates that the tweet reached many users in the network, thus having some degree of influence on other users. The system identifies URLs in tweets, follows the links and downloads the corresponding resources. Unique hyper-links mentioned in multiple tweets are downloaded only once. Duplicated Web-Resources having different URLs are recognized in the following analysis steps.

Text Analysis.

The following steps are applied on the text messages of tweets: Twitter users have the option to tag their tweets with so-called hashtags. Hastags can be understood as a meta-label, describing an topic or aspect of the event with one word or few words. We extract the most common hashtags in all tweets as they represent a special collaborative form of topic summarization from users. We perform a Sentiment-Analysis on the twitter messages. The analysis is based on the popular linguistics based sentiment model SentiStrength [20], which is optimized on short social web texts. SentiStrength provides a positive and a negative sentiment score in range of -4 to +4. To decide whether an event is globally or more locally discussed, we apply in addition to geographical augmentation a language detection. Tweets from HPT are already partially annotated from Twitter with languages. On the remaining messages we apply Bloomberg's language tool NLP<GO>. Tweet-messages are ranked by retweet and favourite counts. These top tweets have been shared or liked by many users of the social network, thus being an important and influential representation of the topic.

Image Analysis.

Images liked within tweets are analyzed with respect to the following aspects: Since user generated content is often very noisy [22], we first apply a image filtering to remove

Figure 2: Most frequently shared images on Twitter for the event 'Wildfire Fort McMurray' in Canada. Top: Filtered images classified as synthetically generated. Bottom: Images classified as natural.

synthetic images such as logos, maps, cartoons which appear often but to not directly show evidence of the observed event. We first generated a dataset by crawling 5.716 synthetic images using web search and randomly sampling the same number of images from the YFCC100M dataset [21]. We used 70% of the dataset for training a linear Support Vector Machine with the representation of the fc6 layer of AlexNet [8]. Testing on the remaining images, we achieved an classification accuracy of 98,6%. Compared to the approach proposed in [12] we accomplish a 2% improvement.

Events with a global impact i.e. with high attention in news and social media have often a set of iconic images that stay in the head of the society. Our goal is to find the most popular ones that are strongly associated with the event in social media. A first approach for this is the selection of images with the most frequent hyper-links in the tweet corpus. However, we observed that popular images are often copied on different servers and thereby getting a new URL. Additional image manipulations such as cropping, scaling and different JPEQ-Quality levels applied on the images make a duplicate detection purely based on image pixels infeasible. We solve this problem by using a high-level representation of the image content and searching for every image for its near duplicates. Our detection algorithm extracts the representation of the fc6 feature layer of the 16 layered CNN VGG-Net from [17] for every image in filtered image set. To detect near duplicates the most similar images are retrieved using the cosine distance. Images above a empirically determined threshold are considered as near duplicates and are together with the searched image removed from image set. The approach is iteratively applied on the remaining images.

We tested features from multiple layers (fc6, fc7, pool5) of state-of-the-art CNN's (AlexNet[8], ResNet [5], VGG-Net[17], and GoogLeNet [19]) for our near duplicate detection approach. Our evaluation is based on the INRIA CopyDays dataset[4] which contains 157 original images and includes three transformations (1) 16% resizing of the image surface plus different JPEQ-Quality compressions (2) cropping of the image surface between 5% to 80% as well as (3) strong transformations such as print, scan and perspective effects. While for the former two transformations nine near duplicate images are generated, the latter one has only 223 images. For an overall evaluation, we therefore only used cropping and JPEQ-Compression transformations. By comparing Mean Average Precision of the top nine images, the fc6-feature of VGG-S-Net delivered the best result.

In order to perform visual sentiment analysis on the fil-

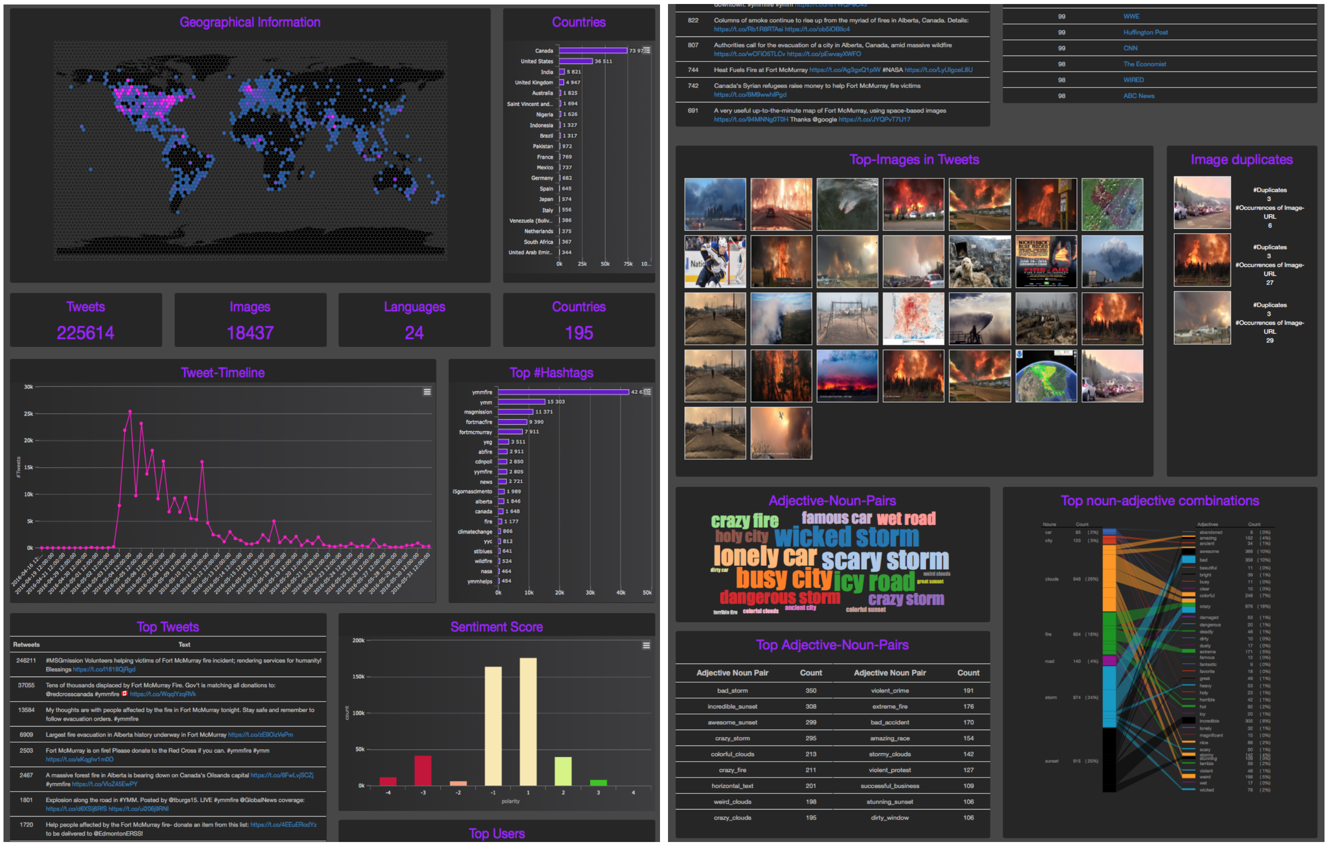

Figure 3: This visualizations shows two sections of our browser. Left: Geographical and temporal activities from social media as well as most influential tweets are visualized for the Wildfire in Fort McMurray are visualized. Left: Most popular images contained in tweets, together with the extracted Adjective-Noun Pairs and near duplicates are visualized. Bottom right shows all combinations of nouns and adjectives on all images

tered images from social media, we apply the concept classifier tool: DeepSentiBank [3]. DeepSentiBank is a fine-tuned Convolutional Neural Network (CNN), that is based on the Visual Sentiment Ontology (VSO) [2]. The VSO consists of 3244 Adjective-Noun-Pairs (ANPs), the classifier is trained on a subset of the VSO with 2089 ANPs. We apply DeepSentiBank on each image and select the top ten ANPs with the highest probability.

Temporal Aggregation.

We aggregate the analysis results from the previous steps for each day which is then aggregated over the entire period. Here, we use the top ranked results of each day as aggregation function for the entire observation period. This allows the system to perform analysis on subsets of days within the observation period.

User Interface.

We developed an online available browser[3] with a suite of visualizations for interacting and exploring the extracted information for an event. Our browser has a entry page which shows all analyzed events and links to the event-specific summary. Visualizations on that website provide an overview of the topic and the contextualized content from social media. Geo-tagged images are visualized on the collected satellite images. Fig. 3 illustrates the geographical, textual and visual extracted and enriched aspects of the example event of the Wildfire in the Fort McMurray, in May 2016.

5. CONCLUSIONS

We presented a scalable system for the contextual enrichment of satellite images by crawling and analyzing multimedia content from social media. We use Twitter as main data source for collecting data and bootstrapping further multimedia content such as images. The focus of our the social media analysis is determined by textual, visual, temporal, geographical and social dimensions. Visualizations show different aspects of the event, allow high-level comprehension and provide deeper insights into the contextualized event from a social media perspective.

6. ACKNOWLEDGMENTS

This work was partially funded by the BMBF project Multimedia Opinion Mining under (MOM: 01WI15002).

[3]http://mom.dfki.de/demos/deep_eye/

7. REFERENCES

[1] H. Becker, D. Iter, M. Naaman, and L. Gravano. Identifying content for planned events across social media sites. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 533–542, New York, NY, USA, 2012. ACM.

[2] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM, 2013.

[3] T. Chen, D. Borth, T. Darrell, and S. Chang. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *CoRR*, abs/1410.8586, 2014.

[4] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid. Evaluation of gist descriptors for web-scale image search. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 19. ACM, 2009.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[6] W. He, S. Zha, and L. Li. Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3):464–472, 2013.

[7] M. E. Keim and E. Noji. Emergent use of social media: a new age of opportunity for disaster resilience. *American journal of disaster medicine*, 6(1):47–54, 2010.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[9] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009.

[10] B. R. Lindsay. Social media and disasters: Current uses, future options, and policy considerations, 2011.

[11] D. Maynard, K. Bontcheva, and D. Rout. Challenges in developing opinion mining tools for social media. *Proceedings of the@ NLP can u tag# usergeneratedcontent*, pages 15–22, 2012.

[12] P. J. McParlane, A. J. McMinn, and J. M. Jose. Picture the scene...;: Visually summarising social media events. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1459–1468. ACM, 2014.

[13] L. Palen. Online social media in crisis events. *Educause Quarterly*, 31(3):76–78, 2008.

[14] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[15] J. Sarasohn-Kahn. The wisdom of patients: Health care meets online social media, 2008.

[16] C. W. Schmidt. Using social media to predict and track disease outbreaks. *Environmental health perspectives*, 120(1):A31, 2012.

[17] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] P. Sobkowicz, M. Kaschesky, and G. Bouchard. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4):470–479, 2012.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[20] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[21] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. Yfcc100m: The new data in multimedia research. *Commun. ACM*, 59(2):64–73, Jan. 2016.

[22] A. Ulges, D. Borth, and T. M. Breuel. Visual concept learning from weakly labeled web videos. In *Video Search and Mining*, pages 203–232. Springer, 2010.

[23] A. Ulges, C. Schulze, D. Borth, and A. Stahl. Pornography detection in video benefits (a lot) from a multi-modal approach. In *Proceedings of the 2012 ACM international workshop on Audio and multimedia methods for large-scale video analysis*, pages 21–26. ACM, 2012.

[24] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59, 2012.