



Selecting User-Generated Videos for Augmented Reality Applications

IEEE BigMM 2016

Hien To, Hyerim Park,
Seon Ho Kim, Cyrus Shahabi.

Integrated Media Systems Center
University of Southern California



Motivation

- ✓ Popularity of augmented-reality (AR)
 - AR browsers, e.g., Layar, Wikitude, and Junaio
- ✓ **Scarcity** of AR content
 - Difficult to create content in AR browsers
 - Adopt multimedia content from social media
- ✓ Lack of **user contexts** causes imprecise registration

Need content source with rich context information



Motivation (Cont...)

Mobile cameras are everywhere!



Any way to use mobile
videos as AR content?





Motivation (Cont...)

- Current most important keywords in video market are as follows: (by **Kleiner Perkins Caufield & Byers (KPCB), 2015**)
 - User-generated
 - Tagged
 - Curated
 - Indexed & Searchable



Challenges

Known issues of mobile videos



1. Out of Control



2. Hard to Search



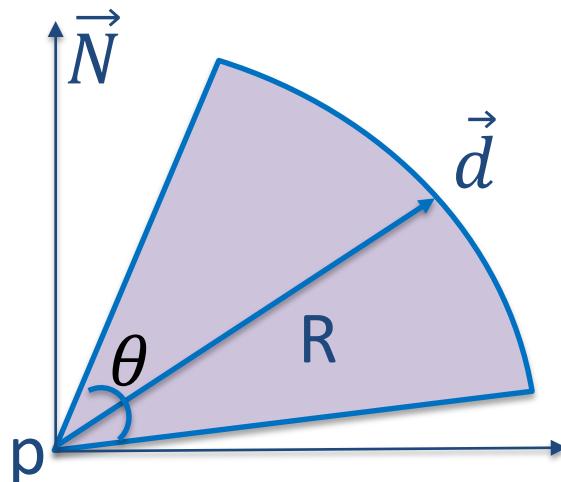
3. Hardly Systematic

Hard to search for interesting videos for AR content



Related Work

- ✓ Content-based retrieval
 - Extensive computation required, still inaccurate
- ✓ Spatial metadata (camera location, orientation)
 - Geo-tagging required, sensors on mobile device



Field Of View (FOV)

p : camera location
 d : camera direction
 θ : viewable angle
 R : viewable distance
 t : timestamp



MediaQ: mobile media management system using spatial metadata from sensors (<http://mediaq.usc.edu>)



Geo-tagged Mobile Videos

✓ Selecting videos for a certain area is now fast!

The mediaQ web interface displays a map of the USC campus. A video thumbnail is shown in the top right corner, with the timestamp "2014-02-24 13:26:22". On the left, there are three sections: "Region" (which is visible), "Range" (where you place the circle), and "Point" (where you place the marker). The map shows various buildings and streets, with several red location markers and blue circular overlays indicating video coverage areas.

The mediaQ mobile application interface shows a navigation bar with "mediaQ", "Search", "Groups", and "USC Server". Below is a section titled "Region which is visible" and a "Range where you place the circle". The main area displays a map of the USC campus with red location markers and blue circular overlays. To the right, there are sections for "Tasks Inquiries", "Pending Tasks", "Collection", "Settings", and "About". At the bottom, a "Video List" is displayed with a table:

Video ID	File Name	Last Modified	Count
2014_12_8_Videotake_14180811802	2014_12_8_Videotake_14180811802.mp4	04	2
2015_1_17_Videotake_14215370788	2015_1_17_Videotake_14215370788.mp4	22	1
2015_1_17_Videotake_14215378197	2015_1_17_Videotake_14215378197.mp4	39	1
2015_1_20_Videotake_14218205940	2015_1_20_Videotake_14218205940.mp4	78	1
2015_2_12_Videotake_14238059193	2015_2_12_Videotake_14238059193.mp4	40	1
2015_2_2_Videotake_142293791617	2015_2_2_Videotake_142293791617.mp4	9	1
2015_2_3_Videotake_142300056960	2015_2_3_Videotake_142300056960.mp4	7	1
2015_2_3_Videotake_142300232931	2015_2_3_Videotake_142300232931.mp4	2	1
2015_2_9_Videotake_142353787555	2015_2_9_Videotake_142353787555.mp4	2	1

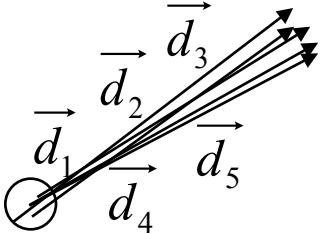
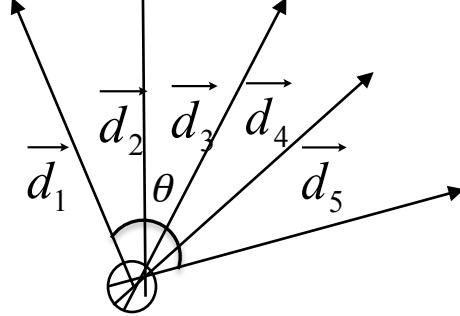
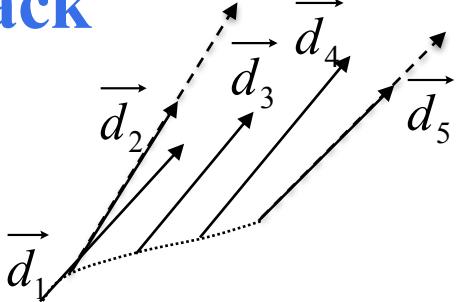
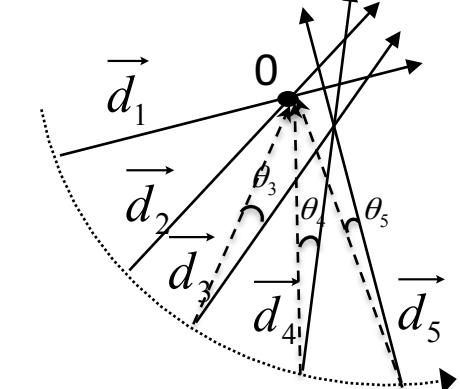


Problem Statement

- ✓ Simple geospatial relevance is not enough.
For example, selecting “**interesting**” videos for AR applications from many at a location?
- ✓ Identify interesting (or significant) videos to AR users as a sequence of FOVs that follow a **particular camera shooting pattern** used in filming such as *tracking*, *zooming*, *panning*, and *arching* scenes

Modeling Significant Video Segments



Position Position	Direction Single	Direction Multiple
Single	$\langle t_{min}, d_{max}, r_{max} \rangle$ Zoom 	$\langle t_{min}, d_{min}, r_{max} \rangle$ Pan 
Multiple	$\langle t_{min}, d_{max} \rangle$ Track 	$\langle t_{min}, d_{max} \rangle$ Arch 



Experiment Using Real Datasets

Statistics	Value
Number of videos with geo-metadata	2,397
Average length per video with content (sec)	72.14
Average camera moving speed (km/h)	4.5
Average camera rotation speed (degrees/sec)	10
Number of users	289
Number of videos by each user	8.29
Number of FOVs	208,978
Number of FOV per second	1.03
Number of FOV per video	74.16

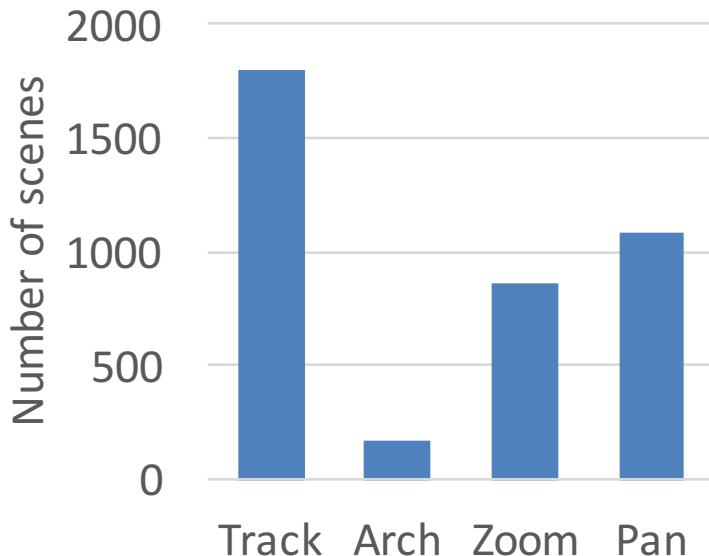
[ACM MMSys 2016 Dataset Paper, USC MediaQ Mobile Video Data]



Significant Video Segments

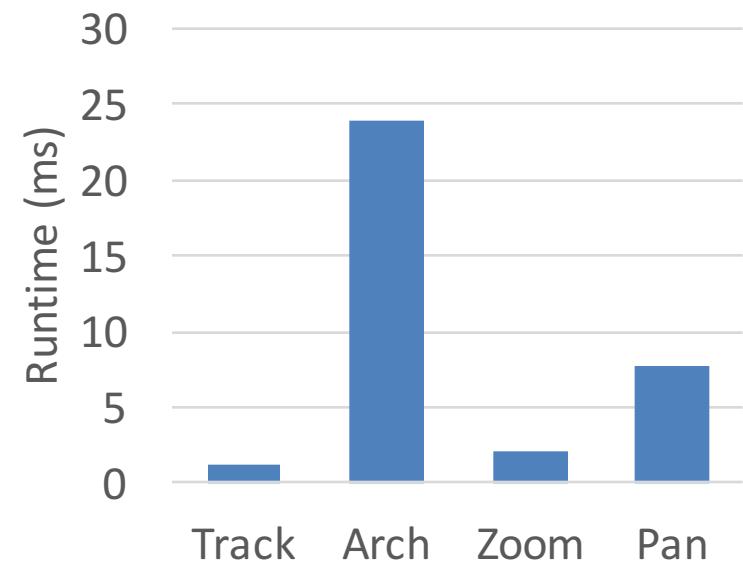
$$t_{min} = 15, d_{max} = 15, d_{min} = 120$$

Number of scenes found



Tracking is the most popular way of capturing mobile videos

Searching time (one video)



Overall search time is fast

Zooming Scene



Panning Scene



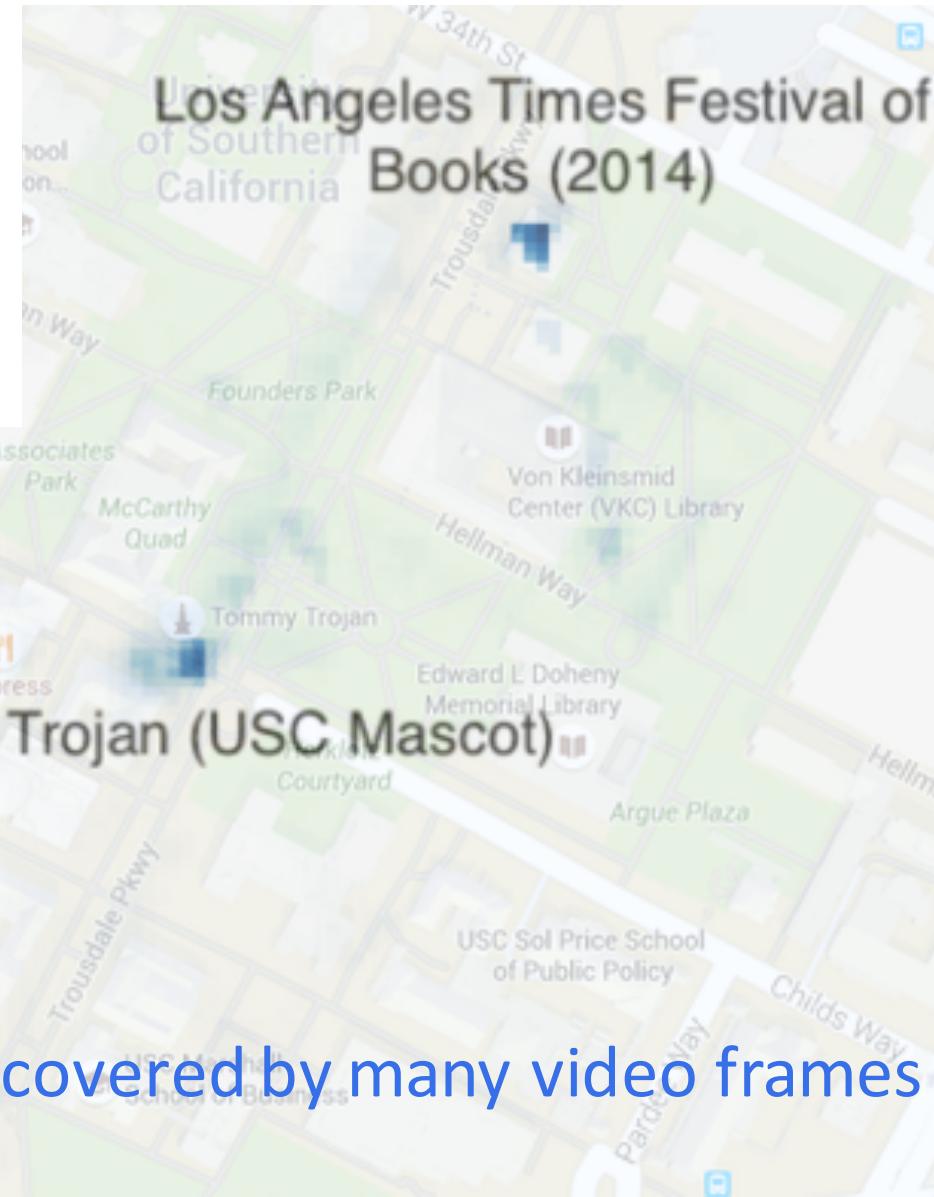
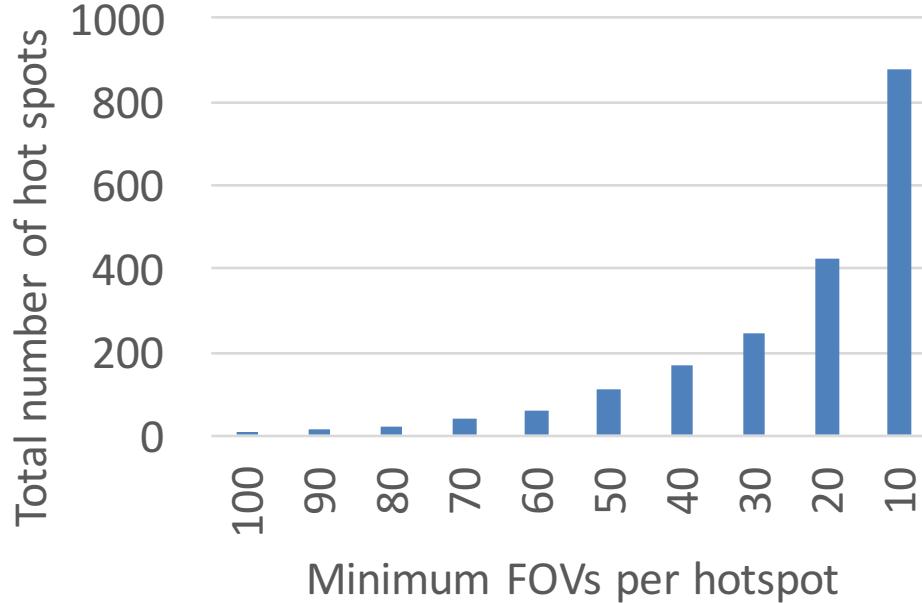
Tracking Scene



Arching Scene



Detecting Hot Spots



PHE (Mediaq -
meeting room)

Hotspots covered by many video frames



Conclusion and Future Work

- ✓ Proposed models and algorithms to identify **interesting video segments** and **hotspots**
- ✓ Experiments on real geo-tagged video data show that the algorithms are fast and able to identify interesting scenes
- ✓ Will consider user mobility and ranking of video segments



Thanks

Seon Ho Kim, Ph.D.
University of Southern California
seonkim@usc.edu