Semi-supervised identification and mapping of surface water extent using street-level monitoring videos

Ruo-Qian Wang & Yangmin Ding

Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA; NEC Laboratories America Inc, 4 Independence Way, Princeton, NJ, USA

ABSTRACT

Urban flooding is becoming a common and devastating hazard, which causes life loss and economic damage. Monitoring and understanding urban flooding in a highly localized scale is a challenging task due to the complicated urban landscape, intricate hydraulic processes, and the lack of high-quality and resolution data. The emerging smart city technology such as monitoring cameras provides an unprecedented opportunity to address the data issue. However, estimating water ponding extents on land surfaces based on monitoring footage is unreliable using the traditional segmentation technique because the boundary of the water ponding, under the influence of varying weather, background, and illumination, is usually too fuzzy to identify, and the oblique angle and image distortion in the video monitoring data prevents georeferencing and object-based measurements. This paper presents a novel semi-supervised segmentation scheme for surface water extent recognition from the footage of an oblique monitoring camera. The semi-supervised segmentation algorithm was found suitable to determine the water boundary and the monoplotting method was successfully applied to georeference the pixels of the monitoring video for the virtual quantification of the local drainage process. The correlation and mechanism-based analysis demonstrate the value of the proposed method in advancing the understanding of local drainage hydraulics. The workflow and created methods in this study have a great potential to study other street-level and earth surface processes.

KEYWORDS

Segmentation; deep learning; monoplotting; smart city; monocular visual data

1. Introduction

Urban flooding is becoming more common and destructive to the society, causing loss of lives, economic damage, social disruption, and housing inequity. As a major component, urban flooding contributes to the general flooding problem, which costs the damage of $9 billion and 71 lives annually (National Academies of Sciences, Engineering, and Medicine, 2019). Communities across the world are facing similar challenges and the
increasing trend will continue with urbanization, the growing number of extreme weathers, and changing climate. Planning and improving toward a safe and resilient urban environment is urgent.

Compared to other types of flooding, urban flooding is known difficult to monitor and quantify, so the understanding of its drainage process, especially on the local scale, is highly limited due to data availability and measurement issues. Anthropogenic drainage systems are primarily designed to mitigate flood risks (American Society of Civil Engineers, Urban Water Resources Research Council, 1992). Despite the importance in managing flooding events, there is little information available regarding the interaction between the runoff, the paved surface, and the design of the drainage systems. The bottleneck that prevents developing and improving urban flood models and forecasting systems is the scarcity of data. Urban flooding events, especially those less dramatic ones, are poorly documented (Galloway et al., 2018). This data scarcity problem is partly due to 1) the high cost of sensing network installation and maintenance for the wide urban area and the traditional sensor-based water-level measurement, which is usually designed for deep water flows and limited to sparse fixed locations, 2) technical difficulties in remote sensing, (e.g. satellite imaging is affected by cloud cover and complex street geometry, and its revisit interval is too long to capture flooding events), and 3) high labor costs associated with traditional geological surveys (e.g. high water mark data collection after floods). These knowledge and data gaps prevent researchers from systematically examining the events, reliably identifying the driving mechanisms, and effectively validating numerical models. Consequently, decision makers cannot be informed about the strategies of preventing, mitigating, and assessing flooding events.

The recent development of smart city technology is providing an exciting new opportunity to advance the data collection and understanding of urban floods. Across urban areas, various traffic cameras are pre-existing to monitor traffic, street activities, and coastal areas to provide a valuable data source to track urban floods (Fan, Jiang, & Mostafavi, 2020; Hu & Wang, 2020; Huang et al., 2019; Jongman et al., 2015; Wang, Hu, Zhou, & Yang, 2020; Wang, Mao, Wang, Rae, & Shaw, 2018). Smartphones have fundamentally changed the lifestyle and business in the world and are becoming a reliable means of citizen science-based data collection, e.g. smartphone Apps and social media were used to collect critical and real-time flooding data (Wang, Mao, Wang, Rae, & Shaw, 2018). These emerging, visual data-based methods could potentially cover a large area and recognize shallow water ponding with affordable costs and abundant details. However, analyzing such visual data is challenging. First, it is difficult to recognize the water pixels in videos or photos because the varying weather condition and daylighting could lead to changing illumination for the monitored scene so that water ponds usually have fuzzy boundaries, the background could significantly change to prevent reliable object-based image analysis, and AI and computer vision-based segmentation algorithms could be challenging to achieve practical accuracy due to heavy labeling and computational loads. In addition, these street-level data sources usually have oblique angles and distorted imagery, which pose further challenges in data processing comparing to the traditional rigorous scientific measurements. The present study is designed to develop a new semi-supervised flood water recognition scheme, which requires a low amount of labeling effort, coupling a visual data projection method called monoplotting to allow a high quality and georeferenced data processing.
The present study is designed to explore the use of street-level surveillance videos to gain high-quality and valuable information of urban flooding. Specifically, a new framework is developed to analyze monitoring videos with semi-supervised segmentation-based urban flooding image processing to access to the flooding information at an unprecedented level—it allows users to estimate and map local flood extents with 1) varying weather and lighting condition, 2) poor contrast between water and the background, 3) affordable training and labeling loads, and 4) oblique perspectives. Section 2 introduces the work and the technical challenges related to this study including segmentation, monoplotting, and their application to urban floods. Section 3 describes the photogrammetry methods, machine learning skills, and the details of the field experiment. The validation of the methods and analysis results are presented in Section 4 and summarized at last.

2. Related work

2.1. Machine learning based object-based image analysis

Object-based Image Analysis (OBIA) is a technology to classify image pixels into segments and measure physical quantities based on the relative position between the classified segments and the background. OBIA has been used to analyze aerial photography for vegetation and urban landscapes, natural disaster damage analysis, and risk management (Blaschke, 2010; Lee, Yang, Nathan, & Blair, 2018; Van der Sande, Jong, & Roo, 2003). Classic manual image segmentation methods (Barrow & Tenenbaum, 1981; Grady, 2006; Kass, Witkin, & Terzopoulos, 1988; Roerdink & Meijster, 2000; Álvarez, Gevers, LeCun, & López, 2012) are often sensitive to subjective inputs and image noise, while the recent deep learning-based algorithms allow more robust extraction of visual features for segmentation analysis, e.g. the supervised learning algorithms of fully convolutional network (FCN) (Long, Shelhamer, & Darrell, 2015), the convolutional neural network (CNN) architectures (Krizhevsky, Sutskever, & Hinton, 2012), VGG16 (Simonyan & Zisserman, 2014), ResNet (He, Zhang, Ren, & Sun, 2016), and the compound methods by Hariharan, Arbeláez, Girshick, and Malik (2014), Arnab, Miksik, and Torr (2017), Bai and Urtasun (2017), and Liu, Kummerow, and Elsaesser (2017), which combined multiple methods for segmentation. But these supervised methods still require heavy labeling burden, and they are difficult to scale up to other applications, especially in the cases where experience and deep expertise are required.

The image segmentation has been applied to analyze remote sensing data in urban studies. A detailed review focusing on land cover can be found in Ma et al. (2017) and Hossain and Chen (2019). However, these classical segmentation algorithms, such as region growing (Jasiewicz, Niesterowicz, & Stepinski, 2016), region merging (Vasuki, Sharma, Ibrahim, & Arciuli, 2017), and clustering (Chen, OuYang, & Chou, 2018), are insufficient and unreliable in analyzing water data that contains heterogeneous textures and uneven/low illumination. Deep learning-based image segmentation can, to an extent, address the technical issues but they show limitations in recognizing fuzzy water pond boundaries under the natural lighting conditions and suboptimal background contracts, making segmentation from a single scene extremely challenging. The labeling burden in such tasks is also daunting as the dramatically changing environment requires dense label
data, which could be generated through a tedious, costly, and inaccurate process. The background change such as the reference object motion adds extra errors to the object-based measurements. For example, Jafari et al. (2021) applied segmentation to estimate flood levels in downtown Houston, Texas during Hurricane Harvey and found the weather and background motion introduced nontrivial errors to the analysis.

Semi-supervised methods, which could alleviate the burden of manual labeling, are gaining a rising attention recently. In this approach, a small set of images are fully annotated but the majority of the image dataset are unavailable. The typical semi-supervised methods include creating surrogate classes (Pickard, 2002), adding entropy regularization (Kull, 2005), using Generative Adversarial Networks (GANs) (Svenningsen, Brandt, Christensen, Dahl, & Dupont, 2015), or averaging ensemble networks (Gimmi, Ginzler, Müller, & Psomas, 2016; Perone, Ballester, Barros, & Cohen-Adad, 2018). Among these methods, self-training is considered an emerging and future machine learning strategy, in which a network is created using the labeled data to train a preliminary network, which is used to predict the labels of the unlabeled data to produce the so-called pseudo labels. Then, the network is retrained using the augmented training set combining the manual labels and the pseudo labels for improved accuracy (Bai and Urtasun, 2017; Pathak, Agrawal, Richhariya, Sadaf, and Hasan, 2015; Rajchl et al., 2016). Although this approach can leverage unlabeled images, mistakes made in the early training process could propagate back to the network and be amplified during training (Chapelle, Scholkopf, & Zien, 2009; Zhu & Goldberg, 2009). Despite several techniques were proposed to overcome this issue, self-supervised learning is still challenging to achieve a high level of accuracy and the training process can be difficult to control to gain the convergence.

### 2.2. Monoplotting

Street-level imagery obtained from smartphones, UAV (Unmanned Aerial Vehicle), and urban and coastal monitoring cameras, tend to provide photos and videos at oblique angles. To retrieve valuable data from such emerging data sources, we need to break through the barrier posed by oblique imagery data. Monoplotting is a potential solution that allows the users to obtain image registration to the 3-D Digital Elevation Model (DEM). Monoplotting was named by Makarovic (1973), who used this method to transform photographs into georeferenced data (Bayr, 2021). This method has several applications and advantages in analyzing geoscience problems: First, the image registration enables the geo-locating of the captured objects and process in the DEM (or a map) for image content quantification. Second, monoplotting requires only a single image and a DEM to obtain 3-D observation of the scene to avoid the data requirement of stereo images to reconstruct 3D information. Monoplotting has been used in extracting long-term earth surface changes using repeat photography (Bayr & Puschmann, 2019; Kull, 2005; Pickard, 2002; Svenningsen, Brandt, Christensen, Dahl, & Dupont, 2015), obtaining 3-D view of the scene with the ground-based photographs, and analyzing historical photographs from the time before aerial photography appeared (Gimmi, Ginzler, Müller, & Psomas, 2016). Recently, this method showed usefulness in the analysis of crowdsourcing images to study emergent and rapid processes such as flooding (Golparvar & Wang, 2020; Triglav-Čekada & Radovan, 2013) and the potential to use the oblique data to communicate with
stakeholders taking the advantage of the similarity with daily-life perception and experience (Triglav-Čekada, Radovan, Gabrovček, & Kosmatin-Fras, 2011). Low-cost sensing technology also benefited from this powerful georeferencing method, e.g. in the monitoring of slow earth surface dynamics such as glacier movement and vegetation changes (Triglav-Čekada, Bric, & Zorn, 2014; Triglav-Čekada, Radovan, Gabrovček, & Kosmatin-Fras, 2011; Wiesmann et al., 2012). Dynamic information obtained in the video could be used to reconstruct 3-D deformation and movement such as in the dam breaking event for forensic analyses (Travelletti et al., 2012; Yuan, Bazzett, Padnani, & Wang, 2021) and multitemporal landslide monitoring (Makarovic, 1983).

Recognizing the great value of monoplotting, handy software and codes have been created for the geoscience community, including the OP-XFORM project (Aschenward, Leichter, Tassner, & Tappeiner, 2001), the JUKE method (Corripio, 2004), Georeferencing oblique terrestrial photography (Mitsitika, Machado, Habib, & Gonçalves, 2004), the 3D Monoplotter (Fluehler, Niederoest, & Akca, 2005), and the DiMoTeP (Conedera et al., 2018). Recently, the WSL Monoplotting Tool (WSL-MPT) developed by the Swiss Federal Research Institute (WSL) is gaining popularity and has been applied to the quantitative analysis of natural hazards (Scapozza, Lambiel, Bozzini, Mari, & Conedera, 2014; Triglav-Čekada, Radovan, Gabrovček, & Kosmatin-Fras, 2011), glacial processes (Stockdale, Bozzini, Macdonald, & Higgs, 2015), and land cover changes (Gabellieri & Watkins, 2019; McCaffrey & Hopkinson, 2017; Stockdale, Bozzini, Macdonald, & Higgs, 2015). A similar tool called Pic2map, leveraging the convenience of the Geographic Information System (GIS) software QGIS, shows a strong rising trend (McCaffrey & Hopkinson, 2020). The present study used Pic2map for the following analysis.

2.3. **Computer vision-based flood data mining**

Computer vision and machine learning have been applied to analyze street-level flood data. The early applications focused on the image level recognition of social media data and georeferencing the data with or without the related text messages (Wang, 2018; Wang, Hu, Zhou, & Yang, 2020; Wang, Mao, Wang, Rae, & Shaw, 2018). At the sub-image level, segmentation has been used to perform a more accurate measurement of flood extents and depths, such as the application to satellite imaging (Gabellieri & Watkins, 2019; Nemni, Bullock, Belabbes, & Bromley, 2020) and those involving thresholding and region growing methods (Arshad, Qureshi, Inam, & Omer, 2020). With deep learning, more accurate measurement became feasible, such as Chaudhary, Taran, Bajaj, and Sengur (2019), Moy de Vitry et al. (2019), and Javid et al. (2020), who applied segmentation to recognize the water area and depth. Additionally, Chang et al. (2019) reviewed various machine learning methods applied in flood forecast modeling and reported that machine learning methods are the key in developing early warning systems for urban flood hazards. However, these studies showed that the accuracy of deep learning methods could be significantly impacted by the distortion and mirror reflection of the image (Song & Tu, 2021) and they are difficult to scale due to the high cost of infrastructure and maintenance constrains to derive precise flood depth data (Moy de Vitry et al., 2019). In addition, the data quality could be compromised due to the difficulty in georeferencing (Golparvar & Wang, 2020).
3. Methods and experiments

3.1. Semi-Supervised image classification

Recognizing the boundary of water ponding on road surface is a challenging task. Mirror reflection, oblique observation angle, and image distortion make the boundary of water ponding fuzzy. The traditional supervised learning methods of water ponding recognition could, to some extent, address the problem but it requires drawing irregular-shape polygons for some, if not all, video frames, which is usually time-consuming, expensive, and tedious. It is also challenging to fit the simplified boundary of the polygon to the fuzzy water ponding edge with varying background and imaging quality. In addition, manual labeling, which depends on the experience and judgement of the label preparer, could be difficult to scale due to the high labor costs.

Recognizing the challenges, we proposed a new semi-supervised segmentation algorithm, which can efficiently recognize semantic areas from an image with minimal labeling inputs to recognize water ponding boundaries without the impact of labeling accuracy. Specifically, the video is first cropped to the interested area with a matrix of $X_{M \times N \times S \times 3}$, where $S$ frames have $M$ rows, $N$ columns, and $3$ bands of pixels. An unsupervised segmentation method, i.e. Felzenszwalb algorithm (Felzenszwalb, Girshick, McAllester, & Ramanan, 2009), was applied to cluster the pixels of a representative frame $X_{M \times N \times 3}$ into $F$ ($\ll M \times N$) segments ($\hat{X}_{F \times 3}$) to generate a reduced-order image. A sparse grid with $L$ ($\ll F$) points was developed covering the representative frame, $\hat{X}_{F \times 3}$. So the $L$ points can be seen as a far smaller subset, $\hat{X}_{L \times 3}$, of $\hat{X}_{F \times 3}$. In fully supervised learning algorithms, users must label $M \times N \times S$ pixels, while in the proposed supervised algorithm, the original image is simplified to $F$ segments and only part of the $L$ points need to be labeled. Therefore, much labeling effort is saved in this new semi-supervised method.

Among the $L$ grid points, we only labeled $L_1$ “permanent” dry points (the purple dots in Figure 1) and $L_2$ “permanent” wet points (the red dots in Figure 1), where $L_1+L_2<L$. So $L_0$ ($=L-L_1-L_2$) points, which are difficult to manually determine or change over time, are left to be classified later. The $L_1$ and $L_2$ labeled points ($\hat{X}_{L_1}$ and $\hat{X}_{L_2}$) were then used to train...
a supervised classification model, i.e. a decision tree model in the present project, to classify the segments containing them into the dry and wet groups (Pedregosa et al., 2011). The supervised classification model is next used to predict the segments that are not labeled. In this approach, the segmentation only relies on the input of the users that have the highest confidence ($\hat{X}_{L_1}$ and $\hat{X}_{L_2}$), so the subjectivity and uncertainty of the labeling input is minimized. The methodology is detailed in Figure 1. In general, the labeling task is reduced from $M \times N \times S$ to $L_1 + L_2$. As an example, the field experiment described below generates a video of $M \times N \times S = 1072 \times 1920 \times 145 \approx 3 \times 10^8$ pixels to label/predict, and they were reduced to the number of $L_1 + L_2 = 16 + 92 = 108$ points to label. For each training of the supervised learning, the dataset was randomly split into training (70%) and testing (30%) sets. The average training accuracy was

3.2. Monoplotting

To obtain a wide observation coverage, monitoring cameras usually employ a wide-angle lens, which distorts the image and poses challenges in georeferencing. We adopted the monoplotting technique to address this issue. Ground Control Points (GCPs) were

![Figure 1. Overview of the data analysis method.](image-url)
identified in the image and the DEM. The best DEM data usually has a resolution of 3 meters and after postprocessing details of local objects are removed. So the traditional monoplotting method that relies on DEM variation would not work for highly localized environments. Here, we instead assume the elevation of the scene is constant and we use satellite images that have information of on-ground objects to determine the GCPs. The principle to obtain the GCPs is to identify a series of widely spread points that can be recognized in both the video and the satellite image and can be registered with one-to-one correspondence. Once the GCPs are determined, an optimization scheme is employed to find the location and parameters of the camera to match the two sets of GCPs from the image and DEM in the projection plane to allow the determination of the pixel coordinates in the 3D space. This involves three steps: First, the GCPs captured in the DEM (called DEM GCPs) are projected to a plane assuming a camera is positioned in a location $(x, y, z)$ using the equation below,

$$
\begin{bmatrix}
  u \\
  v \\
  w
\end{bmatrix} = K_{3 \times 3} R_{3 \times 3} \begin{bmatrix}
  x + T_x \\
  y + T_y \\
  z + T_z
\end{bmatrix}_{3 \times 1}
$$

(1)

where $u$ and $v$ are the horizontal and vertical coordinate of projected points in the image plane, $K$ is the intrinsic camera matrix in which there are at least 4 unknown parameters such as focal length and image center coordinates. $R$ is the rotation matrix and $T$'s components are for the translation vector. Second, the projected GCPs are compared with the GCPs identified in the image (called Image GCPs). The projected GCPs and image GCPs are matched manually to ensure the one-to-one relationship. Note that most monoplotting studies used manual matching method, while Golparvar and Wang (2021) developed a semi-automatic method, which could potentially save the users’ effort. Finally, the camera parameters and location are iteratively adjusted to reproject the DEM GCPs until the best comparison is achieved. The real-world coordinates obtained through monoplotting are then used to project the recognized water ponding distribution to the DEM or a map (Figure 1). This georeferenced data will allow the measurement of the water areas in the real-world length scale. The monoplotting operation was conducted with Pic2map (McCaffrey & Hopkinson, 2020).

This study is probably the first application of monoplotting to obtain highly localized flooding extents from in-street monitoring videos and our innovative improvement to the method allows more flexible applications of the method in other fields such as green infrastructure design (Zhou & Guo, 2022). The assumption that the scene is flat may introduce extra error in estimating flood ponding because of the small elevation difference among the GCPs. Since we used the exact coordinate of the camera, this error is insignificant after relaxing the GCP matching process. In general, this modification allows, to the authors’ knowledge, this first application of monoplotting to estimate the flood extent in a highly localized environment.
3.3. Field experiment

A field experiment was conducted in January of 2022 at a parking lot in Princeton, New Jersey. A solar panel powered video camera (Reolink Go 4 G) was installed on a light pole. This camera transmitted data through the wireless 4 G network and recorded the field every 30 seconds. The camera was zoomed to a drainage well to collect the runoff of the area in the view of the camera.

A major purpose of this study is to understand the interaction between the runoff, the paved surface, and the drainage flow to inform the design of drainage systems. In addition, the water level information in the drainage well can indirectly validate the video-based estimation of flood extents. So we designed the experiment with a water level sensor – an ultrasonic liquid level transmitter with an accuracy of 0.5% of the measurement value – in the well. The water inside the well was connected to an outlet during the experiment. The relative position of the camera to the drainage well is shown in Figure 2.

Sample frames from the captured video are shown in Figure 3. These images show that the water ponding surface had strong mirror reflection of the surrounding environment such as trees and cars. The image illumination contract varied over time depending on the

Figure 2. The setup of the field experiment.

Figure 3. Sample frames of the captured video, in which objects were observed changing and moving in the video with varying illumination and camera blocking.
cloud cover and weather. The background objects such as human and cars had movement and the camera was sometimes blocked by the ongoing traffic. These factors contribute to the uncertainty of the data analysis. The DEM data in 10-meter resolution was collected through the database of New Jersey Department of Environmental Protection Open Data Webpage (https://www.state.nj.us/dep/gis/wmalattice.html) to support the monoplotting operation. Note that monoplotting was originally designed to work with DEM data, but the DEM data is flat for the monitored parking lot and the geospatial scale is too low to capture any features in DEM. We used a satellite image from Google Map overlaying on the DEM data so that the corresponding features could be captured.

4. Results

Sample segmentation results along with their original images are shown in Figure 4 for comparison. The proposed method involves two steps – the supervised learning and unsupervised learning. The supervised learning step was validated automatically in the scikit-learn package and a high accuracy of training was ensured before applying to process the video, while the unsupervised learning was difficult to validate quantitatively.

Here, we instead make a qualitative validation: the flooding extent (grey areas) was overlapped semi-transparently over the original video frames for a manual comparison. We observe that the identified wet areas cover all the major ponding areas of the original images and provide extraordinary new details that human labeling, e.g. the polygon label created by labeling codes such as LabelMe (Russell, Torralba, Murphy, & Freeman, 2008), can hardly capture. The comparison revealed that the majority of the result is satisfactory and the only miss-classification was shown at the edge of the curb due to poor illumination.

The monoplotting method was performed following the illustration in Figure 5. GCPs were identified at the features in the original image and the corresponding points were determined in a map. Their correspondence was shown in Figure 5 and we found the

Figure 4. Sample classification results: the upper panels are the original image and the grey masks in the lower panels are the segmentation results.
camera position and pose could be reliably obtained as shown at the yellow point. Using the obtained geographic coordinates (Figure 2), we reproject the identified water ponding to the real-world coordinate system and a sample flood distribution and the obtained pixel-level geographic coordinates are shown in Figure 2.

From the obtained segmentation results, we can obtain two types of water ponding distribution estimates. The first is SOFI (Moy de Vitry et al., 2019), i.e. the ratio of the water pixel number to the total pixel number. As discussed, this index could bias the estimate of flooding situation because every pixel is treated equally. So we developed a new index called “Perspective Corrected Flood Index” (PCFI), which calculates the ratio of the water area to the total area in the projected 3D coordinate:

\[
P_CFI = \frac{\sum_{i=1}^{n} C_i W_i}{\sum_{i=1}^{n} C_i}
\]

where \(C_i\) obtained through monoplotting is the perspective correction factor for Pixel i, \(W\) is the indicator of water pixel (\(W = 0\) if the pixel is covered by water and \(W = 1\) if it’s not), and \(n\) is the total number of pixels in the image.
The water extent time series: the pixel-based SOFI and area-based PCFI over time, which is compared with the water level inside the drainage well.

The two indices for the video are shown in Figure 6. Due to the heavy uncertainty in the data collection and processing, these two data have strong fluctuation over time. A 5-min moving average filter was applied to remove the strong fluctuation. The pixel-base SOFI is lower than the PCFI. This is expected because each of the far field pixels covers a larger area than the near field. As the far field has more water ponding, the pixel-based SOFI underestimates the ratio of water coverage.

In comparison, the water level data collected inside the drainage well is shown in Figure 6. The water depth value was converted from the measurement of the distance from the transmitter to the water surface with an interval of 5 mins and the initial water level at the beginning of the experiment was marked “0” as the reference. The water depth was found to reach the peak after the first peak of SOFI or PCFI. This could be attributed to the fact that the precipitation peak leads to the peak of surface water ponding. The water on the road surface took about 10 mins to flow to the drainage well and thus the water level in the well has a delay of 10 minutes than the road surface. This observed process is sensational and indirectly proves that the data process was reliable.

From the time series, we can establish a correlation relationship between the water depth in the drainage well and the water ponding extent on the road surface. In Figure 7, the rainfall event is divided into two phases based on the water level change in the well: rising and falling. The trend shows that at the beginning of the rainfall, the flood extent decreases when water level increases in the drainage well, whereas after the peak of the water level in the well, the flood extent starts increased and then remained the same level when water level was falling to the well. This indicates two different drainage processes in the local storm water runoff. In the rising phase, the water ponding was generated by direct rainfall, which flowed into the drainage well so that water level in the well increased and surface water extent decreased. In the falling phase, the surface water extent was mainly formed by collecting the surrounding storm water. In this phase, the water was mainly staying on the road surface without strong runoff to the well and the water in the well exited.
from the outlet in the well to cause the water level’s declination in the well. The study shows that using the advanced machine learning method and monoplotting allows a detailed study in the local drainage and runoff process. The spatial and temporal resolution obtained in this study cannot be reached using the traditional methods.

5. Discussions

The semi-supervised segmentation method shows a good performance. It captures the water extent on the road surface with environmental uncertainties and changing background. Although the noise due to mirror reflection and illumination has been mitigated, we still observe these factors impact the result to a non-trivial extent. The moving average of the data helps further removing such noise but may risk losing physical information in the process. The methods involved in this study still have a potential to improve such as using more advanced unsupervised clustering schemes and supervised classification models. In addition, the resolution and level of labeling should be tested to find the best combination for enhanced accuracy.

The proposed semi-supervised segmentation method is suitable for the task to identify uncertain boundaries based on the image texture. Since only a limited places on the image are required to be labeled, this method significantly reduces the density of labeling inputs. Leveraging on the temporal connection between the frames of the video, especially the fixed-camera footage, the saving of labeling costs could be further increased.

The monoplotting shows a good value to georeference the data and correct the image distortion in this study. However, identifying GCPs needs extensive experience to reach a practical level and the matching process that depends on an optimization scheme could be difficult to converge. Iterative updates of GCPs and camera position Apriori to guide the convergence process are required. There is a hope that a full or semi-automatic
monoplotting scheme could be developed to help the users to perform this task. An effort has been made in Golparvar and Wang (2021), but a large-scale application and validation has not been performed.

This study presents an innovative and perspective-corrected method to estimate the flood/ponding extent during an event. Care and testing are needed to apply the developed method to analyze a real-world problem. We observe that the uncertainty in monoplotting could propagate to the final ponding extent estimate more easily. Higher angle of the view would result in an overestimate of the extent. In addition, missing the labeling of water ponds would result in underestimates of the flood extent. Recognizing the sources of uncertainties, we provide recommendations to potential users and developers in using and developing the framework: 1) if possible, write down all the possible parameters about the camera including the coordinate, height, and angle of view. The camera EXIF information, which could simplify the monoplotting step, should be kept and extracted. 2) Create a dense matrix of labeling points and label the easy ones. Avoid labeling the pixels that are difficult to determine, because this method is designed to leverage the computer’s calculation to classify the pixels. Over-labeling may over-ride the computer’s judgement and lead to errors. Pay attention to the training accuracy and be prepared to update the labels when more training is done and more experience of labeling is gained. 3) Compare different unsupervised learning schemes and use the most reliable one with the best parameters. 4) If possible, measure the dimension of the field and ponding area for validation purposes.

6. Conclusion

This study developed a novel semi-supervised segmentation scheme to recognize the boundary of surface water ponding. The method involves an unsupervised segmentation step to reduce the dimension of clustering, a manual input to label the “permanent” segments, and a step of supervised classification. This scheme shows a satisfactory performance and saves a significant amount of manual input load. This study demonstrates that a semi-supervised segmentation is suitable for fuzzy boundary identification with changing background and image quality. The monoplotting applied in this study shows a good value to reproject the image information to the 3-D real-world coordinate, which enables the quantification of the image information using photogrammetry methods.

The combination of the segmentation and monoplotting methods was shown informative to study the detailed local drainage process and a sample time series and a correlation relationship were obtained for application, which indicates the interaction between surface water ponding and drainage runoff is more complicated than previously anticipated and thus requires systematic investigation in the future.

Acknowledgement

This project was partially supported by the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), Grant No. 69A3551847102, issued to Rutgers, The State University of New Jersey. The authors thank the kind support of Dr. Ting Wang from the NEC Labs in this work.
Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Ruo-Qian Wang, Ph.D., is an Assistant Professor at the Department of Civil and Environmental Engineering, Rutgers University. Prior to joining Rutgers University, he worked as a Lecturer at the University of Dundee, UK. He received his Ph.D. degree in Environmental Fluid Mechanics from Massachusetts Institute of Technology in 2014, his M.S. degree from Singapore Stanford Partnership, and his B.S. degree from Beihang University. He has conducted Postdoctoral research at MIT and University of California Berkeley during 2014 to 2017. Dr. Wang’s research group focuses on developing numerical models to connect big data and decision-making in civil and environmental engineering systems.

Yangmin Ding, Ph.D., is a Researcher in the Department of Optical Networking & Sensing at NEC Labs America, Inc. He received his bachelor’s degree in Civil Engineering from Changsha University of Science and Technology, his master’s degree in Highway and Railway Engineering from Southeast University, and a Ph.D. degree in Civil and Environmental Engineering from Rutgers, the State University of New Jersey. His current research interests include dynamics and vibration, structural health monitoring, distributed fiber optic sensing, and machine learning.

ORCID

Ruo-Qian Wang  http://orcid.org/0000-0003-2052-8948

Data availability statement

The data that support the findings of this study are openly available in figshare at https://doi.org/10.6084/m9.figshare.19714276.v1.

References

Álvarez, J. M., Gevers, T., LeCun, Y., & López, A. M. (2012). Road scene segmentation from a single image. Berlin, Germany: Springer Science+Business Media. American Society of Civil Engineers, Urban Water Resources Research Council, 1992. Design and construction of urban stormwater management systems.

Arnab, A., Miksik, O., & Torr, P. H. S. (2017). On the robustness of semantic segmentation models to adversarial attacks. New York: Computer Vision and Pattern Recognition.

Arshad, M., Qureshi, M., Inam, O., & Omer, H. (2020). Transfer learning in deep neural network based under-sampled MR image reconstruction. Magnetic Resonance Imaging, 76, 96–107. doi:10.1016/j.mri.2020.09.018

Aschenward, J., Leichter, K., Tasser, E., & Tappeiner, U. (2001). Spatio-Temporal landscape analysis in mountainous terrain by means of small format photography: A methodological approach. IEEE Transactions on Geoscience and Remote Sensing, 39(4), 885–893. doi:10.1109/36.917917

Bai, M., & Urtasun, R. (2017). Deep watershed transform for instance segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii. 5221–5229.
Barrow, H. G., & Tenenbaum, J. M., 1981. Computational vision, in Proceedings of the IEEE, New York. 69, 572–595.

Bayr, U. (2021). Quantifying historical landscape change with repeat photography: An accuracy assessment of geospatial data obtained through monoplotting. International Journal of Geographical Information Science, 35(10), 1–21.

Bayr, U., & Pushmann, O. (2019). Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. Ecological Informatics, 50, 220–233. doi:10.1016/j.ecoinf.2019.01.012

Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 65, 2–16. doi:10.1016/j.isprsjprs.2009.06.004

Chang, L. C., Chang, F. J., Yang, S. N., Kao, I., Ku, Y. Y., Kuo, C. L. . . . Bin Mat, M. Z. (2019). Building an intelligent hydroinformatics integration platform for regional flood inundation warning systems. Water, 11(1), 9. doi:10.3390/w11010009

Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-Supervised learning. IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council, 20(3), 542. doi:10.1109/TNN.2009.2015974

Chaudhary, S., Taran, S., Bajaj, V., & Sengur, A. (2019). Convolutional neural network based approach towards motor imagery tasks EEG signals classification. IEEE Sensors Journal, 19, 4494–4500. doi:10.1109/JSEN.2019.2899645

Chen, W., OuYang, B., & Chou, J. J. (2018). Reply. Nature, 562, 1–20. doi:10.1038/s41586-018-0562-8

Conedera, M., Bozzi, C., Ryter, U., Bertschinger, T., & Krebs, P. (2018). Using the monoplotting technique for documenting and analyzing natural hazard events. In J. Simão Antunes Do Carmo (Ed.), IntechOpen. Natural hazards: risk assessment and vulnerability reduction (pp. 107–108). https://doi.org/10.5772/intechopen.77321

Corripio, J. G. (2004). Snow surface albedo estimation using terrestrial photography. International Journal of Remote Sensing, 25(24), 5705–5729. doi:10.1080/01431160410001709002

Fan, C., Jiang, X., & Mostafavi, A. (2020). A network percolation-based contagion model of flood propagation and recession in urban road networks. Scientific Reports, 10, 13481. doi:10.1038/s41598-020-70524-x

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1627–1645. doi:10.1109/TPAMI.2009.167

Fluehler, M., Niederoest, J., & Akca, D. (2005). Development of an educational software system for the digital monoplotting. In The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences.

Gabellieri, N., & Watkins, C. (2019). Measuring long-term landscape change using historical photographs and the WSL monoplotting tool. Landscape History, 40(1), 93–109. doi:10.1080/01433768.2019.1600946

Galloway, G. E., Reilly, A., Ryoo, S., Riley, A., Haslam, H., Brody, S. . . . Rainey, J. (2018). The growing threat of urban flooding: A national challenge.

Gimmi, U., Ginzler, C., Müller, M., & Psomas, A. (2016). Assessing accuracy of forest cover information on historical maps. Prace Geograficzne, 146, 7–18.

Golparvar, B., & Wang, R.-Q., 2020. AI-Supported citizen science to monitor high-tide flooding in newport beach, california. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities. 66–69.

Golparvar, B., & Wang, R. Q. Al-Supported framework of semi-automatic monoplotting for monocular oblique visual data analysis. 2021. arXiv preprint arXiv:2111.14021.

Grady, L. (2006). Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 1768–1783. doi:10.1109/TPAMI.2006.233

Hariharan, B., Arbeláez, P., Girshick, R., & Malik, J. (2014). Simultaneous detection and segmentation. Columbus, OH, USA: Computer Vision and Pattern Recognition.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Las Vegas, Nevada: Computer Vision and Pattern Recognition.
Hossain, M. D., & Chen, D. (2019). Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 115–134. doi:10.1016/j.isprsjprs.2019.02.009

Huang, W., Cao, Z., Huang, M., Duan, W., Ni, Y., & Yang, W. (2019). A new flash flood warning scheme based on hydrodynamic modelling. *Water*, 11, 1221. doi:10.3390/w11061221

Hu, Y., & Wang, R.-Q. (2020). Understanding the removal of precise geotagging in tweets. *Nature Human Behaviour*, 4(12), 1219–1221. doi:10.1038/s41562-020-00949-x

Jafari, N. H., Li, X., Chen, Q., Le, C. Y., Betzer, L. P., & Liang, Y. (2021). Real-Time water level monitoring using live cameras and computer vision techniques. *Computers & Geosciences*. 147. doi:10.1016/j.cageo.2020.104642

Jasiewicz, J., Niesterowicz, J., & Stepinski, T., 2016. Multi-Resolution, pattern-based segmentation of very large raster datasets. GIScience 2016 Short Paper Proceedings, Berkeley, CA.

Jongman, B., Winsemius, H. C., Aerts, J., Perez, E. C., Aalst, M. K., Kron, W., & Ward, P. J., 2015. Declining vulnerability to river floods and the global benefits of adaptation. Proceedings of the National Academy of Sciences. 112, 2271–2280.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1, 321–331. doi:10.1007/BF00133570

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 60(6).

Kull, C. A. (2005). Historical landscape repeat photography as a tool for land use change research. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, 59(4), 253–268. doi:10.1080/00291950500375443

Lee, C., Yang, H. J. N. H., Nathan, S., & Blair, J. E. A. (2018). Broken arrow: Successful retrieval of a dislodged coronary orbital atherectomy microtip. *Catheterization and Cardiovascular Interventions: Official Journal of the Society for Cardio Angiography & Interventions*, 92(3), 511–523. doi:10.1002/ccd.27674

Liu, J., Kummerow, C. D., & Elsaesser, G. S. (2017). Identifying and analyzing uncertainty structures in the TRMM microwave imager precipitation product. *International Journal of Remote Sensing*, 38 (1), 23–42. doi:10.1080/01431161.2016.1259676

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Boston, MA: Computer Vision and Pattern Recognition.

Makarovic, B. (1973). Digital mono-plotters. *ITC Journal*, 4, 583–599.

Makarovic, B. (1983). Data base updating by digital mono-plott ing. *Environmental Assessment and Resource Management*, 433–442.

McCaffrey, D. R., & Hopkinson, C. (2017). Assessing fractional cover in the alpine treeline ecotone using the WSL monoplotting tool and airborne lidar. *Canadian Journal of Remote Sensing*, 43(5), 504–512. doi:10.1080/07038992.2017.1384309

McCaffrey, D. R., & Hopkinson, C. (2020). Modeling watershed-scale historic change in thealpine treeline ecotone using random forest. *Canadian Journal of Remote Sensing*, 46(6), 715–732. doi:10.1080/07038992.2020.1865792

Mitsitaka, E. A., Machado, A. L., Habib, A. F., & Gonçalves, G. (2004, July). 3D monocular restitution applied to small format digital airphoto and laser scanner data. In Proceedings of Commission III, ISPRS Congress, Istanbul.

Moy de Vitry, M., Schneider, M. Y., Wani, O. F., Man ny, L., Leitão, J. P., & Eggimann, S. (2019). Smart urban water systems: What could possibly go wrong? *Environmental Research Letters*, 14(8). doi:10.1088/1748-9326/ab3761

National Academies of Sciences, Engineering, and Medicine. (2019). In *Framing the Challenge of Urban Flooding in the United States*. Washington, DC: The National Academies Press.

Nemni, E., Bullock, J., Belabbes, S., & Bromley, L. (2020). Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16), 2532. doi:10.3390/rs12162532

Pathak, T., Agrawal, T., Richhariya, S., Sadaf, S., & Hasan, G. (2015). Store-Operated calcium entry through orai is required for transcriptional maturation of the flight circuit in drosophila. *The
Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 35(40), 13784–13799. doi:10.1523/JNEUROSCI.1680-15.2015
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. . . Duchesnay, É. (2011). Scikit-Learn: Machine learning in python. Journal of Machine Learning Research, 12, 2825–2830.
Perone, C. S., Ballester, P., Barros, R. C., & Cohen-Adad, J. (2018). Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. Salt Lake City, Utah: Computer Vision and Pattern Recognition.
Pickard, J. (2002). Assessing vegetation change over a century using repeat photography. Australian Journal of Botany, 50(4), 409–414. doi:10.1071/BT010235
Rajchl, M., Lee, M. C. H., Schrans, F., Davidson, A., Passerat-Palmbach, J., Tarroni, G., and Rueckert, D. (2016). Learning under distributed weak supervision. Las Vegas, Nevada: Computer Vision and Pattern Recognition.
Roerdink, J. B. T. M., & Meijster, A. (2000). The watershed transform: Definitions, algorithms and parallelization strategies. Fundamenta Informaticae, 41, 187–228. doi:10.3233/FI-2000-411207
Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. International Journal of Computer Vision, 77(1), 157–173. doi:10.1007/s11263-007-0090-8
Scapozza, C., Lambiel, C., Bozzini, C., Mari, S., & Coneneda, M. (2014). Assessing the rock glacier kinematics on three different timescales: A case study from the southern swiss alps. Earth Surface Processes and Landforms, 39(15), 2056–2069. doi:10.1002/esp.3599
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Columbus, OH: Computer Vision and Pattern Recognition.
Song, Z., & Tuo, Y. (2021). Automated flood depth estimates from online traffic sign images: Explorations of a convolutional neural network-based method. Sensors, 21(16), 5614. doi:10.3390/s21165614
Stockdale, C. A., Bozzini, C., Macdonald, S. E., & Higgs, E. (2015). Extracting ecological information from oblique angle terrestrial landscape photographs: Performance evaluation of the WSL monoplotting tool. Applied Geography, 63, 315–325. doi:10.1016/j.apgeog.2015.07.012
Svenningsen, S. R., Brandt, J., Christensen, A. A., Dahl, M. C., & Dupont, H. (2015). Historical oblique aerial photographs as a powerful tool for communicating landscape changes. Land Use Policy, 43, 82–95. doi:10.1016/j.landusepol.2014.10.021
Travelletti, J., Delacourt, C., Allemand, P., Malet, J.-P., Schmittbuhl, J., Toussaint, R., & Bastard, M. (2012). Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations. ISPRS Journal of Photogrammetry and Remote Sensing, 70, 39–55. doi:10.1016/j.isprsjprs.2012.03.007
Triglav-Čekada, M., Bric, V., & Zorn, M. (2014). HOW TO DECIDE WHICH OBLIQUE IMAGE HAS THE HIGHEST MAPPING POTENTIAL FOR MONOPLOTTING METHOD: A CASE STUDIES ON RIVER EROSION AND FLOODS. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2(5).
Triglav-Čekada, M., & Radovan, D. (2013). Using volunteered geographical information to map the November 2012 floods in slovenia. Natural Hazards and Earth System Sciences, 13(11), 2753–2762. doi:10.5194/nhess-13-2753-2013
Triglav-Čekada, M., Radovan, D., Gabrovec, M., & Kosmatin-Fras, M. (2011). Acquisition of the 3D boundary of the triglav glacier from archived non-metric panoramic images. The Photogrammetric Record, 26(133), 111–129. doi:10.1111/j.1477-9730.2011.00622.x
Van der Sande, C., Jong, S., & Roo, A. (2003). A segmentation and classification approach of IKONOS2 imagery for land cover mapping to improve flood risk and flood damage assessment. International Journal of Applied Earth Observation and Geoinformation, (3), 217–229. doi:10.1016/S0303-2434(03)00003-5
Vasuki, M., Sharma, P. R., Ibrahim, M., & Arciuli, J. (2017). Musicians’ online performance during auditory and visual statistical learning tasks. Frontiers in Human Neuroscience, 11, 114.
Wang, R.-Q., 2018, April. Big data of urban flooding: Dance with social media, citizen science, and artificial intelligence. In EGU General Assembly Conference Abstracts, Vienna, Austria. (p. 404).
Wang, R.-Q., Hu, Y., Zhou, Z., & Yang, K. (2020). Tracking flooding phase transitions and establishing a passive hotline with AI-enabled social media data. *IEEE Access*, 8, 103395–103404. doi:10.1109/ACCESS.2020.2994187

Wang, R.-Q., Mao, H., Wang, Y., Rae, C., & Shaw, W. (2018). Hyper-Resolution monitoring of urban flooding with social media and crowdsourcing data. *Computers & Geosciences*, 111, 139–147. doi:10.1016/j.cageo.2017.11.008

Wiesmann, S., Steiner, L., Pozzi, M., Bozzini, C., Bauder, A., & Hurni, L. (2012, September). Reconstructing historic glacier states based on terrestrial oblique photographs. In Proceedings of the AutoCarto International Symposium on Automated Cartography, Columbus, OH, USA (Vol. 1618).

Yuan, Y., Bazzett, D., Padnani, R., & Wang, R.-Q. (2021). 2020. unlocking data from the online footage of the edenville dam failure. In *Geo-Extreme*. Atlanta, Georgia.

Zhou, Z., & Guo, Q. (2022). Drainage alternatives for rain gardens on subsoil of low permeability: Balance among ponding time, soil moisture, and runoff reduction. *Journal of Sustainable Water in the Built Environment*, 8(3), 05022002. doi:10.1061/JSWBAY.0000988

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1–130.