Flood Depth Estimation from Web Images

Zonglin Meng
Department of Computer Sciences
University of Wisconsin-Madison
Madison, Wisconsin
zmeng32@wisc.edu

Bo Peng
Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin
bo.peng@wisc.edu

Qunying Huang
Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin
qhuang46@wisc.edu

ABSTRACT

Natural hazards have been resulting in severe damage to our cities, and flooding is one of the most disastrous in the U.S and worldwide. Therefore, it is critical to develop efficient methods for risk and damage assessments after natural hazards, such as flood depth estimation. Existing works primarily leverage photos and images capturing flood scenes to estimate flood depth using traditional computer vision and machine learning techniques. However, the advancement of deep learning (DL) methods make it possible to estimate flood depth more accurate. Therefore, based on state-of-the-art DL technique (i.e., Mask R-CNN) and publicly available images from the Internet, this study aims to investigate and improve the flood depth estimation. Specifically, human objects are detected and segmented from flooded images to infer the floodwater depth. This study provides a new framework to extract critical information from large accessible online data for rescue teams or even robots to carry out appropriate plans for disaster relief and rescue missions in the urban area, shedding lights on the real-time detection of the flood depth.

CCS CONCEPTS

• **Computing methodologies** → *Machine learning*; • **Applied computing** → Earth and atmospheric sciences; • **Information systems** → *Geographic information systems*.

KEYWORDS

Mask R-CNN, Detection, Flood depth, Resilient

ACM Reference format:

Zonglin Meng, Peng Bo, and Qunying Huang. 2019. Flood Depth Estimation from Web Images. In *2nd ACM SIGSPATIAL International Workshop on Advances on Resilient and Intelligent Cities (ARIC'19), November 5, 2019, Chicago, IL, USA. ACM, New York, NY, USA, 4 pages.* https://doi.org/10.1145/3356395.3365542

ARIC'19, November 5-8, 2019, Chicago, IL, USA

ACM978-1-4503-6954-1/19/11…$15.00
https://doi.org/10.1145/3356395.3365542

1 Introduction

Every year a number of people suffer from floods which have been unleashing several waves of destruction on our habitat [6]. Floods nowadays are critical concerns among the society because they are becoming common in most advanced countries and directly impact human lives [9]. For example, recent flood disasters in the United States (e.g., the 2018 Hurricane Harvey), Japan (e.g., the 2018 Shikoshu and Western Honshu floods) and Philippines (e.g., the 2017 Visayas and Mindanao floods) have shown that cities are vulnerable to the torrential floods. Thus, efficient assessments of floods should be proposed to make the city resilient to flood and minimize the damage [10].

Flood impact assessment, including identification of inundation depth, is critical in both real-time and longer-term applications in resilient city development. Currently, researchers have been developing various models to estimate the level or depth of flood from photos and images capturing flood scenes. For example, Narayanan, Lekshmy, Rao and Sasidhar detected building objects from flooded images and applied the building's height as the reference to estimate the flood level [5]. Manoj, Mohan and Rao segmented human body from crowd-sourced images to four regions to estimate the depth, which is a significant factor to judge the extent of flood [4]. However, driven by the advancement of state-of-the-art computer vision and machine learning techniques, especially the deep learning (DL) methods, the ability of object detection and prediction from images has made profound applications in the natural disaster estimation, such as the detection of extreme weather events [11], severe weather forecasting [8] and flood impact assessment [3]. Correspondingly, object detection is becoming more and more precise. Existing methods for flood depth estimation based on hand-crafted objects features are no longer satisfying.

Therefore, this paper develops a novel method to accurately evaluate the flood depth from normal camera images, which are publicly available from social media and the Internet. Specifically, an advanced DL model, Mask R-CNN, is used to detect and extract the human body objects in the flood affected area (Figure 1b). Next, the age, gender, ethnicity and body keypoints (Figure 1c) for each human object are inferred and the average height corresponding to different gender, age, and ethnicity will be used as the reference height for each object. Finally, based on the keypoints, the ratio of the height under water and above water for each human object can be calculated and further used to estimate the flood depth (Figure 1d).

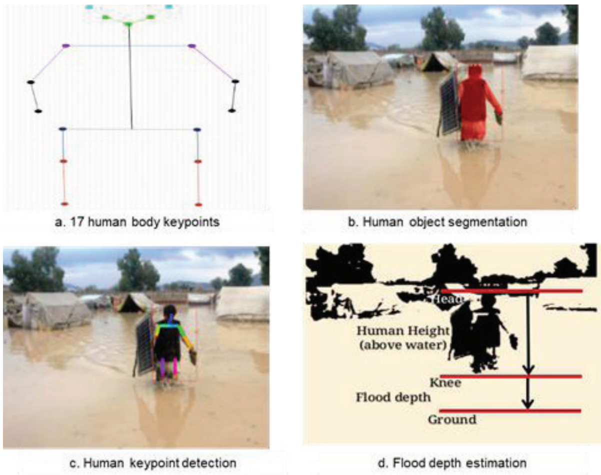

Figure 1. The visualization of flood depth estimation method

The rest of this paper is organized as follows. Section 2 describes the methodology, algorithm of the system to predict the depth of flood. Section 3 describes the experimental results and the current limitation of the work. Section 4 shows the conclusion and the next step.

2 Methodology

As shown in Figure 2, the workflow of our approach consists of three parts. The first is preparing training and testing datasets which are downloaded from search engine and making the datasets ready for the image detection model. The second step is using Mask R-CNN and Face++ Emotional Application Programming Interface (API) to get the prediction data. And the last step is to estimate the depth of flood (Algorithm 1). Although estimating depth based on cars or buildings is also likely to generate a good result, detecting human objects can be more straightforward to save people in such emergent situation, which is one of the advantages of the model.

2.1 Human object detection

The first step is to detect and segment the human body in the flood affected area. Since we estimate the water depth with human height as the reference, accurate detection of the human body can play an important role in the estimation. Due to the high performance of Mask R-CNN, it is relatively easy to segment out human bodies that are partially submerged in the flood. As an extension of Fast-RCNN, Mask R-CNN included a simple but effective layer, RoIAlign, to increase the accuracy of mask and further help segment out the target [1]. Although Mask R-CNN was trained on MS COCO dataset, which did not particularly contain images capturing flood affected areas, the experiment result still proves that the pre-trained model can generate accurate bounding boxes and masks for the partially submerged objects.

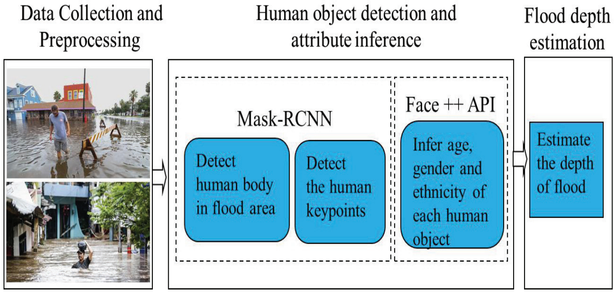

Figure 2. The workflow of the proposed work for flood depth estimation from images

2.2 Human object attributes detection

Since we utilize the human body to estimate the depth of water, the attributes of person (e.g., age, gender and race) are conducive to improve the accuracy of the output. After extracting attributes like age, gender and race from faces, heights of different groups can therefore be estimated. The attributes extraction are accomplished by Face ++ Emotional API that provides the estimation with critical information of age, gender and ethnicity inferred from human faces. This API aims to detect human emotion, but it can also generate additional human object attributes, such as age, gender, head pose or even beauty [2]. After getting the classification results, we use the average height of different groups to predict the result.

2.3 Water depth estimation

Another implementation of Mask R-CNN is particularly illuminating. To estimate the flood depth, it is necessary to calculate the ratio of submerged human body (i.e., the proportion of human body above the water over). In addition, Mask R-CNN can not only detect the human body but also be extended to generate the keypoints on human body (Algorithm 1 line 3). The problem of estimating the ratio is converted to find the remaining human body keypoints above the water. The model is trained with COCO dataset 2017 including annotation files for human keypoints. These keypoints are features of human body, such as head, shoulder, ear, hand, elbow, keen, and feet. By utilizing these keypoints, we can detect the specific parts of the human body that are above the water surface, which can help estimate the ratio (Algorithm 1 line 8). After obtaining the ratio, we can estimate the depth of water (Algorithm 1 line 11).

ALGORITHM 1

```
1  Detect the human in the flooded affected area
2  for each detected object, do
3        Produce keypoints for the detected human body
4        for these keypoint, do
5              Record the deepest point
6              Check which parts of the human body above water
7              Estimate the ratio and return the result
8        end
9  End
10 Estimate the depth of flood by the ratio and average height of each object
```

3 Experiment and results

3.1 Human object detection and attribute inference

For the detection of the human body, we use the pre-trained model, which was trained on COCO 2017 consisting of training images and annotations in JavaScript Object Notation (JSON) format. Providing human-labeled bounding box and Mask, COCO dataset focuses on the common objects in context, in contrast with the testing images downloaded from web search engines using keywords "flood". The testing images are taken from different distances and has various resolutions. In total, there are 155 flood photos, which contain people of different demographic background in different locations. The pre-trained model can generate two types of labels for human body which are bounding boxes and mask (Figure 3). These images are then tested on Face ++ API, which can generate the age, gender and ethnicity (Figure 4).

Figure 3. Result of human detection in flood affected area

3.2 Human keypoints detection and flood depth estimation

For the detection of human keypoints, the model was trained on COCO 2017, where human keypoints are manually labeled and the annotation files contain the corresponding keypoints. However, even though the model was trained on these keypoints, it is hard to get the ideal output mentioned in Mask R-CNN paper. One limitation of our implementations is the detection of keypoints. More specifically, those keypoints are not accurate enough to predict the ratio and some of them are totally incorrect (see Figure 5). As a result, only some of the images can produce the ratio and final output (see Table 1). One reason is the obstacles (i.e., human body is hidden by some other objects). Our implementation can only generate accurate points for human with little or no obstacles. That means we still need to fine-tune the model to make it suitable for our images because a lot of parts of the human keypoints were hidden by the flooding area. If there are multiple objects with different height appearing in the same region, the depth estimation can vary from one object to the other. In this case, all different objects will be detected and have the different results, but the average depth will be recorded.

Figure 4. Result of Face ++ API on test images

We start to train the model with the pre-trained weights and use cross entropy instead of binary cross entropy as the loss evaluation metric. The training dataset is COCO 2017, which contains keypoints annotations. The model is trained with learning rate of 0.002 and a total of 160 epochs with1k iterations per epoch.

4 Conclusion

This paper proposes to detect the flood depth by estimating the height of submerged human body. Such an estimation can be meaningful to real time urban flood assessment and can help to make informed decisions for flood response. To make the city more habitable, we should not only focus on how to protect the ecosystem but also prepare for the unpredictable natural hazards, which are likely to influence how people live, where they live and how cities develop. Flood depth estimation can provide critical information about the flood damage and help with emergency rescue. Since the result can be generated within a few seconds, this approach can be used in near real-time. But more importantly, Mask-RCNN can be deployed to recognize and track objects on real time videos which can strengthen the application of this system. Besides, to determine whether human lives are in danger due to flooding from web images could also be a meaningful topic.

Figure 5. Human keypoint detection for four selected images: only *image a* generate the correct keypoints

This study paves the way to leverage DL and computer vision techniques to estimate the depth of flood inundation, without hydrological modeling and field work during the flood event. The proposed approach is mainly constrained by limited available human annotations of photos in flood scenarios for model training and generalization.

Table 1. Result of human object attribute inference, and flood depth estimation

Image	Age, gender, ethnicity	Height (cm)	Ratio	Water depth (cm)
1	Default value	176	0.695	53.68
2	37, Male, Black	176.9	x	x
3	{16, Female, Asian}, {7, Female, Asian}	162.2 125.4	x	x
4	54, Male, India	176.6	x	x

Since the detection of human keypoints still remains a challenge for photos in flood scenarios, the next step is to generate more accurate human keypoints by fine-tuning the model and training on larger dataset. Besides, we also plan to implement the detection on video to expand its applicability.

5 ACKNOWLEDGMENTS

This research is funded by the National Science Foundation with Project ID: 1940091

REFERENCES

[1] K. He, G. Gkioxari, P. Dollár and R. Girshick. 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), (Oct. 2017), 2980-2988.
DOI: 10.1109/ICCV.2017.32.

[2] Kang, Y, Jia, Q, Gao, S et al. 2019. Extracting human emotions at different places based on facial expressions and spatial clustering analysis. Transactions in GIS. (June. 2019), 23: 450– 480. DOI: 10.1111/tgis.12552

[3] Merz, B., et al. 2010. Review article" Assessment of economic flood damage". Natural Hazards and Earth System Sciences, 2010. 10(8): p. 1697-1724. DOI:10.5194/nhess-13-1351-2013.

[4] M. Geetha, M. Manoj, A. S. Sarika, M. Mohan and S. N. Rao. 2017. Detection and estimation of the extent of flood from crowd sourced images, 2017 International Conference on Communication and Signal Processing (ICCSP), (April. 2017), 0603-0608. DOI: 10.1109/ICCSP.2017.8286429.

[5] R. Narayanan, V. M. Lekshmy, S. Rao and K. Sasidhar. 2014. A novel approach to urban flood monitoring using computer vision, Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), (July, 2014), 1-7. DOI: 10.1109/ICCCNT.2014.6962989.

[6] Smith, A.B. and J.L. Matthews. 2015. Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. Natural Hazards, 2015: p. 1-23. DOI: 10.1007/s11069-015-1678-x.

[7] Yunjie Liu, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, William Collins. 2016. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets.

[8] Will Eadson. 2011. Water and the City: Risk, Resilience and Planning for a Sustainable Future, Housing Studies, 26:04, 635-637, DOI: 10.1080/02673037.2011.568219

[9] Lhomme S., Serre D., Diab Y., and Laganier R. 2012. Analyzing resilience of urban networks: a preliminary step towards more flood resilient cities, Nat. Hazards Earth Syst. Sci., 13, 221-230, https://doi.org/10.5194/nhess-13-221-2013, 2013.

[10] Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M.F., and Collins, W.D. 2016. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. ArXiv, abs/1605.01156.