Human-guided Flood Mapping on Satellite Images

Jiongqian Liang, Peter Jacobs and Srinivasan Parthasarathy

The Ohio State University

Introduction

- Flooding causes substantial loss of life and economy.
- **Flood mapping**: the process of distinguishing flooded areas from non-flooded areas during and after a disaster.
- Flood mapping is very useful:
  - Guiding first responders in a disaster situation.
  - Assessing flood risk in future disaster scenarios.
- **Goal**: mark out flooded areas in satellite images of a region during, and after a flood.
- **Data source**:
  - Satellite images (optical, Radar, etc.)
  - Domain expert knowledge.
  - Other information, such as social media.
- We break down flood mapping into
  - Marking out water and land area
  - Dynamic evolution analysis (see Figure 3).
- **Challenges**:
  - Existing work on image segmentation usually focuses on segmenting one or a few objects out of an image.
  - A typical example is the foreground and background.
  - Flood mapping involves segmenting many areas with water, which can be of different sizes.
  - We require the method to be scalable on high-resolution satellite images (10M+ pixels).

Methodology: Human-Guided Flood Mapping (HUG-FM)

- **HUG-FM** -- Interactive and incremental learning
  - We ask users to **label areas** of an image of a geographical region as either water or land.
  - Adopt these labels to train a classifier and use it to label the rest of the image.
  - Return the flood mapping results in **real time**.
  - Users can decide to adjust the results through an iterative process.
- **HUG-FM** contains two steps:
  - Step 1: **Graph-based Clustering** to divide the image into patches; similar pixels are grouped together.
  - Step 2: **Patch Classification** based on supervision from humans.

Graph-based Clustering

- Formulate the image as a graph and use graph-based clustering approach to generate patches.
- Each pixel in the image is a node. For two nodes i and j, the weight of edge (i, j) is defined as follows:

$$w_{ij} = \begin{cases} e^{-\frac{d(i,j)^2}{\sigma_1^2} - \frac{|F(i)-F(j)|^2}{\sigma_2^2}} & \text{if } d(i,j) < d_{max} \\ 0 & \text{otherwise.} \end{cases}$$

- $d(i,j)$ = Euclidean distance between node i and j.
- $F(i)$ refers to the feature vector of node i (e.g., intensity in gray scale image).
- δ_1, δ_2 and d_{max} are parameters tuned by the user.
- Multi-level Regularized Markov Clustering (MLR-MCL) [3].

Patches Classification

- **Classifier**: k–NN.
- Define a distance function between patch i and patch j:

$$D(i,j) = \left\|\bar{F}(i) - \bar{F}(j)\right\|_2 * \log\left(dist(i,j)\right)$$

  - $\bar{F}(i)$: The feature vector of patch i (e.g., average intensity).
  - $dist(i,j)$: the Euclidean distance between patch i and patch j.
- **Intuition**: If two patches are similar in intensity and close in location, their distance $D(i,j)$ should be small.
- **Classification**: voting by k nearest (most similar) labeled patches.
- **Flood mapping**: compare classification results on post-flood satellite images with pre-flood satellite images to determine the flood areas.

Scalability Analysis

Method	# Markers	Labeling Time (s)
Our Approach (HUG-FM)	2	0.057
Watershed Algorithm	11	0.225
N-cuts Algorithm	0	538.615
Graph method w. post-process	0	558.220

- The results are on the down-scaled satellite image 800×400 (the last two cannot finish in an hour on the original image).
- HUG-FM is much more efficient, which enables real-time and interactive labeling.
- Efficient flood mapping can enable faster response and reduce loss of life and economy.

Experiments and Analysis

Dataset

- Satellite images of Chennai area during the 2015 South Indian Floods (11/08/2015—12/14/2015).
- One Satellite image for every 12 days starting from 10/19/2015.
- Satellite image size: 4500×2500 pixels. Example:

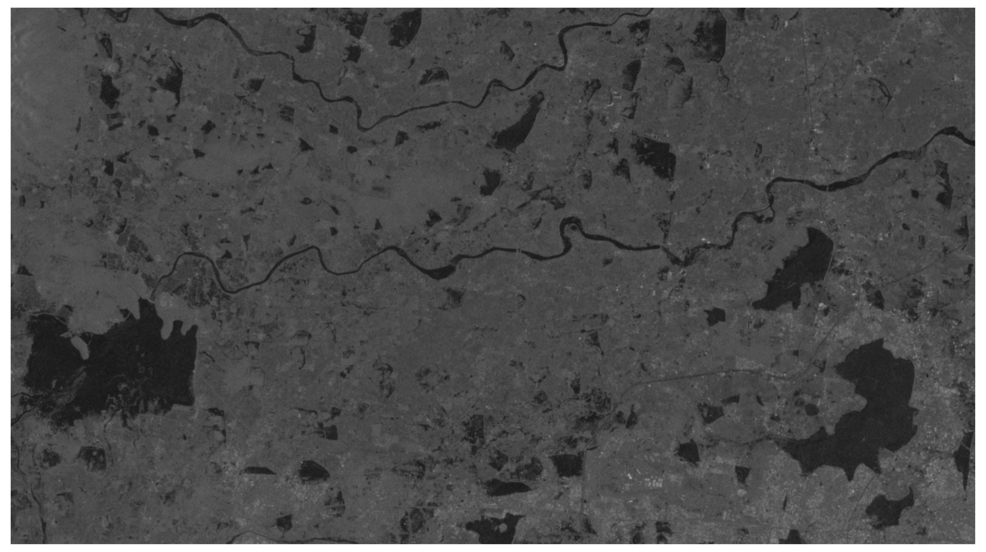
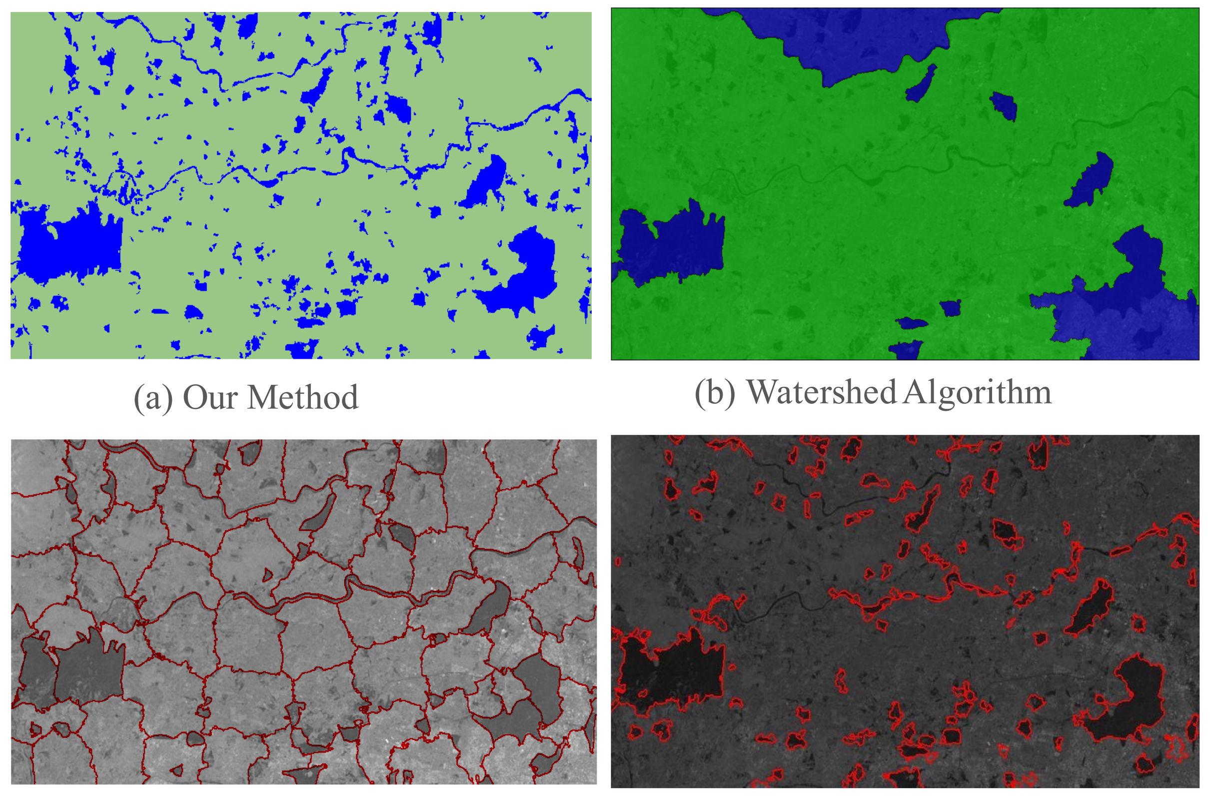


Figure 1. Satellite Image of Chennai area on 11/24/2015

Baselines

- **N-cut** (*Shi et al., 2000 [1]*): graph-based method to minimize the normalized cut.
- **Watershed Algorithm** (*Beucher et al., 1992 [2]*): region growing technique.
- **Graph-based Algorithm with post-processing**: merge similar patches without supervision.

Results

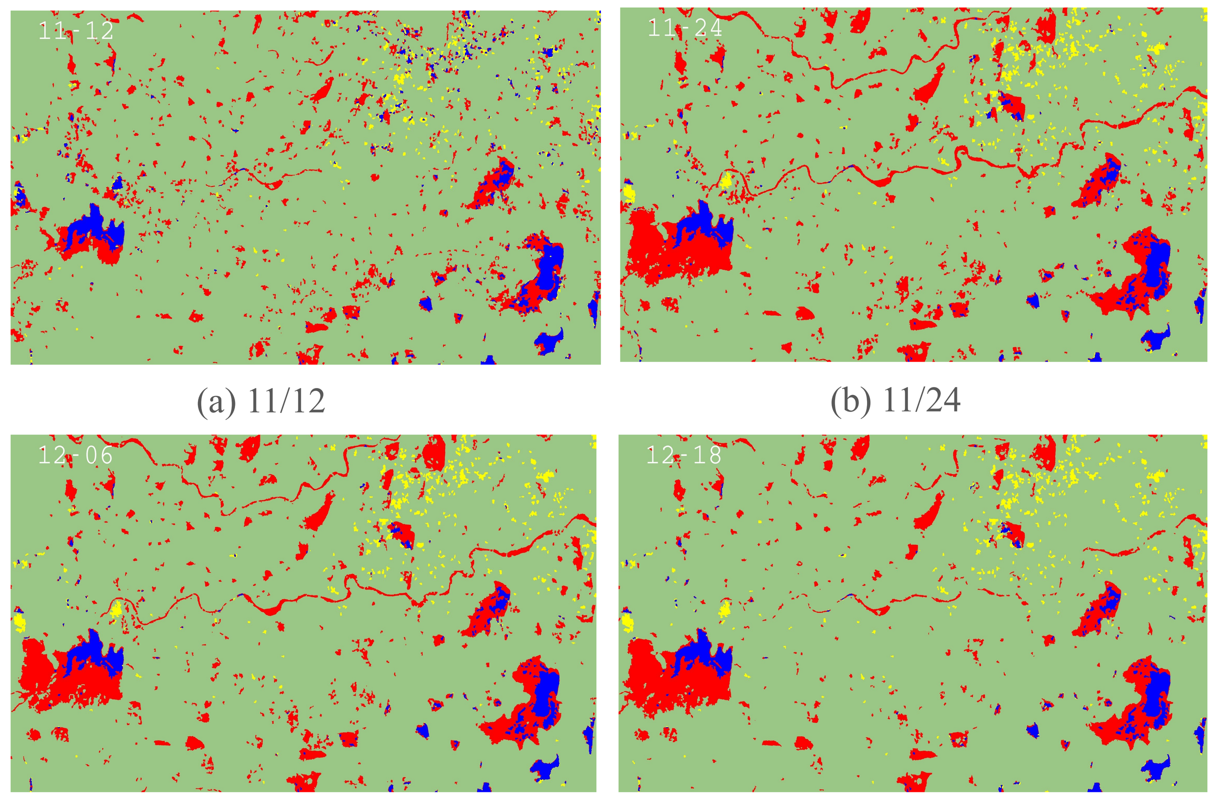


(a) Our Method (b) Watershed Algorithm

(c) Normalized Cuts Algorithm (100 partitions) (d) Graph-based Clustering with Post-processing (100 partitions)

Figure 2. Flood Mapping of Satellite Image on 11/24/2015

- Watershed algorithm, and the graph-based algorithm with post-processing are unable to identify small water regions.
- The N-cut method splits land regions and water regions into multiple segment.
- HUG-FM successfully identifies small water regions, large water regions, rivers, and large land regions.
- HUG-FM is more efficient (real-time, interactive labeling).



(a) 11/12 (b) 11/24

(c) 12/06 (d) 12/18

Figure 3. Flood Mapping Dynamic Analysis.

- Each of the above images is the result of comparing the HUG-FM flood map of the given date with the one on 10/25/2015 (before the flood)
- **Red** shows area that was land on 10/25, but is now water. **Yellow** indicates the opposite case. **Blue** shows area that was water on 10/25, and is still water.
- HUG-FM provides a clear view of the progression of the flood.

Future Work

- Improve the semi-supervised learning method.
  - Adopt more advanced machine learning algorithms: Support Vector Machine (SVM) and Neural Network.
  - Enhance the selection of features, e.g., Scale-invariant feature transform (SIFT).
- Encourage users to label patches leading to maximal information gain.
  - k–NN: ask users to label patches far away from currently labeled patches. SVM: ask to label patches close to the hyper-plane margin.
- Launch crowd-sourcing experiments on Amazon Mechanical Turk.
  - Employ multiple users to label the same image.
  - Adjust the number of employee for one image based on the difficulty, and aggregate the results.
- Incorporate social media information.
  - Relevant posts (e.g., Tweets) on the flood with geo-tag can provide supervision for HUG-FM.

Reference

[1] J. Shi and J. Malik. Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8):888–905, 2000.

[2] Beucher, Serge, and Fernand Meyer. "The morphological approach to segmentation: the watershed transformation. " *OPTICAL ENGINEERING-NEW YORK-MARCEL DEKKER INCORPORATED-* 34 (1992): 433-433.

[3] Satuluri, Venu, and Srinivasan Parthasarathy. "Scalable graph clustering using stochastic flows: applications to community discovery." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 737-746. ACM, 2009.

Acknowledge

- This work is supported by NSF Award NSF-EAR-1520870 and NSF-DMS-1418265. We also thank Desheng Liu and Jiayong Liang for useful discussions and help with data collection.