Proceedings of the Joint Austrian Computer Vision and Robotics Workshop 2020 DOI: 10.3217/978-3-85125-752-6-32

How High is the Tide? Estimation of Flood Level from Social Media

Julia Strebl, Djordje Slijepcevic, Armin Kirchknopf,
Muntaha Sakeena, Markus Seidl, Matthias Zeppelzauer
St. Pölten University of Applied Sciences, Austria
{firstname.lastname}@fhstp.ac.at

Abstract. *The availability of social media data represents an opportunity to automatically detect and assess disasters to better guide emergency forces. We propose a method for flood level estimation from user-generated images to support assessing the severity of flooding events. Furthermore, we provide labeled data for water detection. Results on a public benchmark dataset are promising and motivate further research.*

1. Introduction

The visual estimation of flood levels is a novel task. In this paper we aim at detecting images with a certain water level, i.e. where the water is at least knee-high. Our work is based on preliminary work from the MediaEval 2019 Satellite Task [1]. Our contribution is twofold: we demonstrate the feasibility of visual flood level estimation by combining a supervised water detector with pose estimation and we provide novel image annotations for water detection.

Related work focuses on either visual, textual or multimodal flood level estimation from social media content [1]. Zaffaroni et al. [5], for example, combine multiple pre-trained networks for the estimation of flood level. Further approaches can be found in [4]. We aim at presenting a simple and efficient approach to provide a baseline for future comparison.

2. Methods

Input to our approach are social media images. We propose two approaches that build upon three components: (i) a supervised water detector that predicts whether a certain image or image region contains water, (ii) a pose estimator that detects people and their joints and (iii) a rule-based fusion module that combines the information from the water detector and the pose estimator to make a final decision. The first approach (see Figure 1A) aims at detecting water within the whole image and detecting at least one person with concealed lower body parts. The second approach (see Figure 1B) performs water detection locally around each detected human body. If at least for one body the model detects concealed lower extremities and water in the vicinity, the image is assigned to knee-high water.

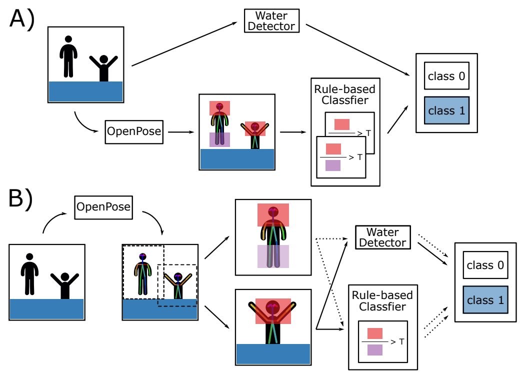

Figure 1: Global (A) and local (B) approach.

We employ ResNet50 (pre-trained on ImageNet) for water detection. Images are resized to the network's input size (227x227) while keeping the original aspect ratio. Horizontal flipping, brightness variations and non-uniform re-scaling of the images are applied for data augmentation. The top five layers are fine-tuned (6 epochs, batch size 256) before the whole network is trained using Adam optimizer (10 epochs, batch size 32, learning rate 10^{-4}). We employ OpenPose [3] to detect body joints from depicted human bodies. To filter out unreliable skeletons, we exclude those with a confidence score (C_U) - calculated from the two most robust upper body parts (head and chest) - below an empirically estimated threshold of 0.6. We calculate a mean confidence score (C_L) over the lower body parts (knees and feet). To determine whether the lower extremities of a skeleton are visible, we employ the following heuristic rule: $C_U / \max(C_L, 10^{-4}) > T$, with T being an empirically determined threshold of 1.5. Finally, positive predictions of the rule-based classifier and the water detector implies a positive detection of a person standing in knee-high water.

3. Datasets

Experiments are carried out on two datasets provided by the MediaEval Benchmark Multimedia Satellite Task 2018 (MMSat18) and 2019 (MMSat19) respectively [1, 2]. All available data is manually annotated (water/no water) and used to train the water detector. A total of 13.761 image annotations (5.395 water, 8.366 no water) along with corresponding image URLs (incl. download tool) as well as our ResNet50 model weights can be accessed publicly[1].

4. Results & Discussion

For experimental evaluations, we randomly split the MMSat19 data into training (80%) and validation (20%) sets preserving class priors. Testing is performed on the (non-public) test set of the MMSat19 benchmark. For the global approach (GA), we used the pipeline in Figure 1A. First, the water detector is trained only on the MMSat19 data (GA-1) and later on both datasets (GA-2). For the local approach (LA), we used the pipeline in Figure 1B with MMSat19 data. Finally, we apply majority voting to all three approaches.

Due to the imbalanced data, we used macro averaged F1-scores as performance measure. The experimental results surpass the random baseline of 0.5, which shows that our models are able to learn useful patterns. The results on the test set show only minor differences between the four approaches. The overall performance is similar on the validation and test sets, which indicates a good generalization ability. The classification accuracy of the water detector is quite high with 88% (not shown in Table 1). The main source of failure are false detections of the pose tracker due to occlusions by foreground objects and reflections in the water (see Figure 2). Potential improvements identified include the use of several pose estimators trained on content from different environments, e.g., rural and urban areas. Additionally, pixel-wise classification (segmentation) of water and human bodies could be useful to deal with occlusions and reflection in the water.

[1]https://tinyurl.com/waterDetectionDataset

Figure 2: Challenges: misleading images (left), water reflections (middle) and occlusions (right).

Approach	Validation (P/R/F1)			Test (F1)
GA-1	0.58	0.67	0.61	0.61
GA-2	0.55	0.60	0.56	0.59
LA	0.58	0.77	0.60	0.59
Majority Voting	0.59	0.68	0.61	0.61

Table 1: Macro-averaged precision (P), recall (R), and f1-scores for visual flood level estimation.

5. Conclusion

Our experiments show that pose estimation and water detection provide useful clues for the assessment of flood levels. By building upon skeletons, the presented approach is invariant to gender, age and height. Main challenges for robust water level estimation represent occlusions and reflections. For future work, a larger, more balanced and more heterogeneous dataset is needed.

Acknowledgments

This work was supported by the Austrian Research Promotion Agency (FFG), Project nos. 855784, 856333, and 865973.

References

[1] B. Bischke, P. Helber, S. Brugman, E. Basar, Z. Zhao, and M. Larson. The multimedia satellite task at MediaEval 2019: Estimation of flood severity. In *Working Notes Proc. of MediaEval Wshp. (to appear)*.

[2] B. Bischke, P. Helber, Z. Zhao, J. de Bruijn, and D. Borth. The multimedia satellite task at MediaEval 2018: Emergency response for flooding events. In *Working Notes Proc. of the MediaEval Wshp.*, 2018.

[3] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *Proc. of CVPR*, 2017.

[4] M. Larson, P. Arora, C.-H. Demarty, M. Riegler, B. Bischke, E. Dellandrea, M. Lux, A. Porter, and G. J. Jones. *"Working Notes Proc. of the MediaEval 2019 Wshp. (to appear)"*.

[5] M. Zaffroni, L. Lopez-Fuentes, A. Farasin, P. Garza, and H. Skinnemoen. AI-based flood event understanding and quantification using online media and satellite data. In *"Working Notes Proc. of the MediaEval 2019 Wshp. (to appear)"*.