IEEE Big Data 2015, October 31, 2015

Crowdsourcing the Acquisition and Analysis of Mobile Videos for Disaster Response

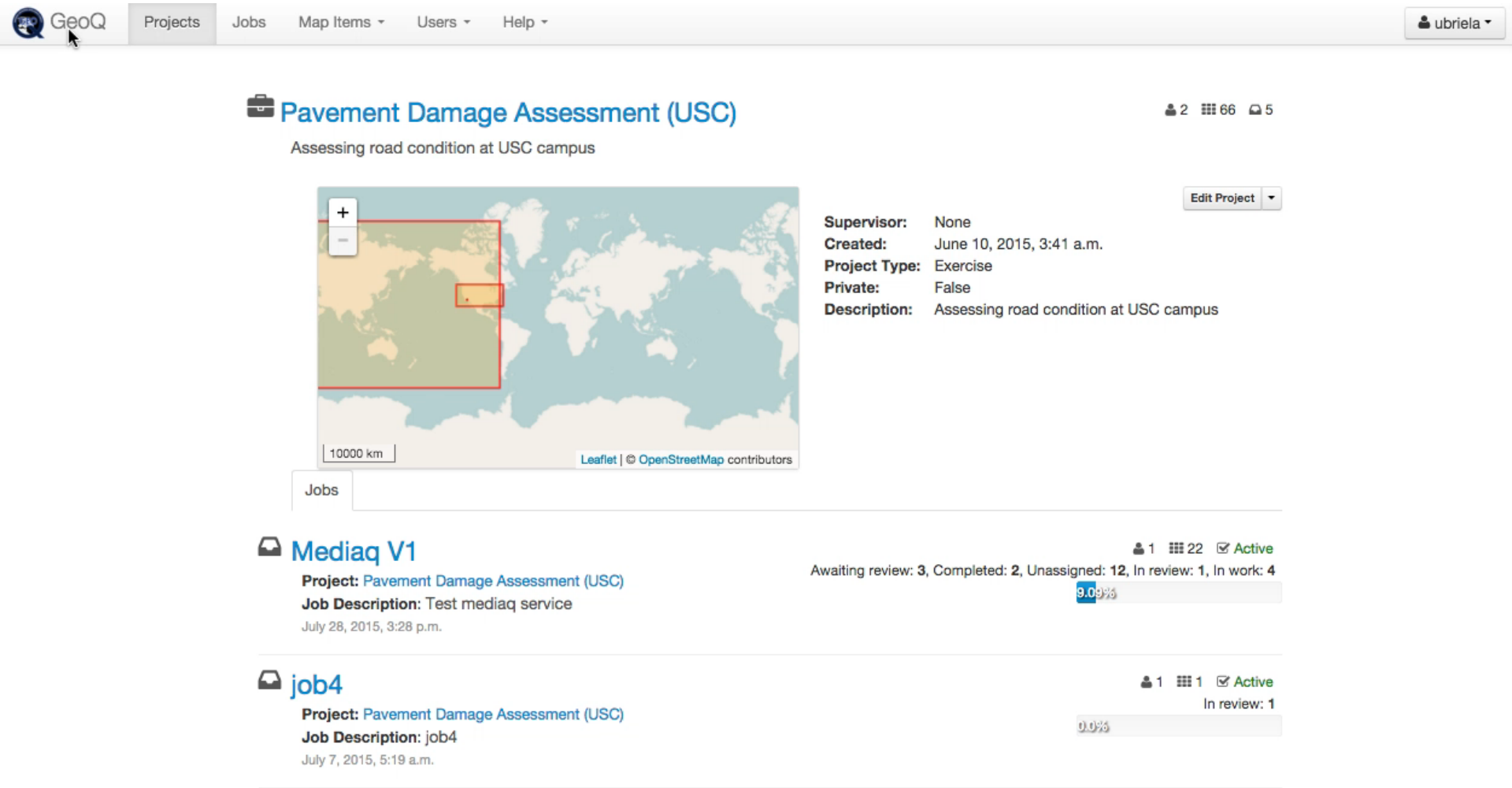
*Presented by **Hien To***

Dr. Seon Ho Kim

Prof. Cyrus Shahabi

Integrated Media Systems Center
University of Southern California

GeoQ by National Geospatial-Intelligence Agency (NGA)



Motivation

- ✓ Obtain real-time information to enhance **situational awareness** during and after **disasters**
- ✓ Live streaming of **mobile video**, e.g., YouTube, Periscope
- ✓ **Prioritize data collection** with limited transmission
- ✓ **Analyze collected data** in timely manner

Disasters

Assess damage across **large geographical area**

Haiti Earthquake

Cable Cuts in Vietnam

Disrupted communication networks or data overload!

Haiti earthquake, 85% Haitians could still access to their mobile phones but 70% cell towers were destroyed

Crowdsourcing Disaster Response

Early disaster data collection focus on VGI	*[Goodchild et al. 2010]*
Disaster surveillance and response system	*[Chu et al. 2012]*
Processing and integration of disaster data by leveraging Linked Open Data	*[Ortmann et al. 2011]*
Combine human and machine intelligence	*[Schulz et al. 2012]*
Infrastructure damage caused by Earthquake	*[Kobayashi 2014]*
Resilient communication technologies, e.g., drones	*[Nemoto and Hamaguchi 2014]*

Our study:

- ✓ Focuses on both data collection and data analysis
- ✓ Collects huge amount of videos concerning disasters
- ✓ Improves resilience and responsiveness of computer systems

Crisis Response Efforts

Ushahidi

USGS ShakeMap
Did You Feel It?

AI for DR

USC MediaQ

NGA GeoQ

Skybox Imaging

esri Disaster Response

Google Crisis Response

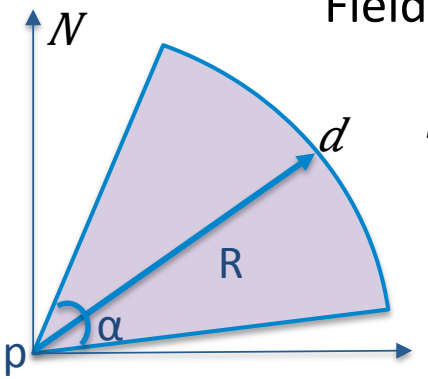
Planetary Response Network

Highlight GeoQ and MediaQ

MediaQ and GeoQ

Data Collection

MediaQ [1]: mobile crowdsourcing platform to collect **videos/images**



Field Of View (FOV)

FOV Metadata:

p: camera location

d: camera direction

α: viewable angle

R: viewable distance

t: timestamp

Data Analysis

GeoQ [2]: crowdsourcing platform to analyze disaster data

Work cells are assigned to **analysts**

GeoQ's Work Cells used in Nepal Earthquake 2015

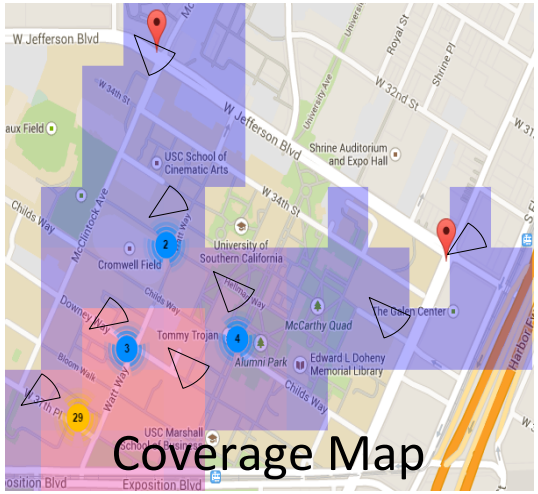
- ✓ MediaQ provides visual data to GeoQ.
- ✓ MediaQ workers are **on-site people**; GeoQ workers are **off-site analysts.**

[1] By *USC Integrated Media System Center*

[2] By *US National Geospatial-Intelligence Agency* (NGA)

Proposed Techniques

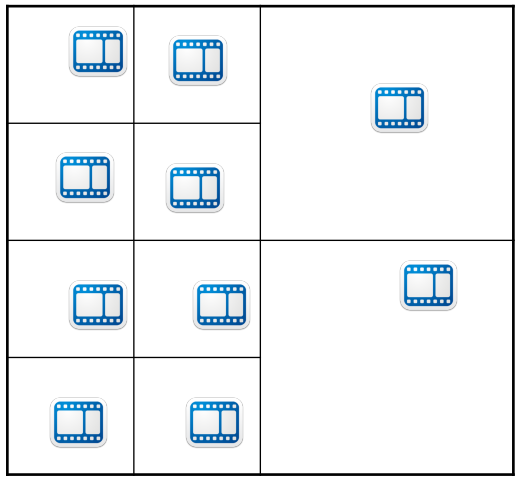
Collect data fast because time is critical



Coverage Map

Video metadata

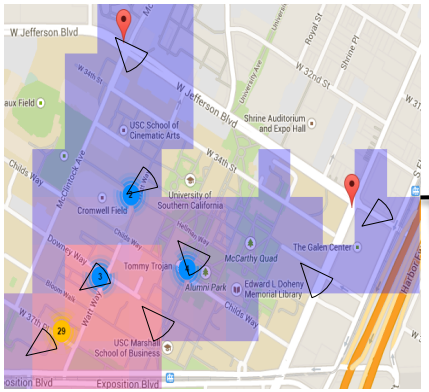
Analyze the collected data effectively



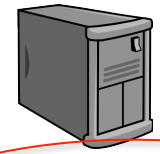
Video content

- ✓ Upload **metadata (FOV) first** then identify videos with high **visual awareness** to be uploaded
- ✓ Partition space based on video locations

Proposed Framework



Coverage Map

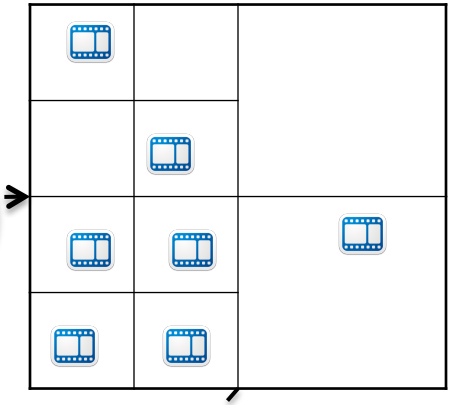


Server

2. Upload relevant videos

3. Assign videos to analysts

Space Decomposition



Analyst

1. Crowdsource videos from community

4. Evaluate work cells

0	0	0
1	3	
2	5	2
0	0	

Urgency Map

Video metadata

Video content

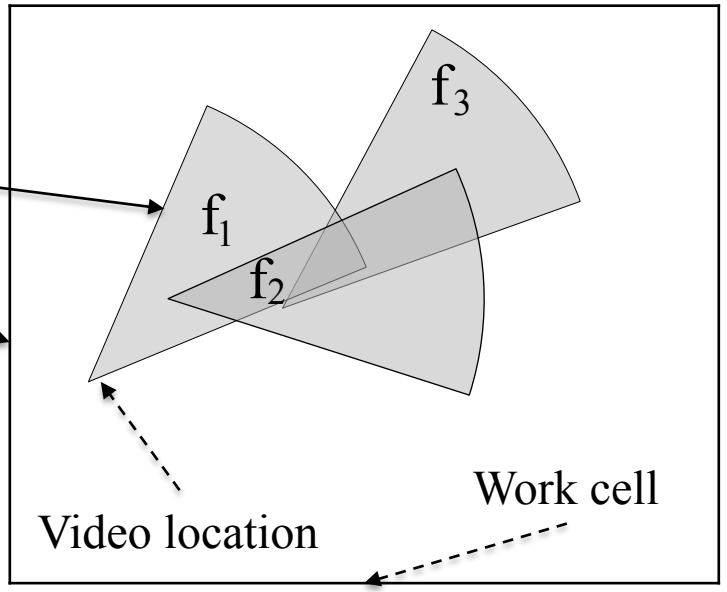
1. Acquisition of Video Metadata and Content

- ✓ Hold video in the device and have server ping the device when it wants the owner to upload the video
- ✓ Real-time acquisition of metadata via Internet, SMS, WLAN
- ✓ Size of metadata is thousands times smaller than content
- ✓ Real-time analysis on the uploaded metadata, e.g., visualization
- ✓ Support data governance, preserve privacy and strict access control *[Dey and Chinchwadkar, ICDE 2015]*

2. Visual Awareness Maximization

Visual Awareness (VA) of a Video

$$VA(v) = Urgency(w)\frac{area(v)}{area(w)}$$



3-seconds video within a work cell

A video has high VA if

- ✓ Enclosing work cell is urgent
- ✓ Aggregate video coverage is large

Selects **videos** that maximize total VA without exceeding budget (i.e., constrained bandwidth)

This problem is 0-1 Knapsack.

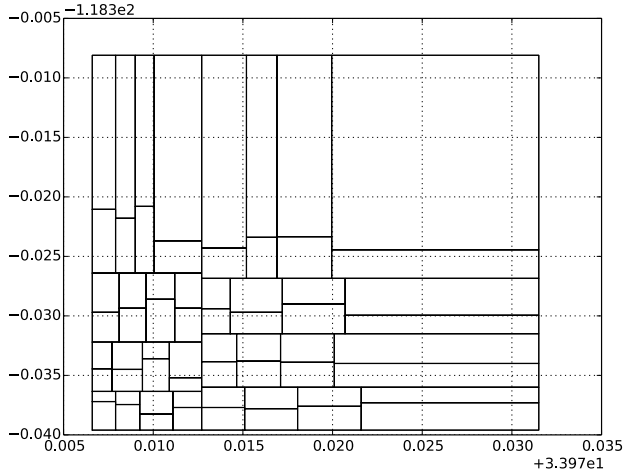
3. Assign Videos to Analysts

GeoQ uses of grid-based partition

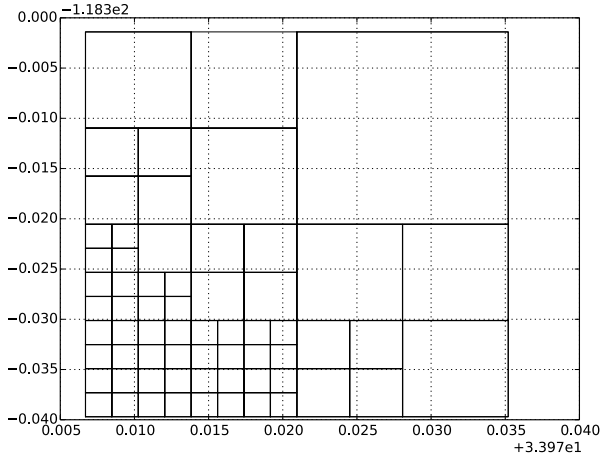
- ✓ Some analysts may be overloaded ☹

Space partitioning algorithm based on Quadtree and Kd-tree

- ✓ Each analyst handles equal number of videos ☺
- ✓ Terminates when #workcells > #analysts-3



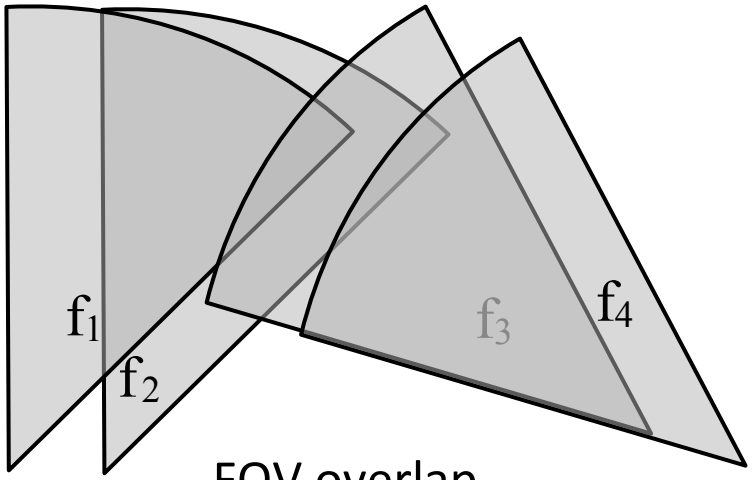
Kd-tree (Gaussian dist.)



Quadtree (Gaussian dist.)

Minimum Redundant Coverage

Upload **key frames** and metadata to server and **minimize redundant coverage**



FOV overlap

Select **frames** that maximizes total Visual Awareness without exceeding budget

This problem is Maximum Coverage Problem.

Simulation Results

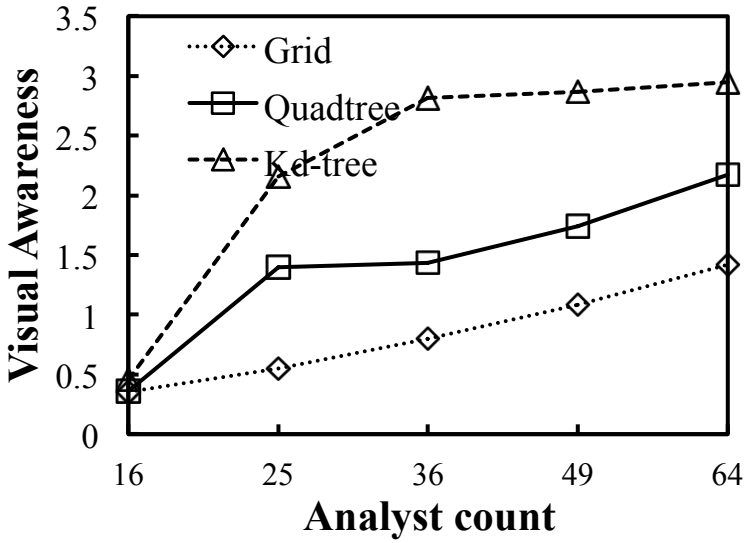
1. Partitioning techniques: Grid, Quadtree, Kd-tree
2. Visual awareness at video level vs. frame level

Synthetic Datasets *[Ay et. Al. SIGSPATIAL'2010]*

Statistics	Value
Spatial distributions	Uniform, Gaussian and Zipfian
#videos	1000 with varying length
Region size	10 × 10 square km in Los Angeles area
Unit grid cells	500×500
FOV # per second	1
Viewable angle	60 degrees
Visible distance	200 meters

Visual Awareness Maximization at **Video Level**

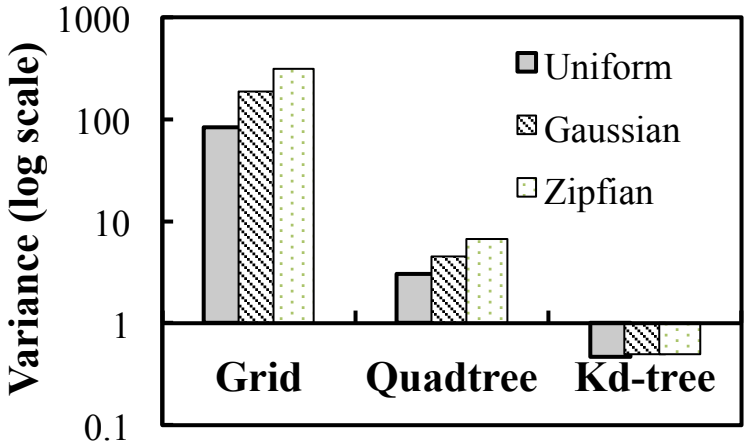
Varying the number of analysts



Uniform Dataset

Kd-tree increases visual awareness by up to 300%

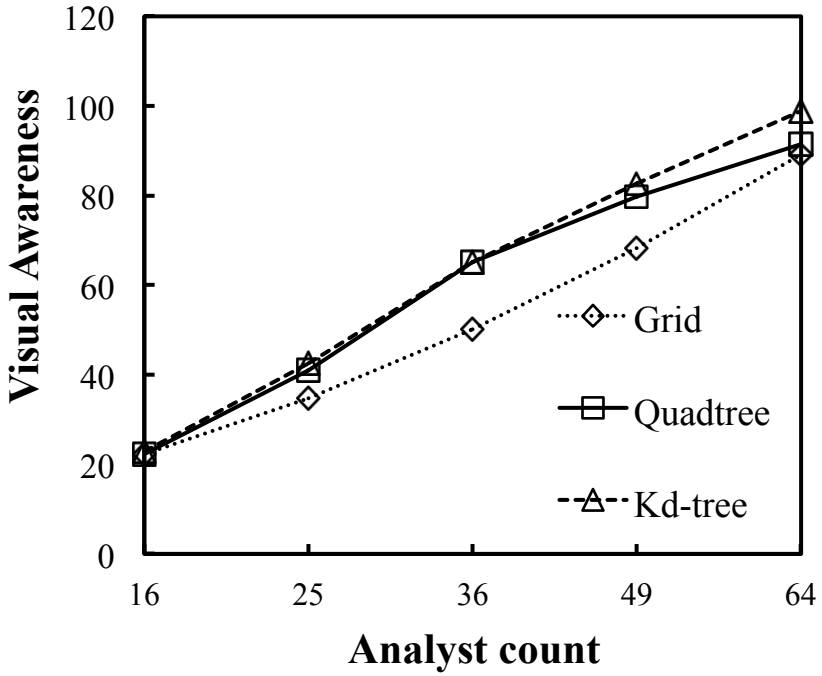
Variance of #videos per analyst



Gaussian Dataset

Kd-tree produces similar #videos per work cell

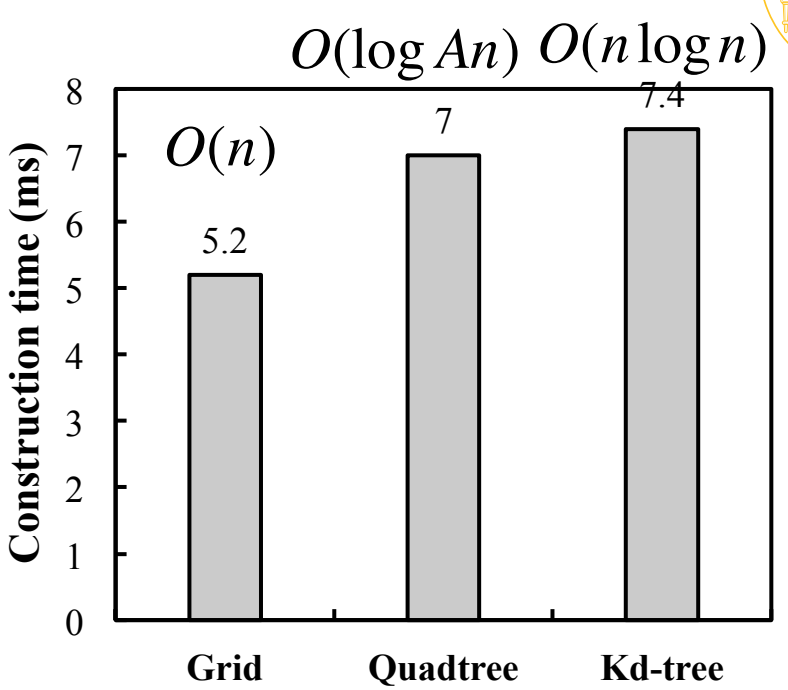
Frame Level



Uniform Dataset

Obtained visual awareness is an order of magnitude higher in comparison to video-level opt

Runtime Measurements



Uniform Dataset

Construction times are small and the differences are insignificant

Conclusion

- ✓ Proposed crowdsourcing framework for collection and analysis of video data under disaster situations
- ✓ Introduced a mechanism to enable time-sensitive acquisition and analysis of spatial metadata
- ✓ Developed an analytical model to quantify the visual awareness of a particular video or frame
- ✓ Introduced two variants visual awareness maximization problem: uploading the entire videos and individual frames
- ✓ Introduced more effective spatial decomposition to assign videos to analysts

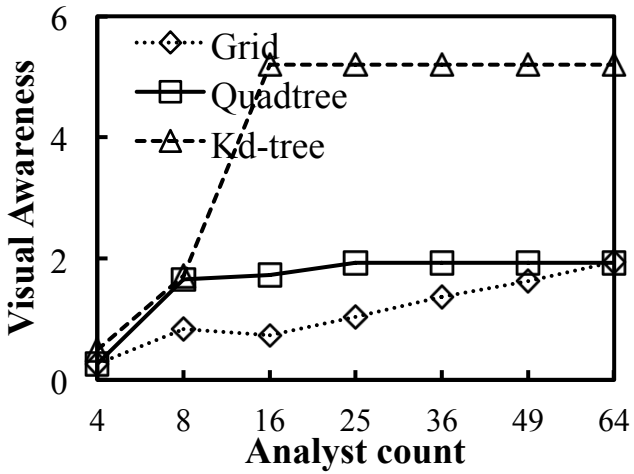
Q/A

Hien To

University of Southern California

hto@usc.edu

MediaQ Dataset



Location/Region Entropy

Location entropy (LE)

- Measures the *diversity* of unique visitors of a *location*
- High if *many* users were observed at the location with *equal* proportion

	u_1	u_2	LE
l_1	1000	1	**0.35**
l_2	500	500	**0.69**

$$LE(l) = \sum_{u \in U_l} P_l(u) \log P_l(u)$$

where $P_l(u) = \frac{|O_{u,l}|}{|O_l|}$

$Freq(l) = |O_l|$: total number of visits

$UserCount(l) = |O_{u,l}|$: number of visits by worker u

Cranshaw, J., et al., (2010). *Bridging the gap between physical locations and online social networks.* UBICOMP, 119-128

Work Cell Urgency

Compute **urgency value** of a work cell

$$Urgency(w) = \text{Priority}(w)\alpha + \text{RegionEntropy}(w.r)\beta$$

The priority of an area, e.g., nuclear plant areas are more important than residence areas.

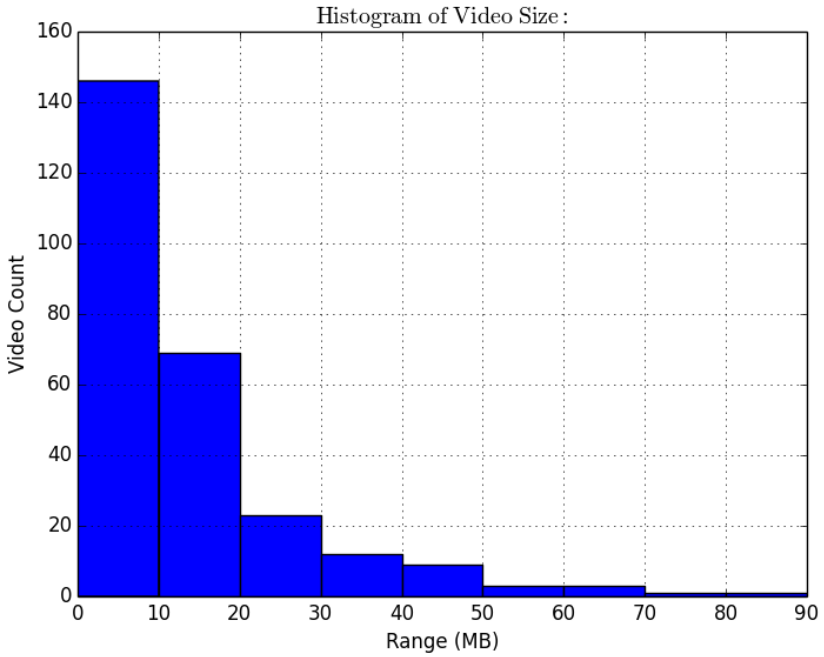
Region Entropy – **geosocial factor** of a region. RE is high if the region are visited by many people many times, e.g., school, hospital.

Nuclear plant

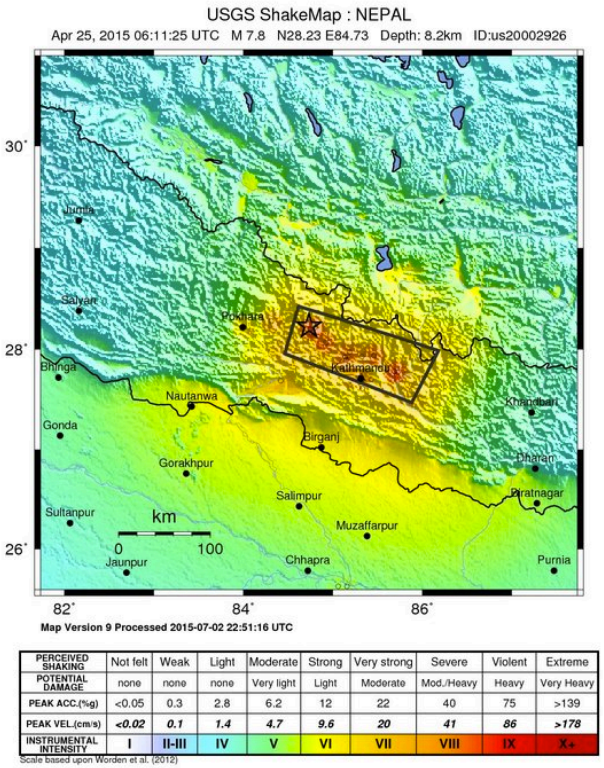
USC Campus

Future Work

1. Use user-generated videos from MediaQ



2. Use near-real-time ShakeMap from USGS



Data Analysis

Work Cell Analysis

- ✓ Analysts measure the severity of their assigned work cells and set urgency values to them
- ✓ Urgency of the work cells can be automatically calculated