Social Street View: Blending Immersive Street Views with Geo-tagged Social Media

Ruofei Du and Amitabh Varshney
The Augmentarium| UMIACS | University of Maryland, College Park
{ruofei, varshney} @ cs.umd.edu
www.SocialStreetView.com

UNIVERSITY OF MARYLAND

UMIACS

Introduction

Social Media

image courtesy: plannedparenthood.org

Introduction

Social Media - Interactivity

image courtesy: adweek.com

Introduction

Social Media - Versatility

image courtesy: forbes.com

Introduction

Social Media - Social Relevance

image courtesy: huffingtonpost.com


METADATA

image courtesy: B&T Magazine

Introduction

Meta data - Time of Creation

image courtesy: Brian Clegg

Introduction

Meta data - Location of Creation

image courtesy: squarespace.com

Introduction

Meta data - Camera Parameters

Image courtesy: naldzgraphics.net

Related Work

Linear narrative visualization

image courtesy: instagram.com, facebook.com, twitter.com

Related Work

2D layout

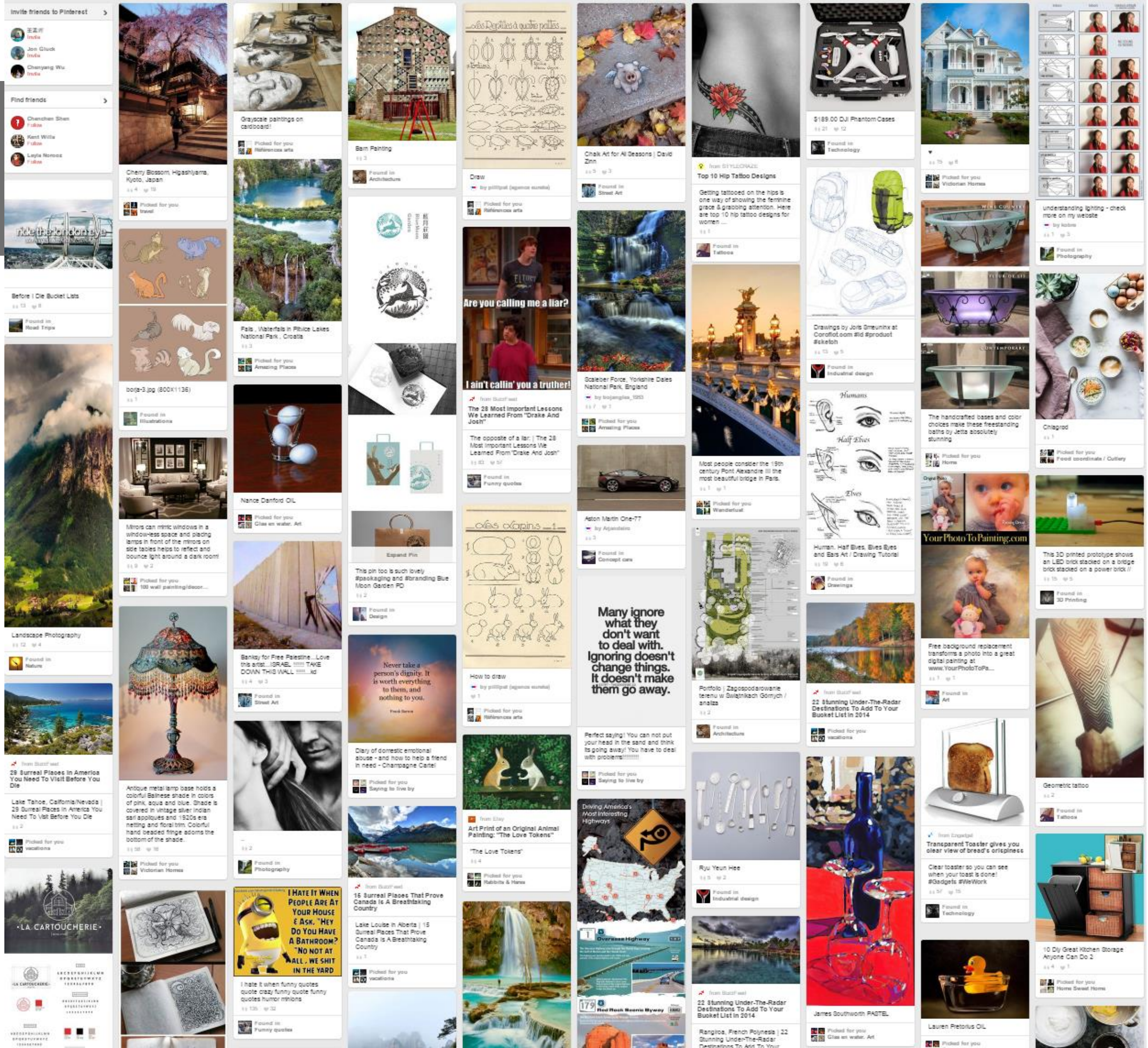


image courtesy: pinterest.com

Related Work

Natural Immersive Virtual Reality?

image courtesy: drivenhealthy.com

Related Work

Karnath et al. and *Loomis et al.*

Related Work

Karnath et al. and *Loomis et al.*

Spatial awareness is a function of the temporal not the posterior parietal lobe

Hans-Otto Karnath, Susanne Ferber & Marc Himmelbach

Department of Cognitive Neurology, University of Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Our current understanding of spatial behaviour and parietal lobe function is largely based on the belief that spatial neglect in humans (a lack of awareness of space on the side of the body contralateral to a brain injury) is typically associated with lesions of the posterior parietal lobe. However, in monkeys, this disorder is observed after lesions of the superior temporal cortex[1], a puzzling discrepancy between the species. Here we show that, contrary to the widely accepted view, the superior temporal cortex is the neural substrate of spatial neglect in humans, as it is in monkeys. Unlike the monkey brain, spatial awareness in humans is a function largely confined to the right superior temporal cortex, a location topographically reminiscent of that for language on the left[2]. Hence, the decisive phylogenetic transition from monkey to human brain seems to be a restriction of a formerly bilateral function to the right side, rather than a shift from the temporal to the parietal lobe. One may speculate that this lateralization of spatial awareness parallels the emergence of an elaborate representation for language on the left side.

Spatial neglect is a characteristic failure to explore the side of space contralateral to a brain lesion. Patients with this disorder behave as if one side of the surrounding space had ceased to exist. Since the early post-mortem studies, we have believed that, in humans, lesions located predominantly in the posterior parietal lobe are critical for this disorder. Analyses of computerized tomography scans of right-hemispheric stroke patients with neglect found that superimposed lateral projections of these lesions centred on the inferior parietal lobule (IPL)[3,4] and the temporo-parieto-occipital (TPO) junction[4]. More recent studies have confirmed the validity of this conclusion although evidence for additional pathology leading

...illan Magazines Ltd

Related Work

Karnath et al. and *Loomis et al.*

Spatial awareness is a function of the temporal not the posterior parietal lobe

Hans-Otto Karnath, Susanne Ferber & Marc Himmelbach

Department of Cognitive Neurology, University of Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Our current understanding of spatial behaviour and parietal lobe function is largely based on the belief that spatial neglect in humans (a lack of awareness of space on the side of the body contralateral to a brain injury) is typically associated with lesions of the posterior parietal lobe. However, in monkeys, this disorder is observed after lesions of the superior temporal cortex[1], a puzzling discrepancy between the species. Here we show that, contrary to the widely accepted view, the superior temporal cortex is the neural substrate of spatial neglect in humans, as it is in monkeys. Unlike the monkey brain, spatial awareness in humans is a function largely confined to the right superior temporal cortex, a location topographically reminiscent of that for language on the left[2]. Hence, the decisive phylogenetic transition from monkey to human brain seems to be a restriction of a formerly bilateral function to the right side, rather than a shift from the temporal to the parietal lobe. One may speculate that thi temporal to the parietal lobe. One may speculate that this lateralization of spatial awareness parallels the emergence of a elaborate representation for language on the left side.

Spatial neglect is a characteristic failure to explore the side space contralateral to a brain lesion. Patients with this disor behave as if one side of the surrounding space had ceased to ex Since the early post-mortem studies, we have believed that humans, lesions located predominantly in the posterior par lobe are critical for this disorder. Analyses of computerized to raphy scans of right-hemispheric stroke patients with neglect that superimposed lateral projections of these lesions centred inferior parietal lobule (IPL)[3,4] and the temporo-parieto-o (TPO) junction[4]. More recent studies have confirmed the va this conclusion although evidence for additional pathology

Behavior Research Methods, Instruments, & Computers
1999, 31 (4), 557-564

Immersive virtual environment technology as a basic research tool in psychology

JACK M. LOOMIS and JAMES J. BLASCOVICH
University of California, Santa Barbara, California

and

ANDREW C. BEALL
Massachusetts Institute of Technology, Cambridge, Massachusetts

Immersive virtual environment (IVE) technology has great promise as a tool for basic experimental research in psychology. IVE technology gives participants the experience of being surrounded by the computer-synthesized environment. We begin with a discussion of the various devices needed to implement immersive virtual environments, including object manipulation and social interaction. We review the benefits and drawbacks associated with virtual environment technology, in comparison with more conventional ways of doing basic experimental research. We then consider a variety of examples of research using IVE technology in the areas of perception, spatial cognition, and social interaction.

Human history records a progression of artifacts for representing and recreating aspects of external reality, ranging from language, drawings, and sculpture in earlier times to the more modern artifacts of photographs, movies, television, and audio recordings. Relatively recently, the digital computer and its associated technologies, including three-dimensional (3-D) graphics, have given rise to increasingly realistic artifacts that blur the distinction between reality and its representation (Ellis, 1995).

The ultimate representational system would allow the observer to interact "naturally" with objects and other individuals within a simulated environment or "world," an experience indistinguishable from "normal reality." Although such a representational system might conceivably use direct brain stimulation in the future, it will more likely use digitally controlled displays that stimulate the human sensory organs, the natural conduits to the brain.

Displays of this type, referred to as *virtual displays* (VDs), although far from ideal, exist today. Following the terminology of others (e.g., Durlach & Mavor, 1995; Stanney & Salvendy, 1998), we refer to the corresponding environment represented and stored in the computer and experienced by the user as a *virtual environment* (VE). *Virtual environment technology* (VET) refers inclusively both to VDs and to the VEs so created, including VEs produced by using conventional desktop computer displays.

With the support of Grant N00014-95-1-0573 from the Office of Naval Research (to J.M.L.) and National Science Foundation Grants SBR 9872084 (to J.J.B.) and SBR 9873432 (to J.M.L. and J.J.B.), and with additional support from the University of California, Santa Barbara, the authors have developed several immersive virtual displays, use of which has stimulated many of the ideas expressed here. The authors thank Florence Gaunet and Patrick Péruch for comments on an earlier version of the article. Correspondence concerning this article should be addressed to J. M. Loomis, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: loomis@psych.ucsb.edu).

(*Virtual reality* is widely used as an alternative term, but we prefer VE.) An immersive virtual environment (IVE) is one in which the user is perceptually surrounded by the VE. Ivan Sutherland (1965), one of the originators of 3-D computer graphics, was the first person to conceive and build an immersive VD system. For the history of IVEs, see Ellis (1995), Kalawsky (1993), and Rheingold (1991).

There are two usual implementations of an IVE. The first of these involves placing multiple projection screens and loudspeakers around the user. A popular design is the CAVE (Cruz-Neira, Sandin, & DeFantini, 1993), which involves back-projecting the computer-generated visual imagery onto the translucent walls, floor, and ceiling of a moderately sized cubical room, in which the user is free to move; shutter glasses provide stereoscopic stimulation, so that one sees the VE not as projections on the room surfaces, but as solid 3-D structures within and/or outside of the cube. The second and more common implementation of an IVE involves the use of a head-mounted display (HMD), used in conjunction with a computer and a head tracker (Barfield & Furness, 1995; Biocca & Delaney, 1995; Burdea & Coiffett, 1994; Durlach & Mavor, 1995; Kalawsky, 1993). The head tracker measures the changing position and orientation of the user's head within the physical environment, information that is communicated to the rendering computer, which has stored within it a 3-D representation of the simulated environment (Meyer, Applewhite, & Biocca, 1992). At any given moment, the computer generates and outputs the visual and auditory imagery to the user's HMD from a perspective that is based on the position and orientation of the user's head. The HMD consists of earphones and video displays attached to a support worn on the head; the video display component is based on cathode ray tube (CRT) displays, liquid crystal displays, or laser-based retinal scanners (Barfield, Hendrix, Bjorneseth, Kaczmarek, &

Related Work

Visualization of Geo-tagged Information

Related Work

Visualization of Geo-tagged Information

TwitterStand: News in Tweets*

Jagan Sankaranarayanan†
jagan@cs.umd.edu

Hanan Samet†
hjs@cs.umd.edu

Benjamin E. Teitler†
bteitler@cs.umd.edu

Michael D. Lieberman†
codepoet@cs.umd.edu

Jon Sperling‡
Jon.Sperling@hud.gov

ABSTRACT

Twitter is an electronic medium that allows a large user populace to communicate with each other simultaneously. Inherent to Twitter is an asymmetrical relationship between friends and followers that provides an interesting social network-like structure among the users of Twitter. Twitter messages, called tweets, are restricted to 140 characters and thus are usually very focused. We investigate the use of Twitter to build a news processing system, called *TwitterStand*, from Twitter tweets. The idea is to capture tweets that correspond to late breaking news. The result is analogous to a distributed news wire service. The difference is that the identities of the contributors/reporters are not known in advance and there may be many of them. Furthermore, tweets are not sent according to a schedule: they occur as news is happening, and tend to be noisy while usually arriving at a high throughput rate. Some of the issues addressed include removing the noise, determining tweet clusters of interest bearing in mind that the methods must be online, and determining the relevant locations associated with the tweets.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Storage and Retrieval

General Terms

Algorithms, Design, Performance

*This work was supported in part by the National Science Foundation under Grants EIA-08-12377, CCF-08-30618, and IIS-07-13501, as well as NVIDIA Corporation, Microsoft Research, Google, the E.T.S. Walton Visitor Award of the Science Foundation of Ireland, and the National Center for Geocomputation at the National University of Ireland at Maynooth.

†Department of Computer Science, Center for Automation Research, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA.

‡HUD Office of Policy Development & Research (PD&R), 451 7th St. SW, Room 8146, Washington, DC 20410, USA.

Keywords

Twitter, News, Geotagging, Online clustering

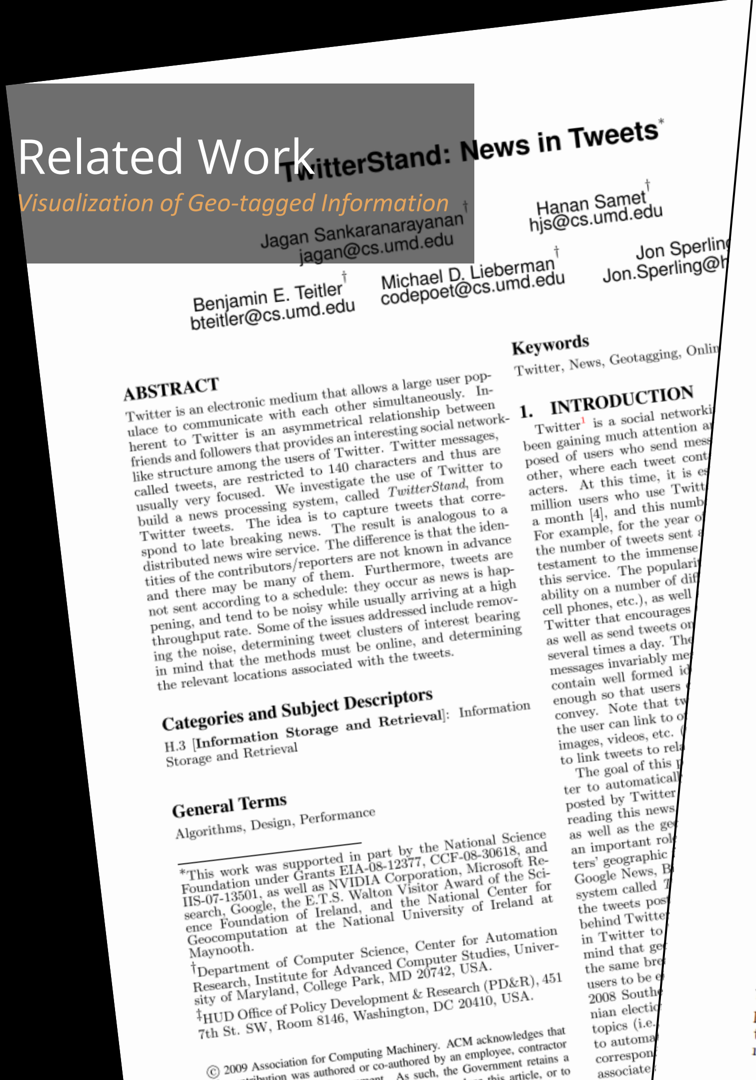
1. INTRODUCTION

Twitter[1] is a social networking website that recently has been gaining much attention and following. Twitter is composed of users who send messages (termed *tweets*) to each other, where each tweet contains a maximum of 140 characters. At this time, it is estimated that there are 6 to 7 million users who use Twitter a total of 134 million times a month [4], and this number is increasing at a rapid rate. For example, for the year of 2008, Twitter grew in terms of the number of tweets sent at a rate of 1382% [12] which is a testament to the immense popularity and wide adoption of this service. The popularity of Twitter stems from its availability on a number of different electronic devices (e.g., web, cell phones, etc.), as well as the prevalence of a subculture in Twitter that encourages users to acquire a large friend pool, as well as send tweets on a wide variety of subjects, typically several times a day. The restriction on the lengths of Twitter messages invariably means that the tweets do not necessarily contain well formed ideas, being rather brief, yet complete enough so that users can make sense of the ideas that they convey. Note that tweets also have a mechanism by which the user can link to other objects on the web such as articles, images, videos, etc. (termed *artifacts*) which is typically used to link tweets to related material on the Internet.

The goal of this paper is to demonstrate how to use Twitter to automatically obtain breaking news from the tweets posted by Twitter users, and to provide a map interface for reading this news, since the geographic location of the user as well as the geographic terms comprising the tweets play an important role in *clustering* tweets and establishing clusters' geographic foci. In contrast to news aggregators such as Google News, Bing News, and Yahoo! News, we introduce a system called *TwitterStand* that works exclusively with only the tweets posted by the users of Twitter. The key novelty behind TwitterStand is one of mobilizing the millions of users in Twitter to be our eyes and ears in the world, bearing in mind that geographically proximate users often tweet about the same breaking news. In other words, we rely on Twitter users to be either providers of original news content (e.g., the 2008 Southern California earthquake [13] and the 2009 Iranian election [3]), or expressers of opinions on current news topics (i.e., mini blogs), both of which enable TwitterStand to automatically identify current news topics and cluster the corresponding tweets into appropriate news stories. We also associate an importance score with each news topic which can

Related Work

Visualization of Geo-tagged Information



TwitterStand: News in Tweets*

Jagan Sankaranarayanan[†]
jagan@cs.umd.edu

Hanan Samet[†]
hjs@cs.umd.edu

Benjamin E. Teitler[†]
bteitler@cs.umd.edu

Michael D. Lieberman[†]
codepoet@cs.umd.edu

Jon Sperling[‡]

ABSTRACT

Twitter is an electronic medium that allows a large user populace to communicate with each other simultaneously. Inherent to Twitter is an asymmetrical relationship between friends and followers that provides an interesting social network-like structure among the users of Twitter. Twitter messages, called tweets, are restricted to 140 characters and thus are usually very focused. We investigate the use of Twitter to build a news processing system, called *TwitterStand*, from Twitter tweets. The idea is to capture tweets that correspond to late breaking news. The result is analogous to a distributed news wire service. The difference is that the identities of the contributors/reporters are not known in advance and there may be many of them. Furthermore, tweets are not sent according to a schedule: they occur as news is happening, and tend to be noisy while usually arriving at a high throughput rate. Some of the issues addressed include removing the noise, determining tweet clusters of interest bearing in mind that the methods must be online, and determining the relevant locations associated with the tweets.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Storage and Retrieval

General Terms

Algorithms, Design, Performance

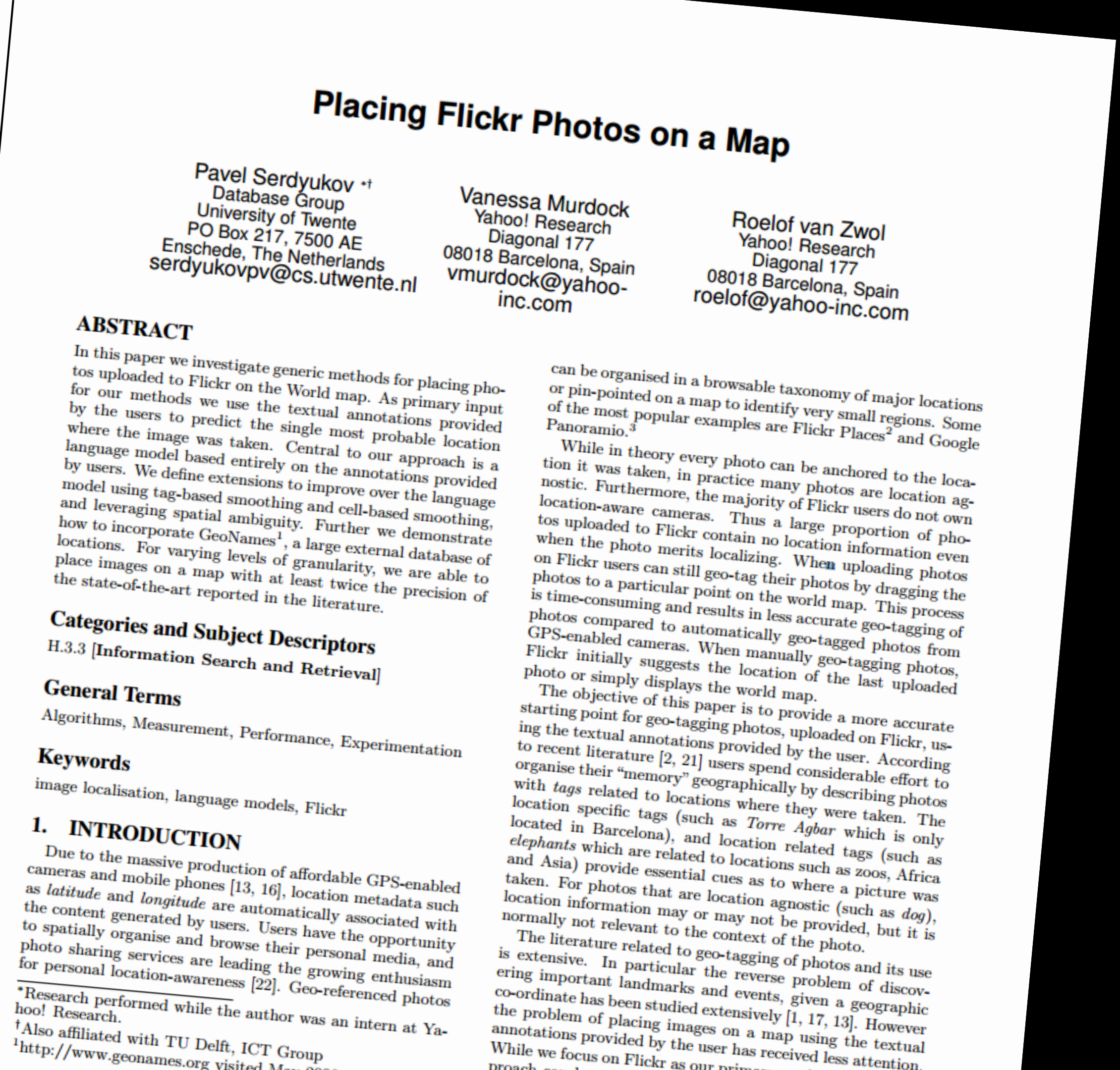
Keywords

Twitter, News, Geotagging, Onlin

*This work was supported in part by the National Science Foundation under Grants EIA-08-12377, CCF-08-30618, and IIS-07-13501, as well as NVIDIA Corporation, Microsoft Research, Google, the E.T.S. Walton Visitor Award of the Science Foundation of Ireland, and the National Centre for Geocomputation at the National University of Ireland at Maynooth.

[†]Department of Computer Science, Center for Automation Research, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA.

[‡]HUD Office of Policy Development & Research (PD&R), 451 7th St. SW, Room 8146, Washington, DC 20410, USA.



Placing Flickr Photos on a Map

Pavel Serdyukov [*,†]
Database Group
University of Twente
PO Box 217, 7500 AE
Enschede, The Netherlands
serdyukovpv@cs.utwente.nl

Vanessa Murdock
Yahoo! Research
Diagonal 177
08018 Barcelona, Spain
vmurdock@yahoo-inc.com

Roelof van Zwol
Yahoo! Research
Diagonal 177
08018 Barcelona, Spain
roelof@yahoo-inc.com

ABSTRACT

In this paper we investigate generic methods for placing photos uploaded to Flickr on the World map. As primary input for our methods we use the textual annotations provided by the users to predict the single most probable location where the image was taken. Central to our approach is a language model based entirely on the annotations provided by users. We define extensions to improve over the language model using tag-based smoothing and cell-based smoothing, and leveraging spatial ambiguity. Further we demonstrate how to incorporate GeoNames[1], a large external database of locations. For varying levels of granularity, we are able to place images on a map with at least twice the precision of the state-of-the-art reported in the literature.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]

General Terms

Algorithms, Measurement, Performance, Experimentation

Keywords

image localisation, language models, Flickr

1. INTRODUCTION

Due to the massive production of affordable GPS-enabled cameras and mobile phones [13, 16], location metadata such as *latitude* and *longitude* are automatically associated with the content generated by users. Users have the opportunity to spatially organise and browse their personal media, and photo sharing services are leading the growing enthusiasm for personal location-awareness [22]. Geo-referenced photos can be organised in a browsable taxonomy of major locations or pin-pointed on a map to identify very small regions. Some of the most popular examples are Flickr Places[2] and Google Panoramio.[3]

While in theory every photo can be anchored to the location it was taken, in practice many photos are location agnostic. Furthermore, the majority of Flickr users do not own location-aware cameras. Thus a large proportion of photos uploaded to Flickr contain no location information even when the photo merits localizing. When uploading photos on Flickr users can still geo-tag their photos by dragging the photos to a particular point on the world map. This process is time-consuming and results in less accurate geo-tagging of photos compared to automatically geo-tagged photos from GPS-enabled cameras. When manually geo-tagging photos, Flickr initially suggests the location of the last uploaded photo or simply displays the world map.

The objective of this paper is to provide a more accurate starting point for geo-tagging photos, uploaded on Flickr, using the textual annotations provided by the user. According to recent literature [2, 21] users spend considerable effort to organise their "memory" geographically by describing photos with *tags* related to locations where they were taken. The location specific tags (such as *Torre Agbar* which is only located in Barcelona), and location related tags (such as *elephants* which are related to locations such as zoos, Africa and Asia) provide essential cues as to where a picture was taken. For photos that are location agnostic (such as *dog*), location information may or may not be provided, but it is normally not relevant to the context of the photo.

The literature related to geo-tagging of photos and its use is extensive. In particular the reverse problem of discovering important landmarks and events, given a geographic co-ordinate has been studied extensively [1, 17, 13]. However the problem of placing images on a map using the textual annotations provided by the user has received less attention.

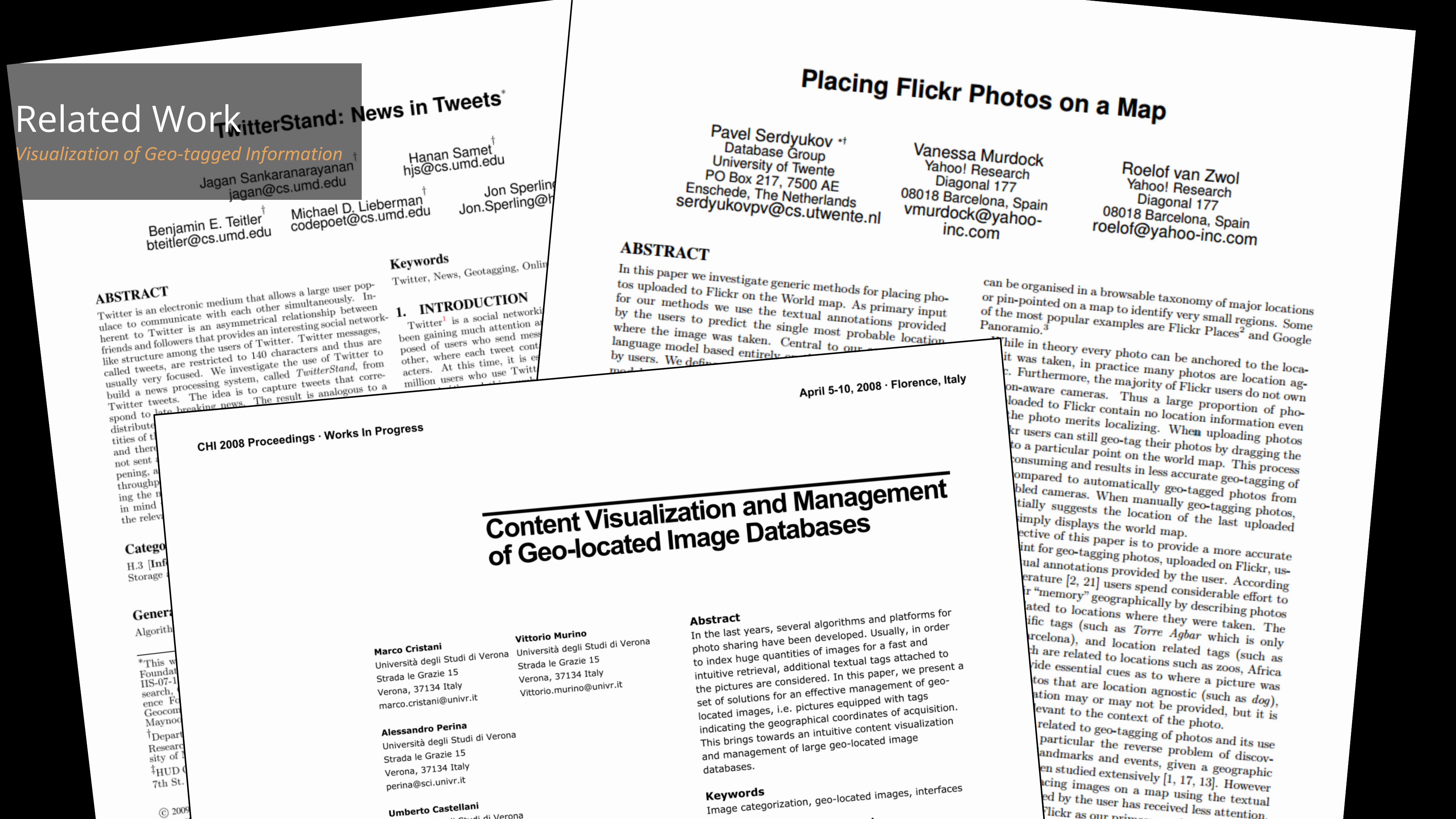
*Research performed while the author was an intern at Yahoo! Research.

[†]Also affiliated with TU Delft, ICT Group

[1]http://www.geonames.org

Related Work

Visualization of Geo-tagged Information



Related Work

Visualization of Geo-tagged Information

Related Work

Panoramio - Photos of the World
Cuenca and Manchón, 2005 - Now

image courtesy: panoramio.com

Related Work

TwitterStand: News in Tweets

Sankaranarayanan et al. SIGGIS 2009

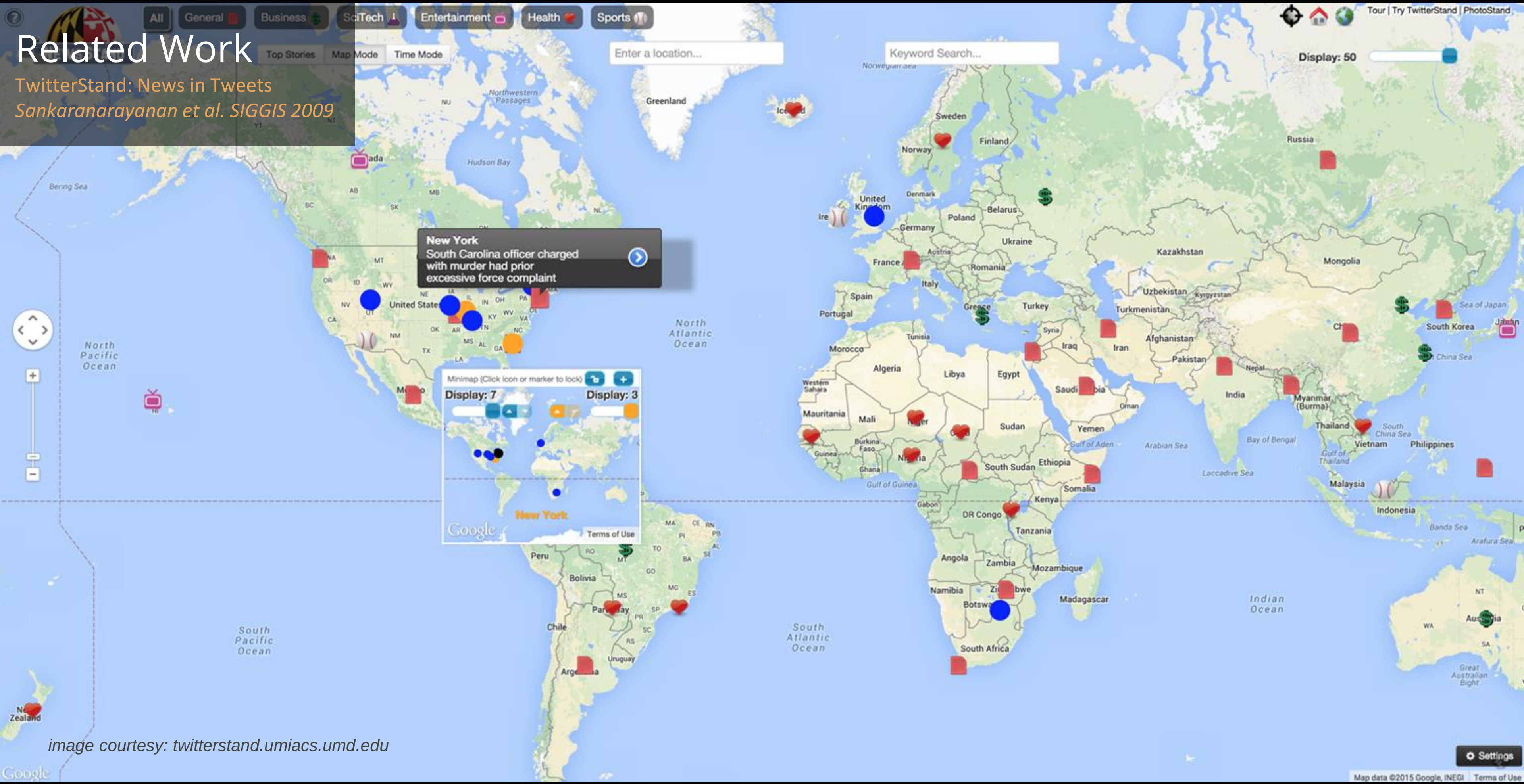


image courtesy: twitterstand.umiacs.umd.edu

Related Work

Flickr - World Map

Serdyukov et al. SIGIR 2009

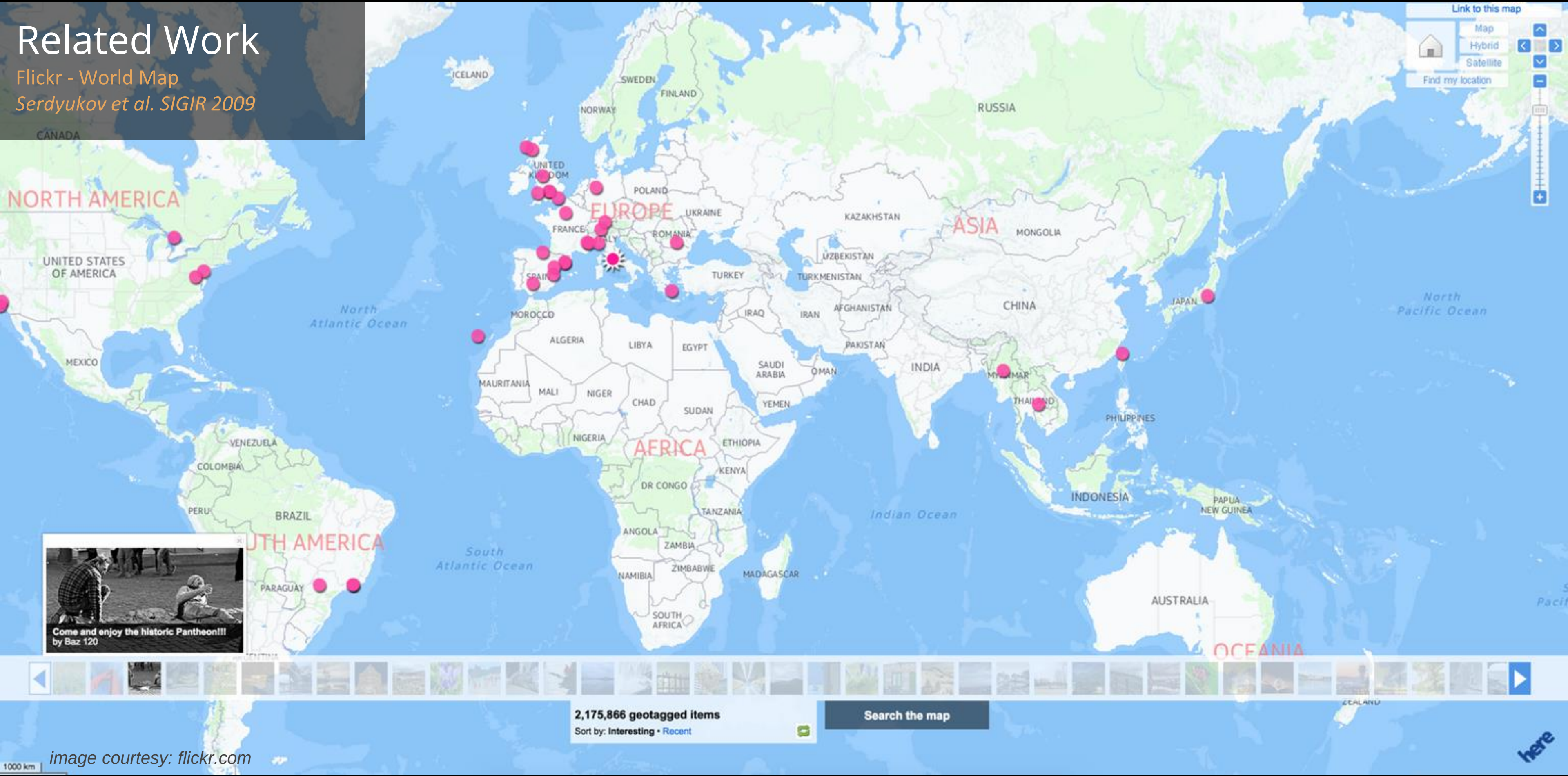


image courtesy: flickr.com

Related Work

PhotoStand for News Photos

Samet et al. VLDB 2013

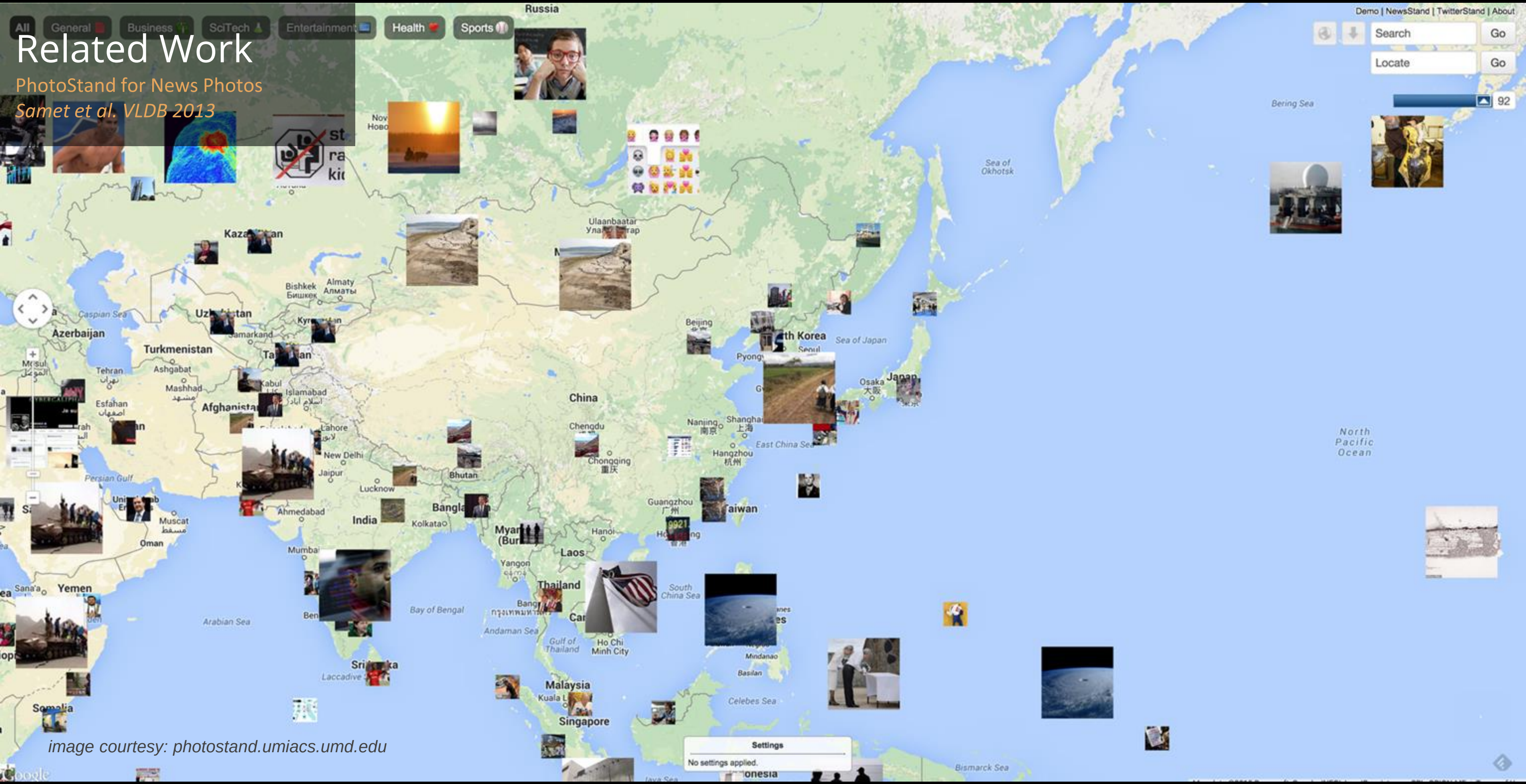


image courtesy: photostand.umiacs.umd.edu

Related Work

Visualization of Geo-tagged Information (cont.)

Related Work

Visualization of Geo-tagged Information (cont.)

UIST 2001 (ACM Symp. on User Interface Software and Technology), Orlando, FL, November 11-14, 2001, pp. 101-110

View Management for Virtual and Augmented Reality

Blaine Bell *Steven Feiner* *Tobias Höllerer*
Department of Computer Science
500 W 120th St., 450 CS Building
Columbia University
New York, NY 10027
{bell,feiner,htobias}@cs.columbia.edu

ABSTRACT
We describe a view-management component for interactive 3D user interfaces. By *view management*, we mean maintaining visual constraints on the projections of objects on the view plane, such as locating related objects near each other, or preventing objects from occluding each other. Our view-management component accomplishes this by modifying selected object properties, including position, size, and transparency, which are tagged to indicate their constraints. For example, some objects may have geometric properties that are determined entirely by a physical simulation and which cannot be modified, while other objects may be annotations whose position and size are flexible.

We introduce algorithms that use upright rectangular extents to represent on the view plane a dynamic and efficient approximation of the occupied space containing the projections of visible portions of 3D objects, as well as the unoccupied space in which objects can be placed to avoid occlusion. Layout decisions from previous frames are taken into account to reduce visual discontinuities. We present augmented reality and virtual reality examples to which we have applied our approach, including a dynamically labeled and annotated environment.

CR Categories and Subject Descriptors: H.5.1 **[Information Interfaces and Presentation]** Multimedia Information Interfaces—*Artificial, augmented, and virtual realities*; H.5.2 **[Information Interfaces and Presentation]** User Interfaces—*Graphical User Interfaces, Screen design*; I.3.6 **[Computer Graphics]** Methodology and Techniques—*Interaction Techniques*; I.3.7 **[Computer Graphics]** Three-Dimensional Graphics and Realism—*Virtual Reality*.

Additional Keywords and Phrases: view management, environment management, annotation, labeling, wearable computing, augmented reality, virtual environments

1. INTRODUCTION

Designing a 3D graphical user interface (UI) requires creating a set of objects and their properties, arranging them in a scene, setting a viewing specification, determining lighting and rendering parameters, and deciding how to update these decisions for each frame. Some of these decisions may be fully constrained; for example, a simulation may determine the position and shape of certain ... the viewing specification may be explicitly ... In contrast, other decisions must be ... especially interested in

Related Work

Visualization of Geo-tagged Information (cont.)

UIST 2001 (ACM Symp. on User I...

View Manage...

ABSTRACT

We describe a view-mana... 3D user interfaces. By... maintaining visual constr... on the view plane, such a... other, or preventing obje... view-management con... modifying selected obj... size, and transparency... constraints. For examp... properties that are... simulation and which... objects may be anno... flexible.

We introduce algo... extents to represent... efficient approximat... the projections of v... the unoccupied sp...

Photo Tourism: Exploring Photo Collections in 3D

Noah Snavely, University of Washington
Steven M. Seitz, University of Washington
Richard Szeliski, Microsoft Research

(a) (b) (c)

Figure 1: Our system takes unstructured collections of photographs such as those from online image searches (a) and reconstructs 3D points and viewpoints (b) to enable novel ways of browsing the photos (c).

Abstract

We present a system for interactively browsing and exploring large unstructured collections of photographs of a scene using a novel 3D interface. Our system consists of an image-based modeling front end that automatically computes the viewpoint of each photograph as well as a sparse 3D model of the scene and image to model correspondences. Our *photo explorer* uses image-based rendering techniques to smoothly transition between photographs, while also enabling full 3D navigation and exploration of the set of images and world geometry, along with auxiliary information such as overhead maps. Our system also makes it easy to construct photo tours of scenic or historic locations, and to annotate image details, which are automatically transferred to other relevant images. We demonstrate our system on several large personal photo collections as well as images gathered from Internet photo sharing sites.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Modeling and recovery of physical attributes

Keywords: image-based rendering, image-based modeling, photo browsing, structure from motion

1 Introduction

...goal of image-based rendering is to evoke a visceral sense ... of photographs of a scene. The ... this goal

is that these approaches will one day allow virtual tourism of the world's interesting and important sites.

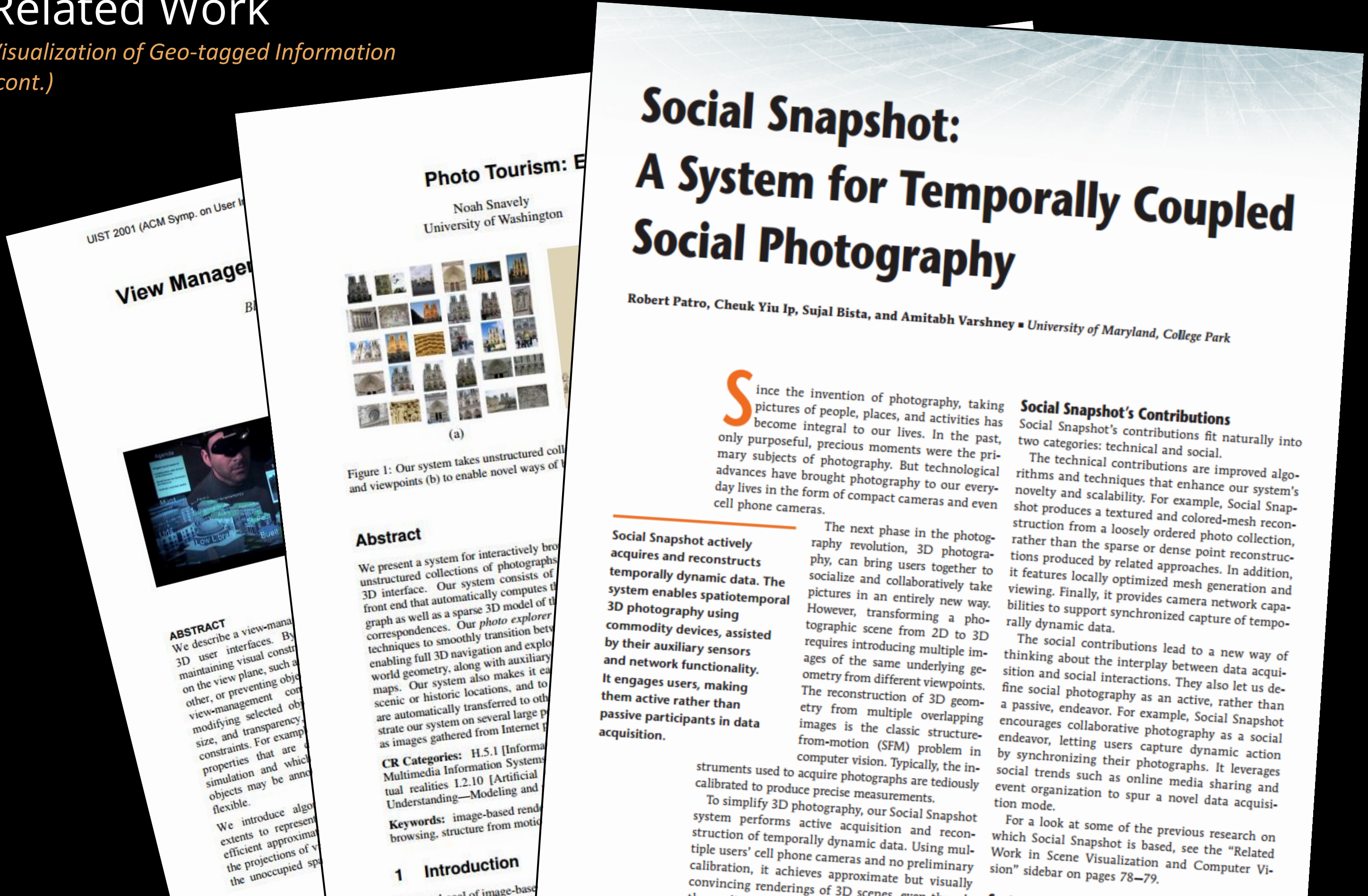
During this same time, digital photography, together with the Internet, have combined to enable sharing of photographs on a truly massive scale. For example, a Google image search on "Notre Dame Cathedral" returns over *15,000* photos, capturing the scene from myriad viewpoints, levels of detail, lighting conditions, seasons, decades, and so forth. Unfortunately, the proliferation of shared photographs has outpaced the technology for browsing such collections, as tools like Google (www.google.com) and Flickr (www.flickr.com) return pages and pages of thumbnails that the user must comb through.

In this paper, we present a system for browsing and organizing large photo collections of popular sites which exploits the common 3D geometry of the underlying scene. Our approach is based on computing, from the images themselves, the photographers' locations and orientations, along with a sparse 3D geometric representation of the scene, using a state-of-the-art image-based modeling system. Our system handles large collections of unorganized photographs taken by different cameras in widely different conditions. We show how the inferred camera and scene information enables the following capabilities:

- **Scene visualization.** Fly around popular world sites in 3D by morphing between photos.
- **Object-based photo browsing.** Show me more images that contain this object or part of the scene.
- **Where was I?** Tell me where I was when I took this picture.
- **What am I looking at?** Tell me about objects visible in this image by transferring annotations from similar images.

Related Work

Visualization of Geo-tagged Information (cont.)



Social Snapshot: A System for Temporally Coupled Social Photography

Robert Patro, Cheuk Yiu Ip, Sujal Bista, and Amitabh Varshney ▪ *University of Maryland, College Park*

Since the invention of photography, taking pictures of people, places, and activities has become integral to our lives. In the past, only purposeful, precious moments were the primary subjects of photography. But technological advances have brought photography to our everyday lives in the form of compact cameras and even cell phone cameras.

Social Snapshot actively acquires and reconstructs temporally dynamic data. The system enables spatiotemporal 3D photography using commodity devices, assisted by their auxiliary sensors and network functionality. It engages users, making them active rather than passive participants in data acquisition.

The next phase in the photography revolution, 3D photography, can bring users together to socialize and collaboratively take pictures in an entirely new way. However, transforming a photographic scene from 2D to 3D requires introducing multiple images of the same underlying geometry from different viewpoints. The reconstruction of 3D geometry from multiple overlapping images is the classic structure-from-motion (SFM) problem in computer vision. Typically, the instruments used to acquire photographs are tediously calibrated to produce precise measurements.

To simplify 3D photography, our Social Snapshot system performs active acquisition and reconstruction of temporally dynamic data. Using multiple users' cell phone cameras and no preliminary calibration, it achieves approximate but visually convincing renderings of 3D scenes, even th…

Social Snapshot's Contributions

Social Snapshot's contributions fit naturally into two categories: technical and social.

The technical contributions are improved algorithms and techniques that enhance our system's novelty and scalability. For example, Social Snapshot produces a textured and colored-mesh reconstruction from a loosely ordered photo collection, rather than the sparse or dense point reconstructions produced by related approaches. In addition, it features locally optimized mesh generation and viewing. Finally, it provides camera network capabilities to support synchronized capture of temporally dynamic data.

The social contributions lead to a new way of thinking about the interplay between data acquisition and social interactions. They also let us define social photography as an active, rather than a passive, endeavor. For example, Social Snapshot encourages collaborative photography as a social endeavor, letting users capture dynamic action by synchronizing their photographs. It leverages social trends such as online media sharing and event organization to spur a novel data acquisition mode.

For a look at some of the previous research on which Social Snapshot is based, see the "Related Work in Scene Visualization and Computer Vision" sidebar on pages 78–79.

Related Work

Visualization of Geo-tagged Information (cont.)

UIST 2001 (ACM Symp. on User I...

View Manage...

Photo Tourism: E...

Noah Snavely
University of Washington

Figure 1: Our system takes unstructured coll... and viewpoints (b) to enable novel ways of...

Social...
A Syst...
Social...

Robert Patro, Ch...

3D Wikipedia: Using online text to automatically label and navigate reconstructed geometry

Bryan C. Russell[1] Ricardo Martin-Brualla[2] Daniel J. Butler[2] Steven M. Seitz[2] Luke Zettlemoyer[2]

[1]Intel Labs [2]University of Washington

Figure 1: *Given a reference text describing a specific site, for example the Wikipedia article above for the Pantheon, we* **automatically** *create a labeled 3D reconstruction, with objects in the model linked to where they are mentioned in the text. The user interface enables coordinated browsing of the text with the visualization (see video).*

Abstract

We introduce an approach for analyzing Wikipedia and other text, together with online photos, to produce *annotated* 3D models of famous tourist sites. The approach is completely automated, and leverages online text and photo co-occurrences via Google Image Search. It enables a number of new interactions, which we demonstrate in a new 3D visualization tool. Text can be selected to move the camera to the corresponding objects, 3D bounding boxes provide anchors back to the text describing them, and the overall narrative of the text provides a temporal guide for automatically flying through the scene to visualize the world as you read about it. We show compelling results on several major tourist sites.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities I.2.7 [Artificial Intelligence]: Natural Language Processing—Text analysis I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Modeling and recovery of physical attributes

Keywords: image-based modeling and rendering, Wikipedia, natural language processing, 3D visualization

Links: DL PDF

1 Introduction

Tourists have long relied on guidebooks and other reference texts to learn about and navigate sites of interest. While guidebooks are packed with interesting historical facts and descriptions of site-specific objects and spaces, it can be difficult to fully *visualize* the scenes they present. The primary cues come from images provided with the text, but coverage is sparse and it can be difficult to understand the spatial relationships between each image viewpoint. For example, the Berlitz and Lonely Planet guides [Berlitz International 2003; Garwood and Hole 2012] for Rome each contain just a single photo of the Pantheon, and have a similar lack of photographic coverage of other sites. Even online sites such as Wikipedia, which do not have space restrictions, have similarly sparse and disconnected visual coverage.

Instead of relying exclusively on static images embedded in text, suppose you could create an interactive, photorealistic visualization, where, for example, a Wikipedia page is shown next to a detailed 3D model of the described site. When you select an object (e.g., "Raphaels tomb") in the text, it flies you to the corresponding location in the scene via a smooth, photorealistic transition. Similarly, when you click on an object in the visualization, it highlights the corresponding descriptive text on the Wikipedia page. Our goal is to create such a visualization **completely automatically** by analyzing the Wikipedia page itself, together with many photos of the site available online (Figure 1).

Related Work

Visualization of Geo-tagged Information (cont.)

Our Approach?

Social Street View

Social Street View

An immersive social media navigation system in mixed-reality!

Demonstration

The Augmentarium, UMIACS
6000 x 3000 pixels

Natural Immersive Virtual Reality?

almost never in a natural immersive virtual reality settings.

Conception, architecting & implementation

1

Social Street View

A mixed reality system that can depict geo-tagged social media in immersive 3D web environments

Blending multiple modalities of

2

Street View + Social Media

Depth maps, normal maps, and road orientation
GPS coordinates and time creation

Enhancing visual augmentation

3

Maximal Poisson-disk sampling

Evaluated by image saliency metrics

Achieving cross-platform compatibility by

4

WebGL + Three.js

smartphones, tablets, desktop, high-resolution large-area wide field of view tiled display walls, as well as head-mounted displays.

Technical Challenges?

System Overview

Architecture

Social Street View System Flowchart

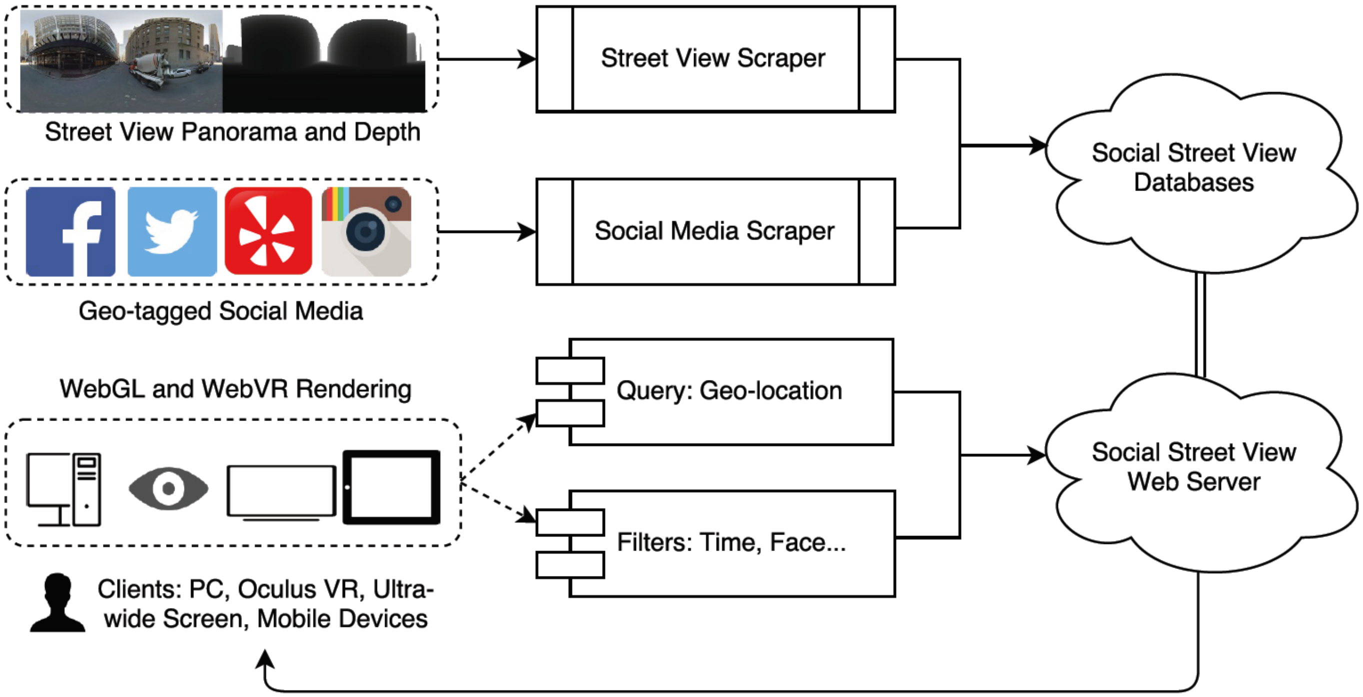


image courtesy: Google

Street View Cars - Cameras, Lasers and GPS

Image courtesy from Google Street View

image courtesy: Google

image courtesy: Google

Tiles of Panoramic Images

Image courtesy from Google Street View

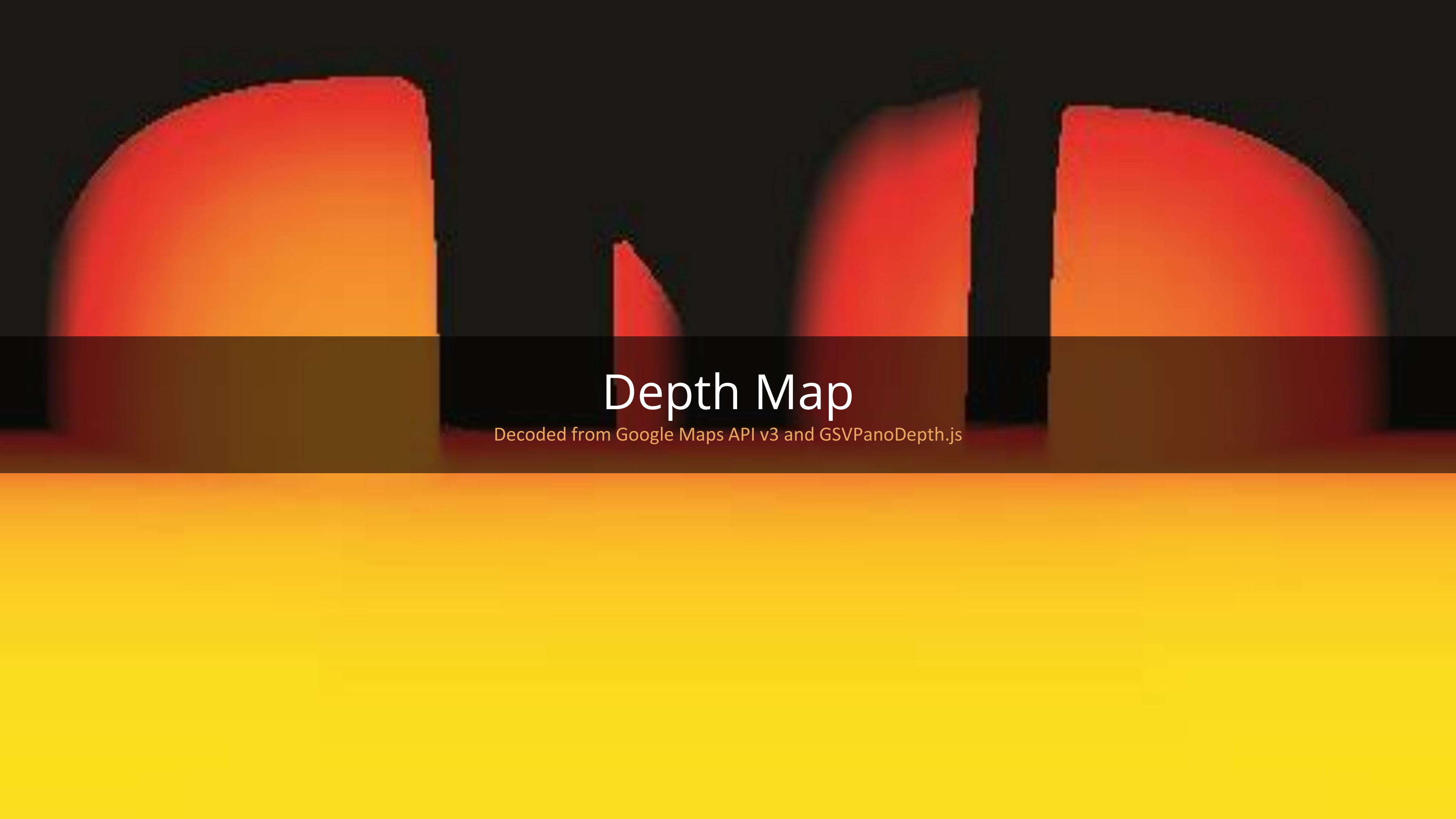
image courtesy: Google

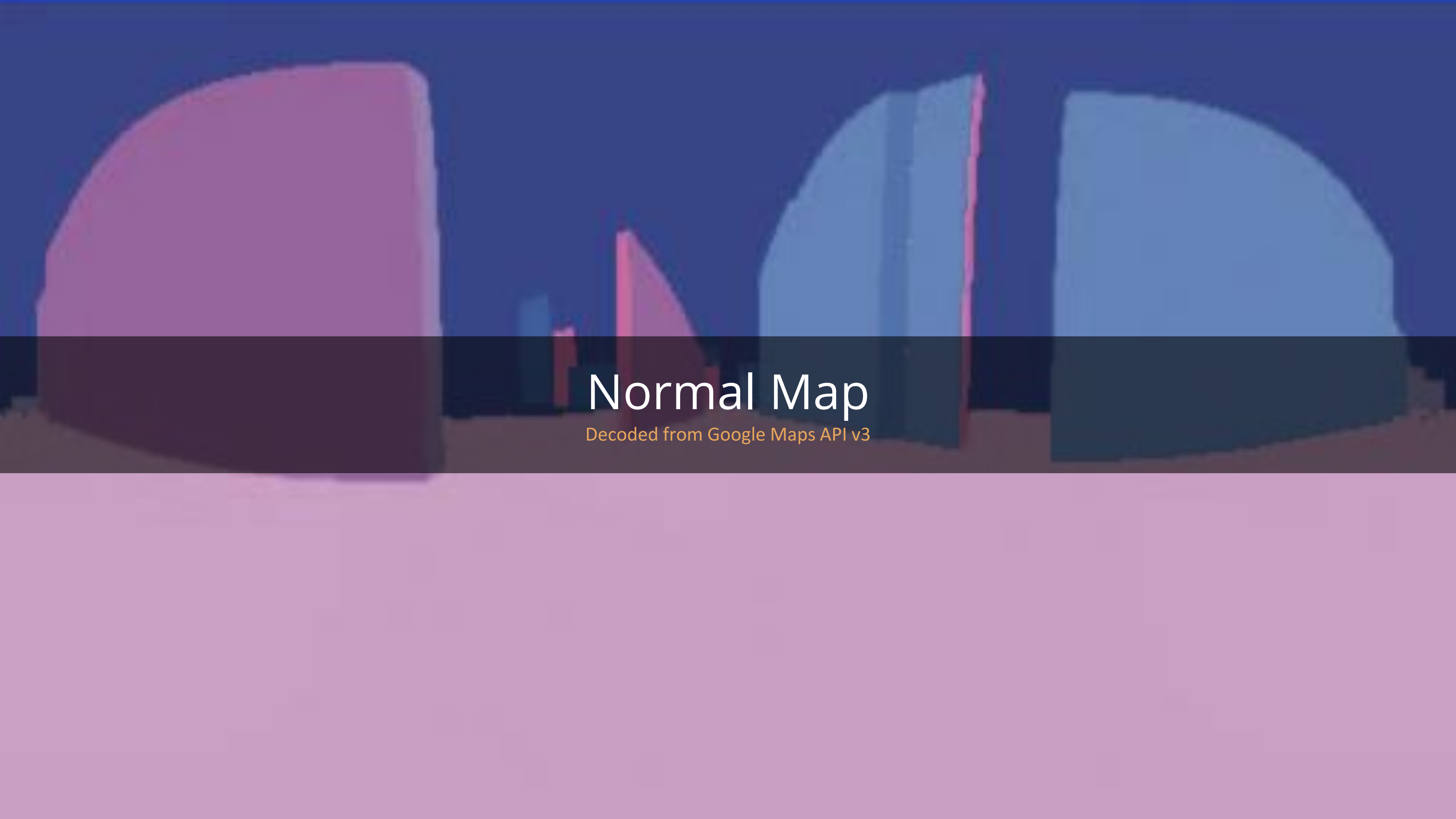

PED XING

Panorama

Depth Map

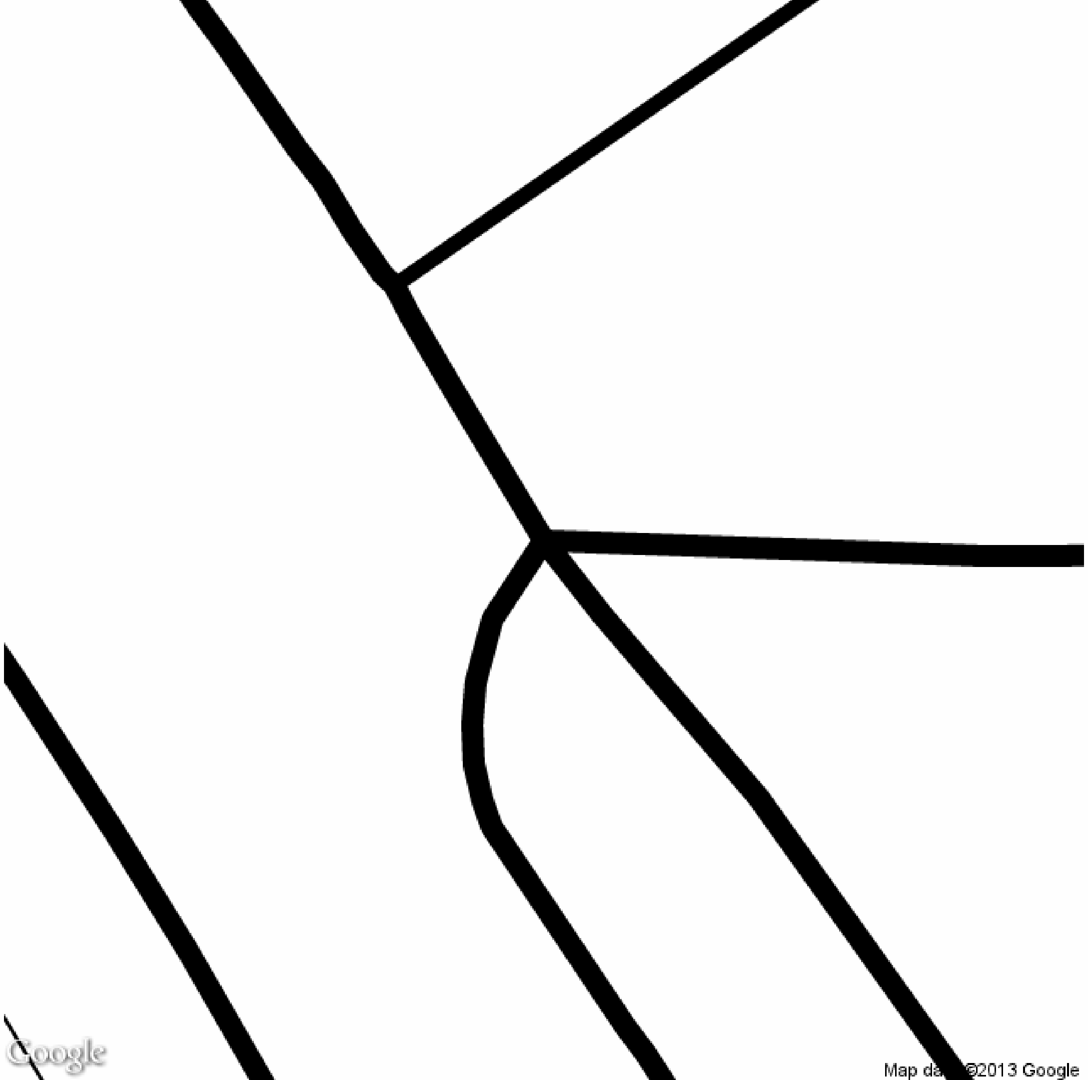
Decoded from Google Maps API v3 and GSVPanoDepth.js





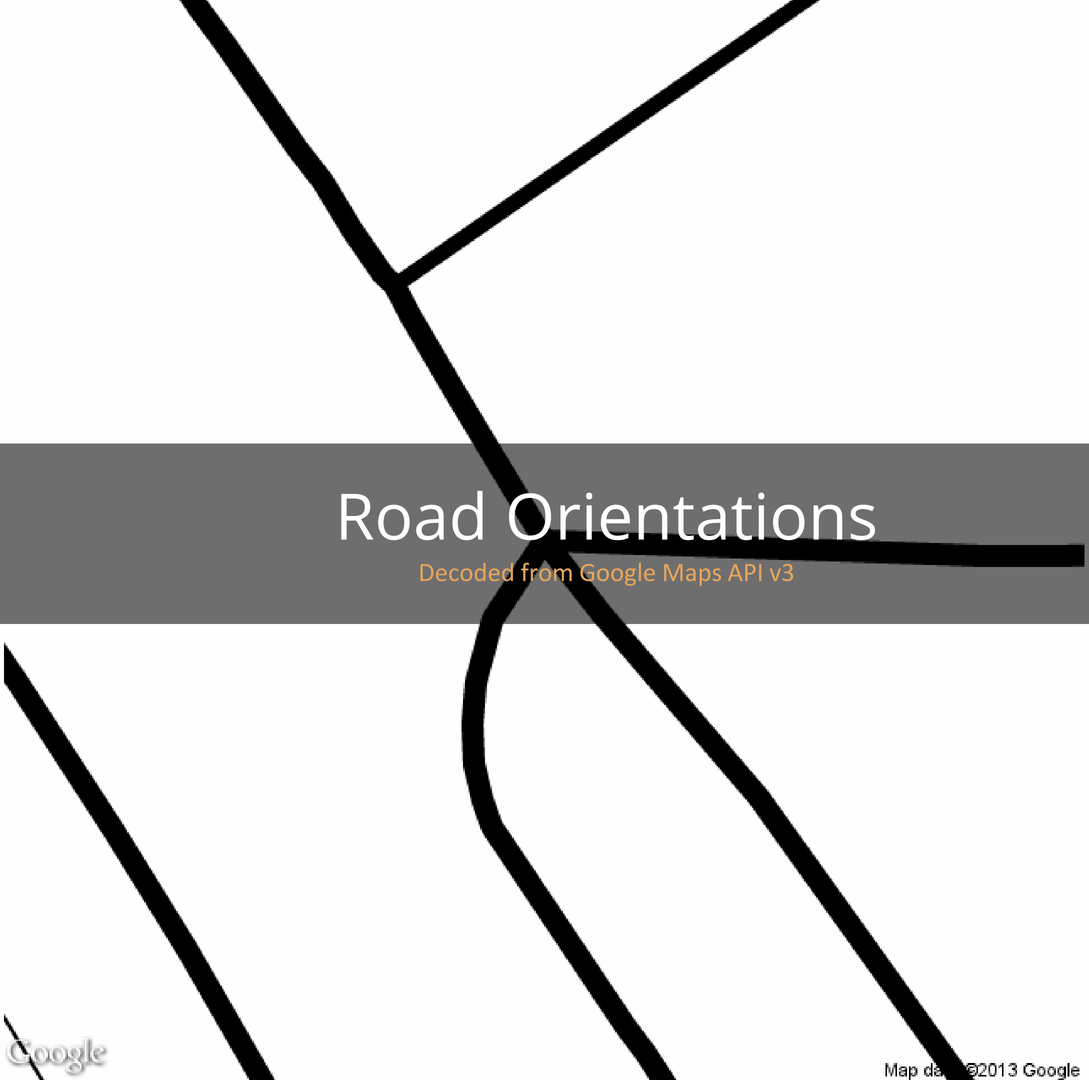
Normal Map

Decoded from Google Maps API v3


Map da ©2013 Google

Road Orientations

Decoded from Google Maps API v3



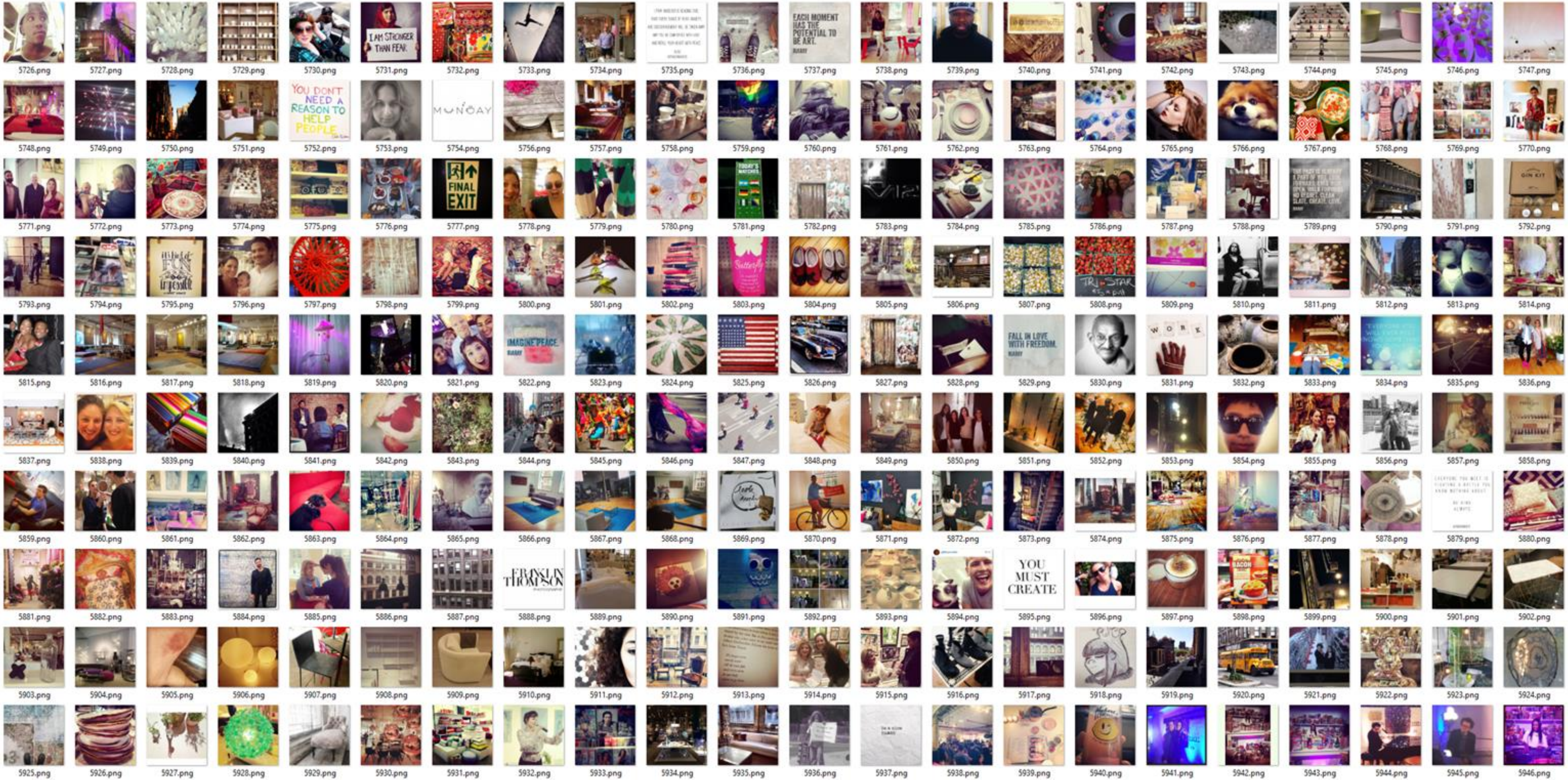


image courtesy: Instagram

Multi-level Strategy for Streaming Social Media

Image courtesy from Instagram users in public domains

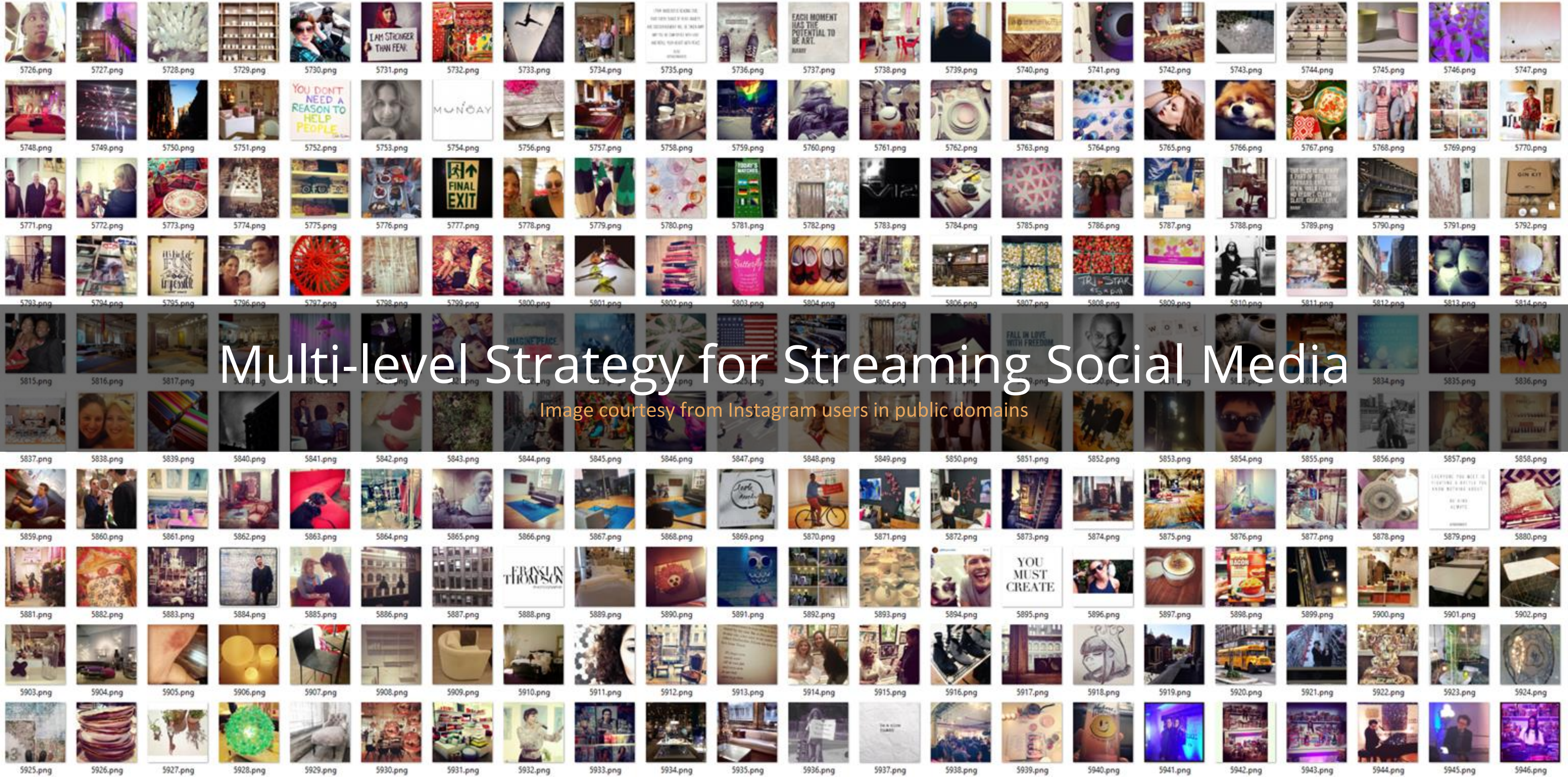
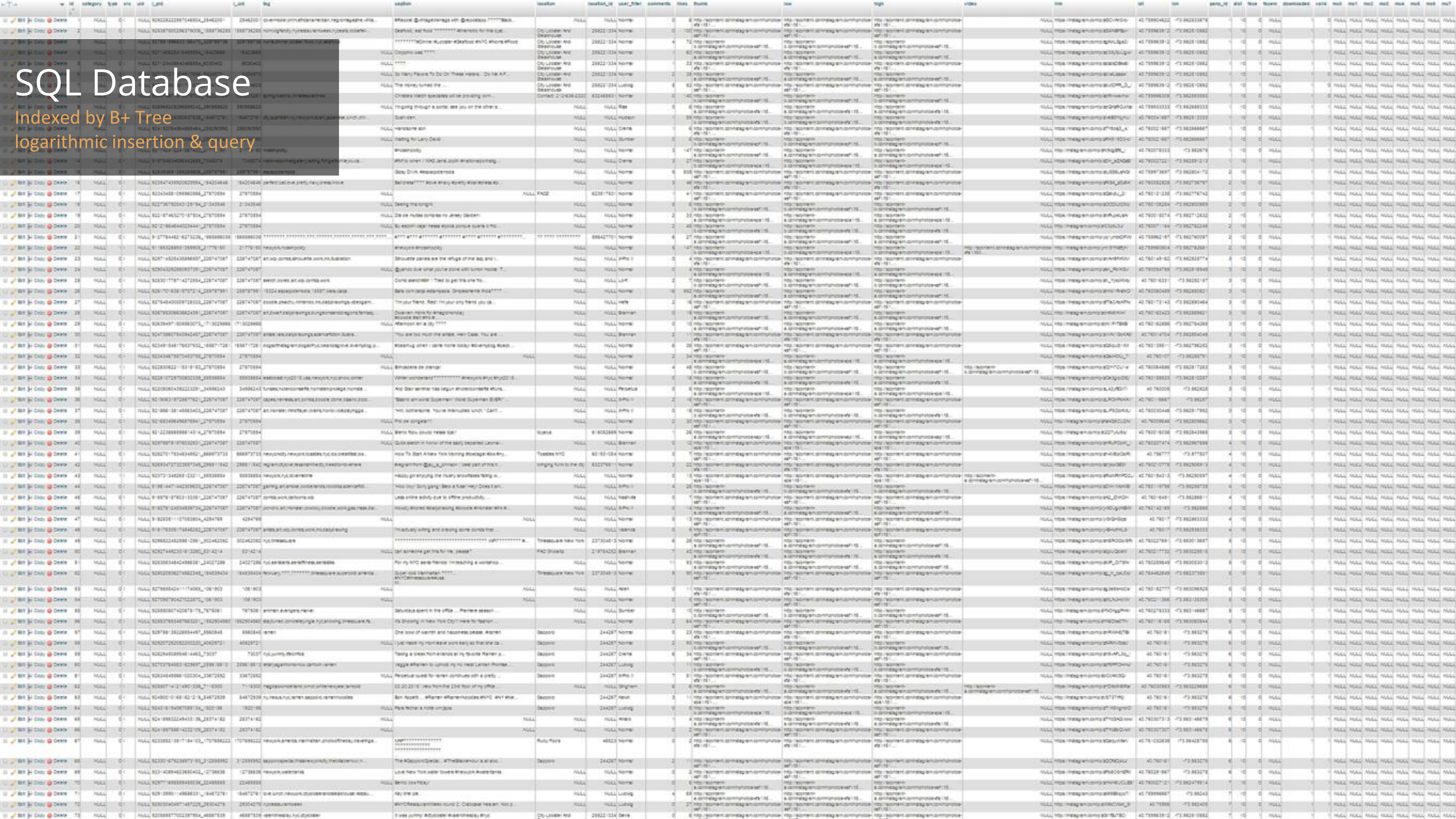


image courtesy: Instagram

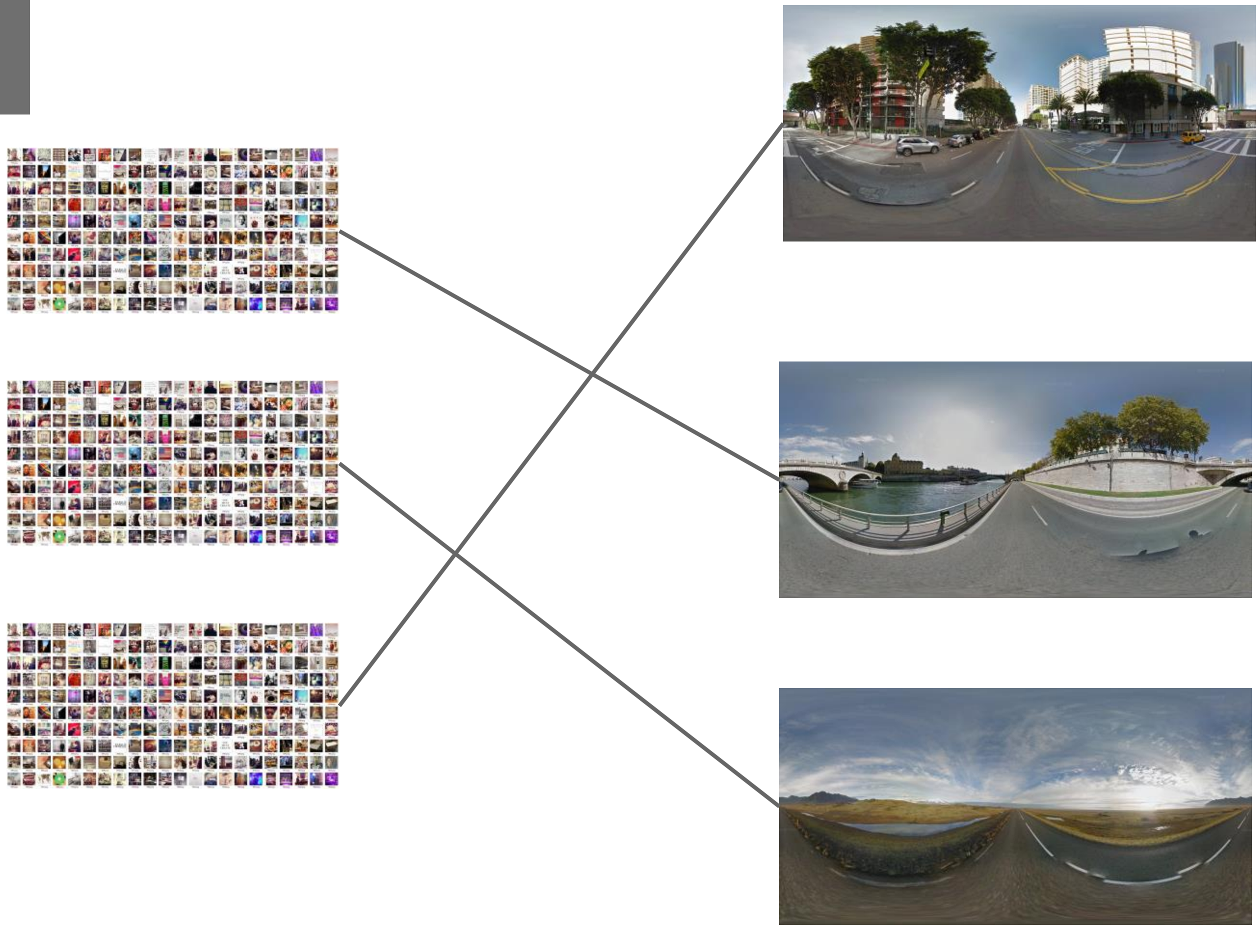
SQL Database

Indexed by B+ Tree
logarithmic insertion & query



Bipartite Graph

G = V(V, S, E)



Haversine Formula

Andrew, 1805

$$\alpha_{ij} = \sin^2(\frac{\varphi_i - \varphi_j}{2}) + \cos\varphi_i \cdot \varphi_j \cdot \sin^2(\frac{\lambda_i - \lambda_j}{2})$$

$$\beta_{ij} = 2 \cdot atan2(\sqrt{\alpha_{ij}}, \sqrt{(1 - \alpha_{ij})})$$

$$d_{ij} = R \cdot \beta_{ij}$$

Interface

Social Street View
Augmentarium Depth
Augmentarium Baseline
Oculus VR
Address, Geo-location or HashID
Go
Keywords
Search
Urban
Rural
Indoor
Terrain
Misc
NYC
London
Paris
Rome
Tokyo
Month
1 - 9
Hour
3 - 20
Distance
0 - 64
2D
Depth
Side
Front
Model
Enhance Dark
None
Enhance Bright
Spring
Summer
Autumn
Winter
All
Face-only
None-face-only
All
No-text-only
With-text-only
Include-video
DirLight
Unlimited
AmbLight
Unlimited
BloomEffect
Unlimited
DarkEffect
Unlimited
LightColor
Unlimited
Number
Unlimited
Radius
Unlimited
Help
Credits
Map
Satellite
Welcome to Social Street View!
Address: 129 W 49th St, New York

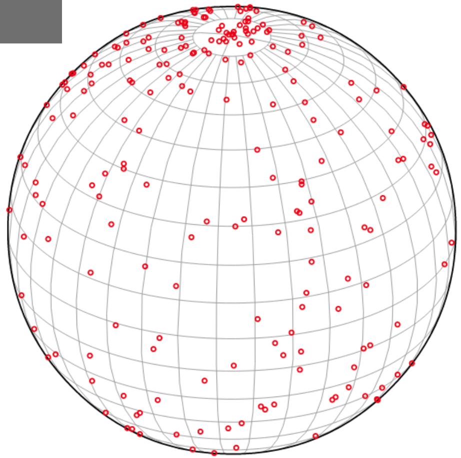
Algorithm

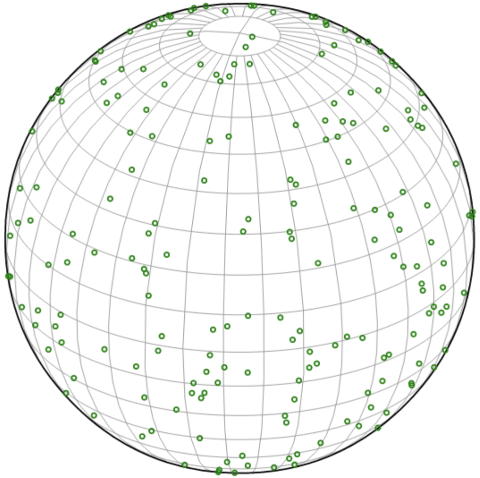
Adding depth & normal map & maximal Poisson-disk sampling

Social Street View enables users to see-through the nearby restaurants.

Baseline

Random Uniform Sampling





$$\varphi_i = (u_i - \frac{1}{2})\pi,\ \lambda_i = (2v_i - 1)\pi$$

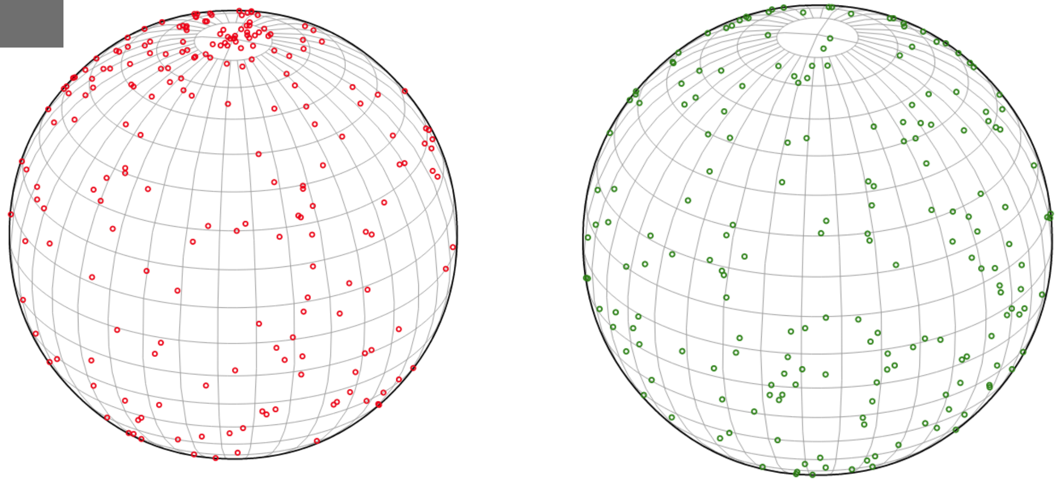
$$x_i = \cos\varphi_i \cos\lambda_i,\ y_i = \sin\varphi_i,\ z_i = \cos\varphi_i sin\lambda_i$$

Without uniform sampling

Accumulation

Baseline

Random Uniform Sampling

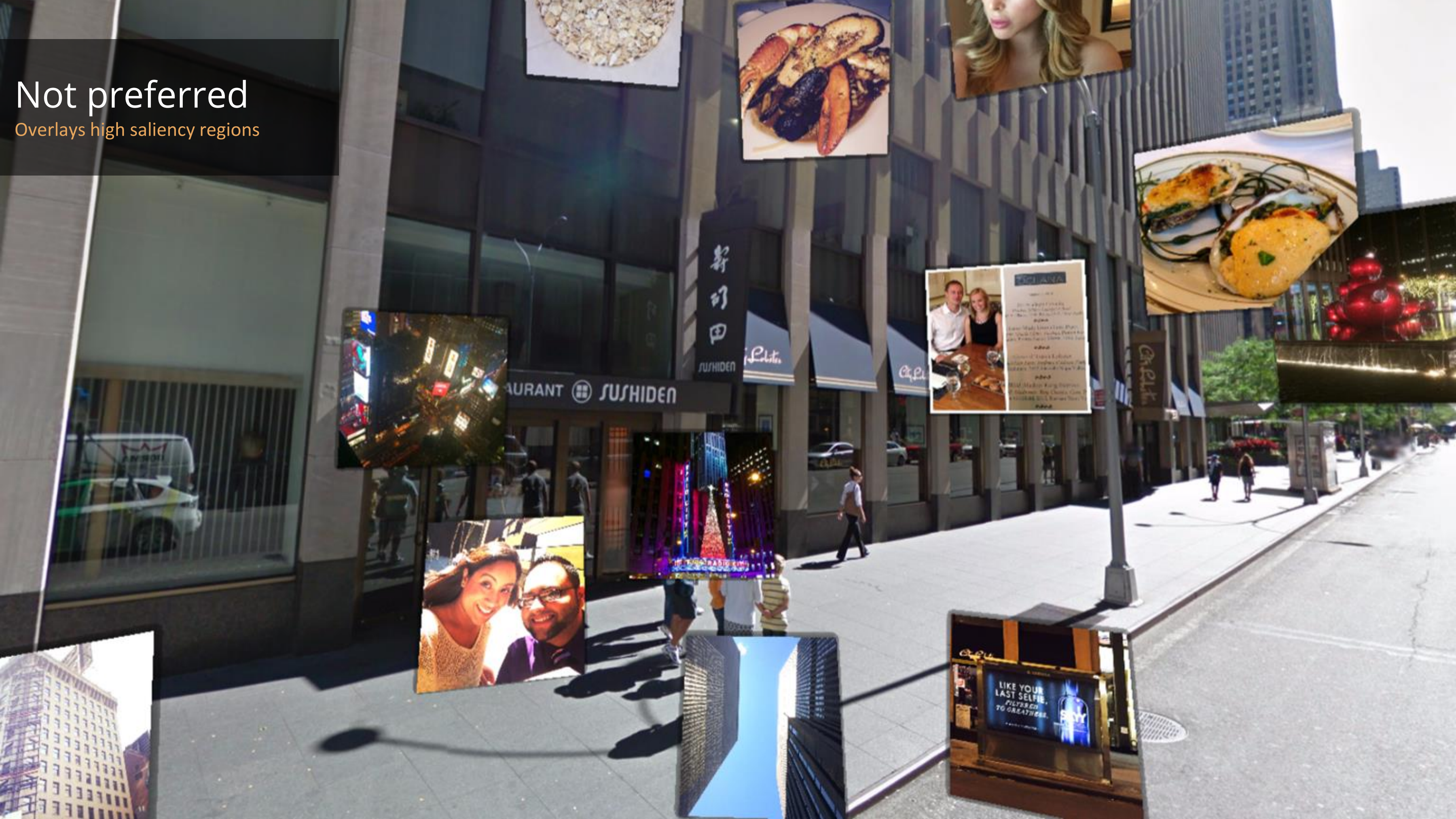


$$\varphi_i = (u_i - \frac{1}{2})\pi,\ \lambda_i = (2v_i - 1)\pi$$

$$x_i = \cos\varphi_i \cos\lambda_i,\ y_i = \sin\varphi_i,\ z_i = \cos\varphi_i sin\lambda_i$$

Not preferred

Overlays high saliency regions



Add depth map

Remove sky and ground (most)

$$\Omega_s = \{q_i \mid \forall q_i \in \Omega \wedge d_i = \infty\}$$

$$\forall \tilde{p}_i \in T, D_{min} < \tilde{d}_i < D_{max}$$

Add depth map

Remove sky and ground (most)

(a) Uniform Random Sampling on Sphere

(b) After Adding Depth Map

(c) Depth Map

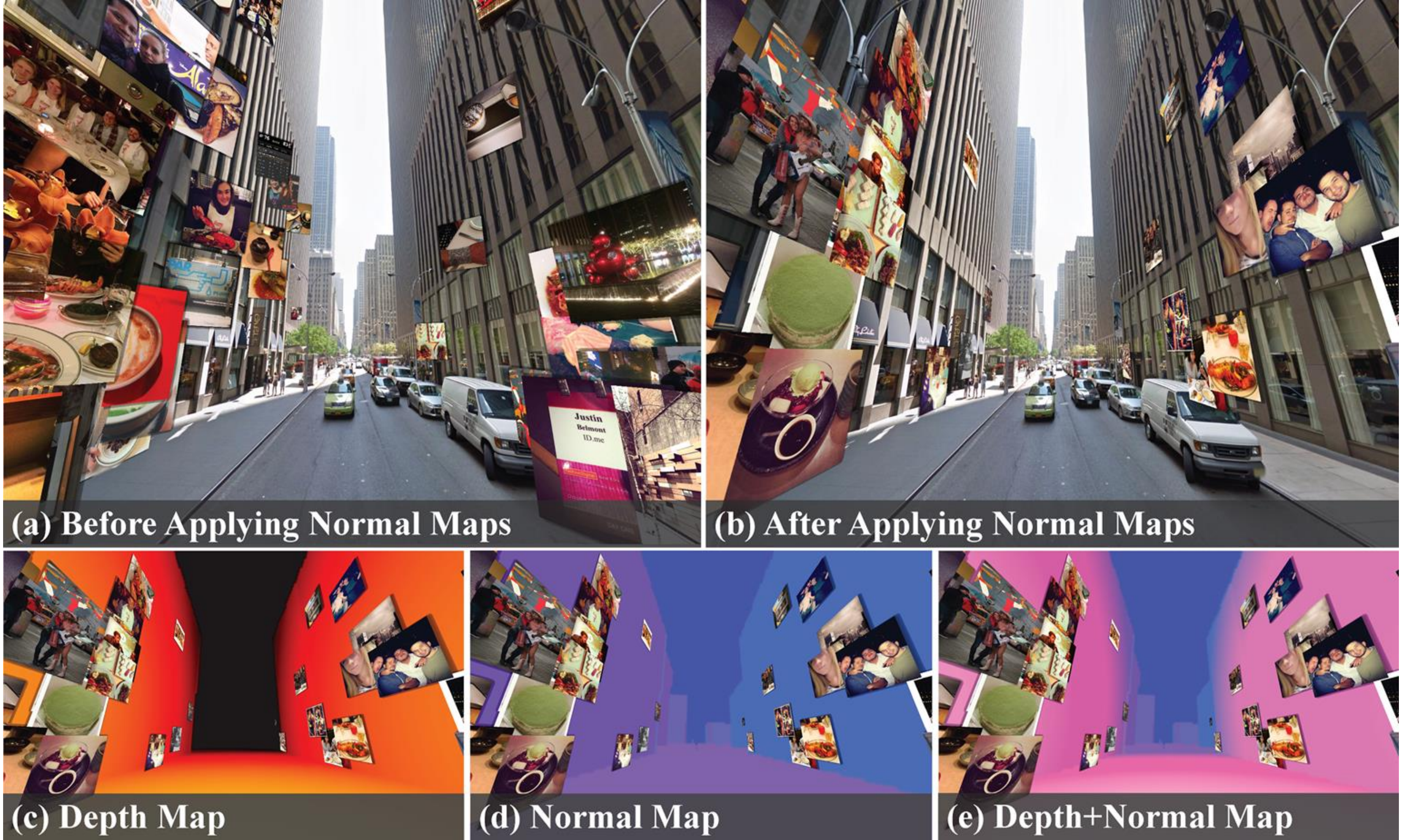
(d) Depth Overlay

Can we ensure each image aligns with the building geometries?

Add normal map

Remove all ground, align images

$$\Omega_g = \{q_i \mid \forall q_i \in \Omega \wedge \|\mathbf{n}_i - \mathbf{n}_g\| < \delta\}$$


(a) Before Applying Normal Maps
(b) After Applying Normal Maps
(c) Depth Map
(d) Normal Map
(e) Depth+Normal Map
Justin
Belmont
ID.me

Can we reduce visual clutter and occlusion?

Maximal Poisson-disk Sampling

Gamito et al. Remove visual clutter and occlusion

$$\forall \tilde{p}_i \in \tilde{P}, \tilde{P} \subseteq T, \forall S \subseteq \Omega : Pr(\tilde{q}_i \in S) = \int_S \mathrm{d}i \tag{10}$$

$$\forall \tilde{p}_i, \tilde{p}_j \in \tilde{P}, \tilde{p}_i \neq \tilde{p}_j : \|p_i - p_j\| \geq r \tag{11}$$

$$S(X) = \{\tilde{p}_j \in T : \|\tilde{p}_i - \tilde{p}_j\| \geq r, \tilde{p}_i \in \tilde{P}\} : S(X) = \emptyset \tag{12}$$

Dart-throwing Algorithm

PixelPie by Ip et al. using vertex and fragment shaders

image courtesy:
PeterPan23@Wikipedia

Pixel-Pie Algorithm

Remove when occlusion occurs

Algorithm 1 Maximal Poisson-disk sampling by dart-throwing

Input: The minimum distance r between sampled points
Output: A set $\tilde{P}$ of points which satisfy equation (10)-(12)

Set $\tilde{P} \leftarrow \emptyset$, empty region $\tilde{R} \leftarrow T$
repeat
  Generate some random points $\tilde{P}' \subseteq R$ by rasterizing them as circular disks into a depth map in vertex shader.
  Remove any point $\tilde{p} \in \tilde{P}'$ whose corresponding point $\tilde{q}$ violates $\tilde{q} \in \Omega_g \vee D_{min} < \tilde{d} < D_{max}$
  Identify and remove the occluded disks from $\tilde{P}$ by reading the depth map in the shader.
  $\tilde{P} \leftarrow \tilde{P} \cup \tilde{P}'$
  Update the empty region R in the fragment shader.
until $R \leftarrow \emptyset$

Project Social Media Pictures

By Maximal Poisson-disc Sampling

ALGORITHM 1: Social Media Layout using Poisson-disk Samples

Input: N sorted social media images $\hat{S} = \{s_i \mid i = 1 \ldots N\}$, acquired from SSV servers.

Output: A set of image planes to display social media: $I = I_1 \ldots I_M, M \leq N$.

Generate the set of candidate sample points $\tilde{\mathbf{P}}$ by the PixelPie algorithm;

Sort points in $\tilde{\mathbf{P}}$ in descending order according to their corresponding values in the depth map D so that the closest sample point is laid out first;

Set $I \leftarrow \emptyset$;

for $i \leftarrow 1 \ldots \min(N, |\tilde{P}|)$ **do**

  Place I_i with texture from $s_i \in \hat{S}$ at the projected position $\tilde{\mathbf{q}}_\mathbf{i} \leftarrow \mathscr{P}(\tilde{\mathbf{p}}_\mathbf{i})$;

  Rescale I_i according to the corresponding depth value: $\tau_i \leftarrow \tau / d_i$ for perspective visual effects;

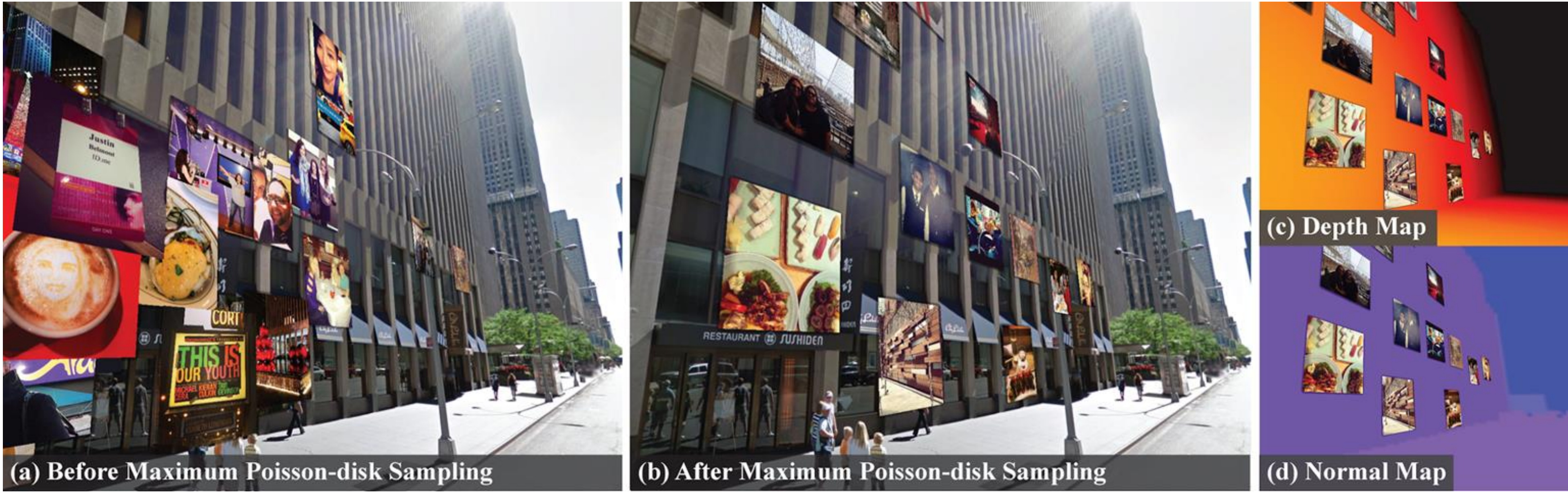
  Rotate I_i so that it is perpendicular to the normal vector $\mathbf{n_i} \leftarrow \mathbf{N}(u_i, v_i)$;

  Add I_i to the result set: $I \leftarrow I \cup I_i$;

end

Sampling Comparison

Remove visual clutter and occlusion



(a) Before Maximum Poisson-disk Sampling
(b) After Maximum Poisson-disk Sampling
(c) Depth Map
(d) Normal Map

Scenic Landscapes

Using orientation of the road

ALGORITHM 2: Social Media Layout using Road Orientations

Input: $|O|$ road orientations with $o_i \in [0, 2\pi]$. K social media to be placed for each orientation. Typically, $|O| = 2$ for a road with two orientations.

Output: A set of image planes to display social media: $I = I_1 \ldots I_M, M \geq K \cdot |O|$.

Set $I \leftarrow \emptyset$;

for $i \leftarrow 1 \ldots |O|$ **do**

  Set the position $\mathbf{q_i} \leftarrow (KR\cos o_i,\ h,\ KR\sin o_i)$ at height h and radius R;

  (Optional based on user's preference) Add a frontal image plane to I at $\mathbf{q_i}$;

  Set the translation $\mathbf{t} \leftarrow (T\cos(o_i + \frac{\pi}{2}), 0, T\sin(o_i + \frac{\pi}{2}))$ with constant T;

  for $k \leftarrow 1 \ldots K$ **do**

    Set $\tilde{\mathbf{q}} \leftarrow (kR\cos o_i,\ h,\ kR\sin o_i)$;

    Add a left side image plane to I at position $\mathbf{q}' \leftarrow \tilde{\mathbf{q}} + \mathbf{t}$;

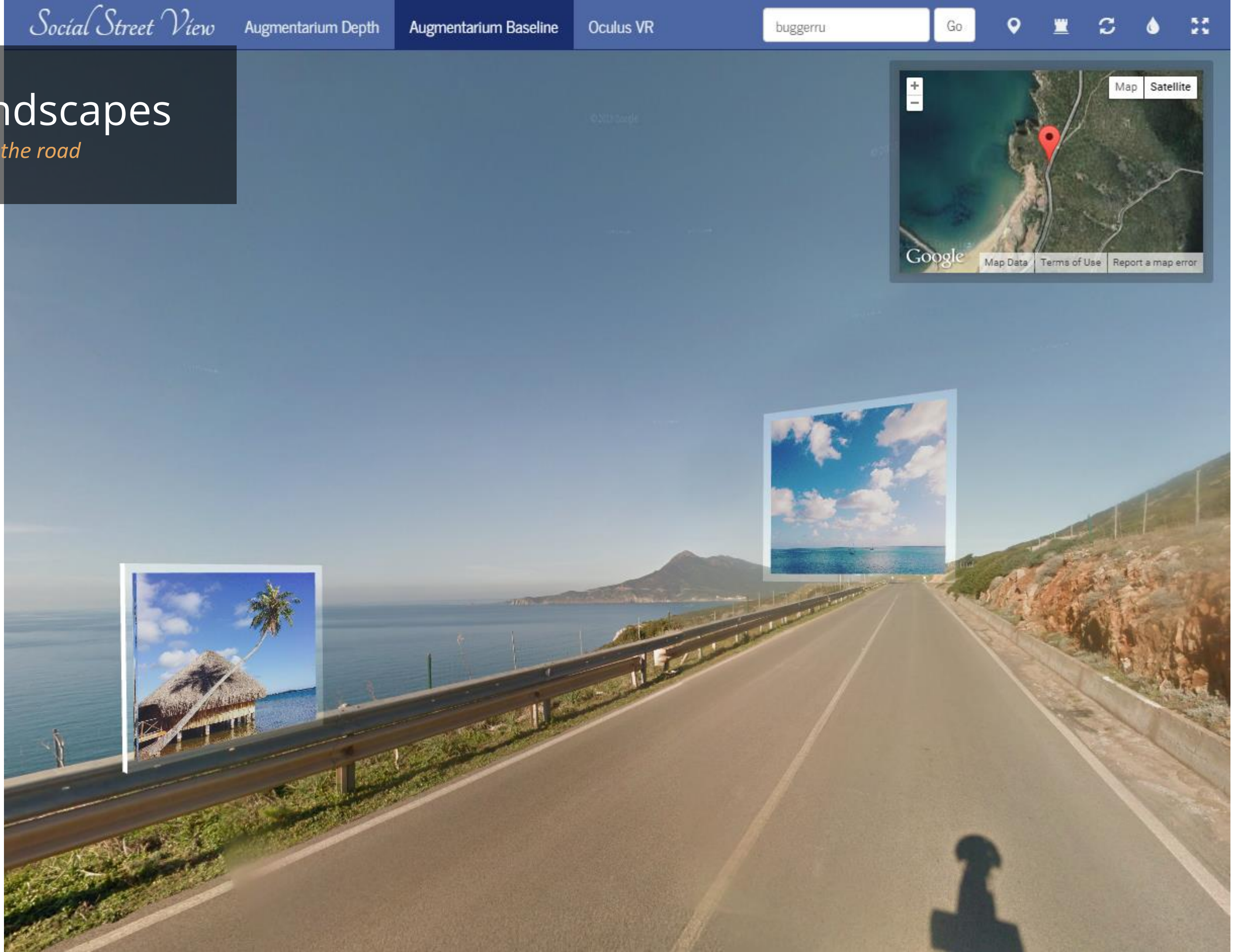
    Add a right side image plane to I at position $\mathbf{q}' \leftarrow \tilde{\mathbf{q}} - \mathbf{t}$;

  end

end

Scenic Landscapes

Using orientation of the road



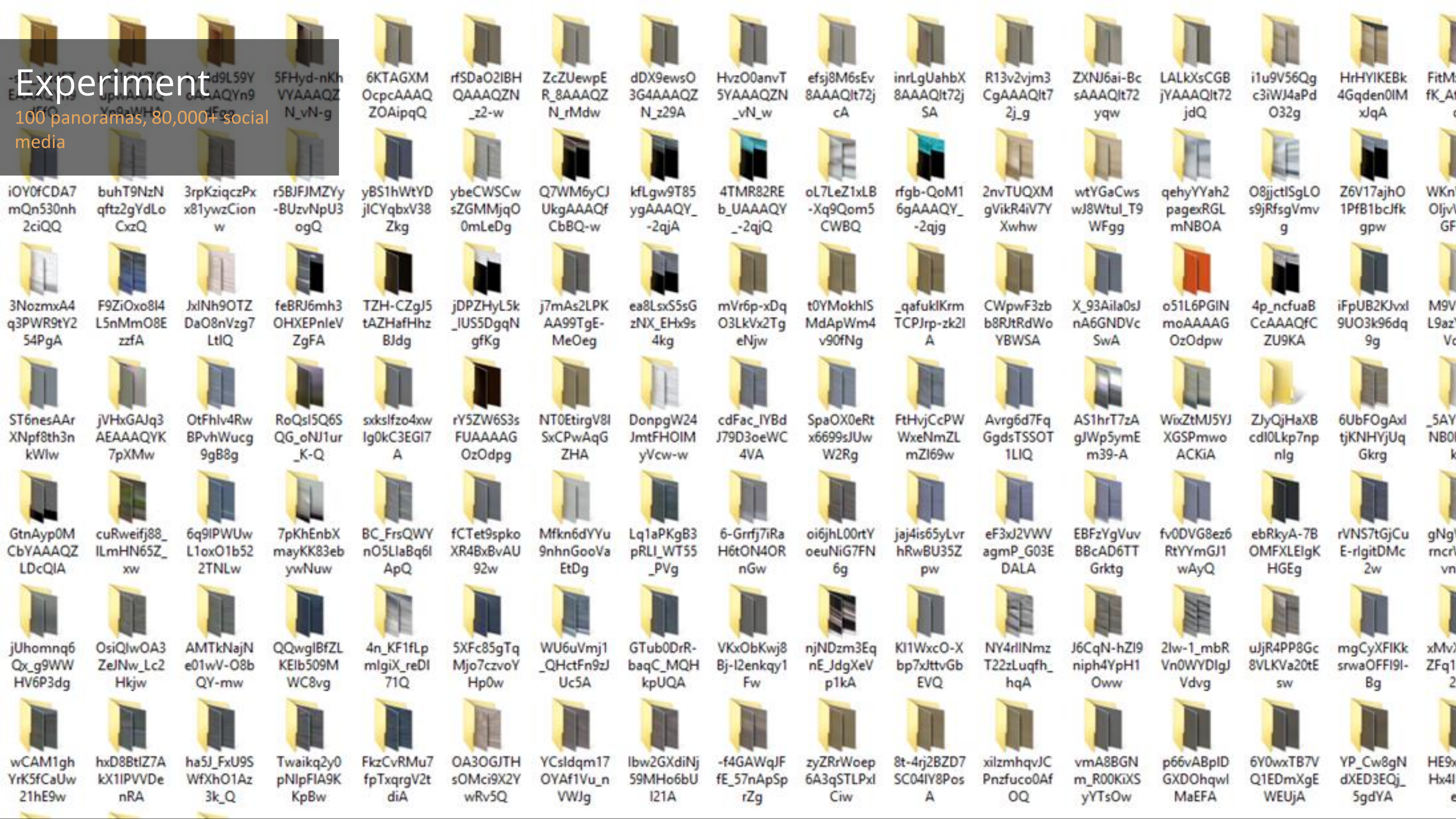
Scenic Landscapes

Using orientation of the road

In addition, our system allows users to walk around and explore live social media streams.

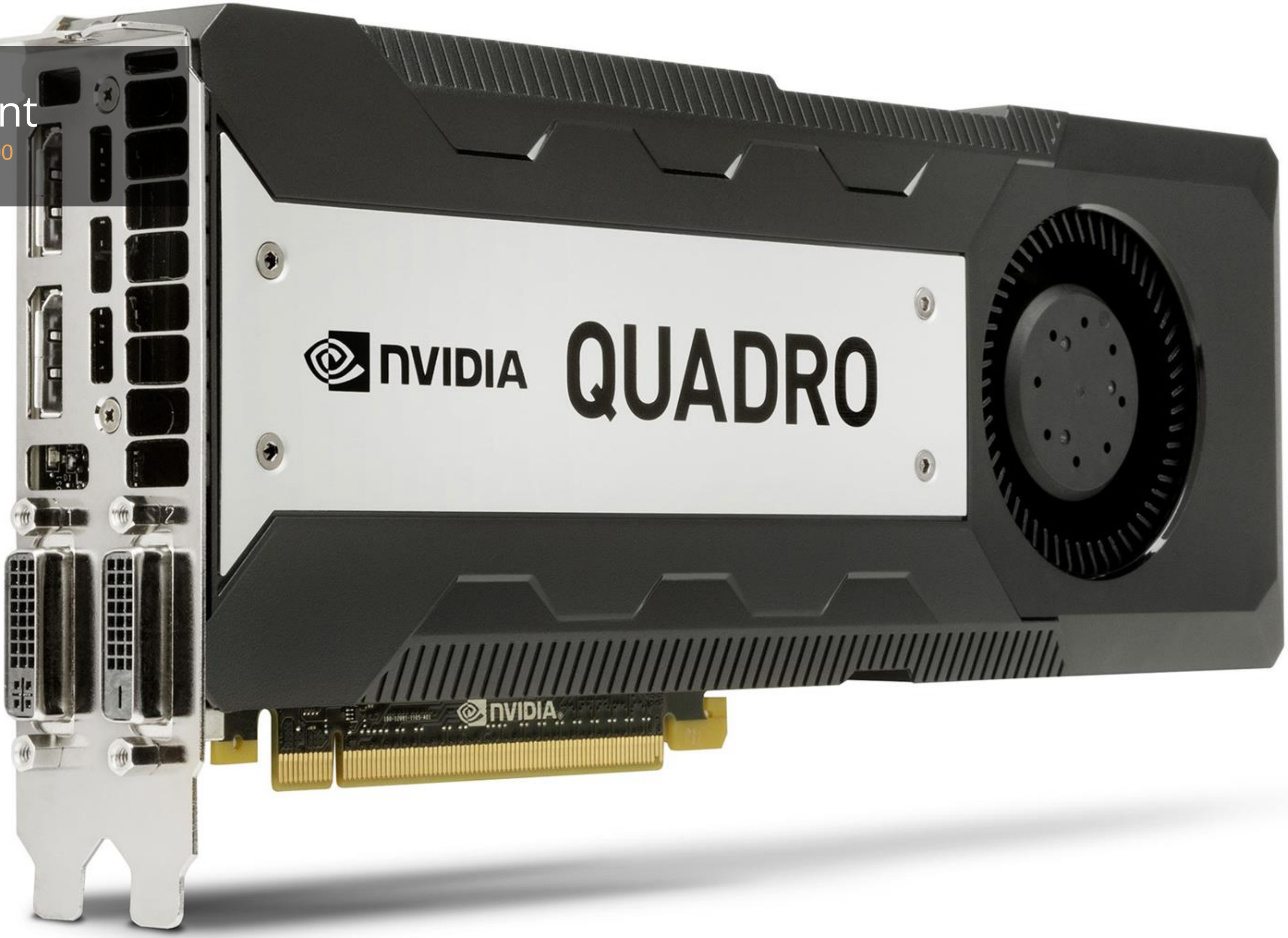
Experiment

100 panoramas, 80,000+ social media



Experiment

Nvidia Quadro K6000



Experiment

Conditions of panoramic images

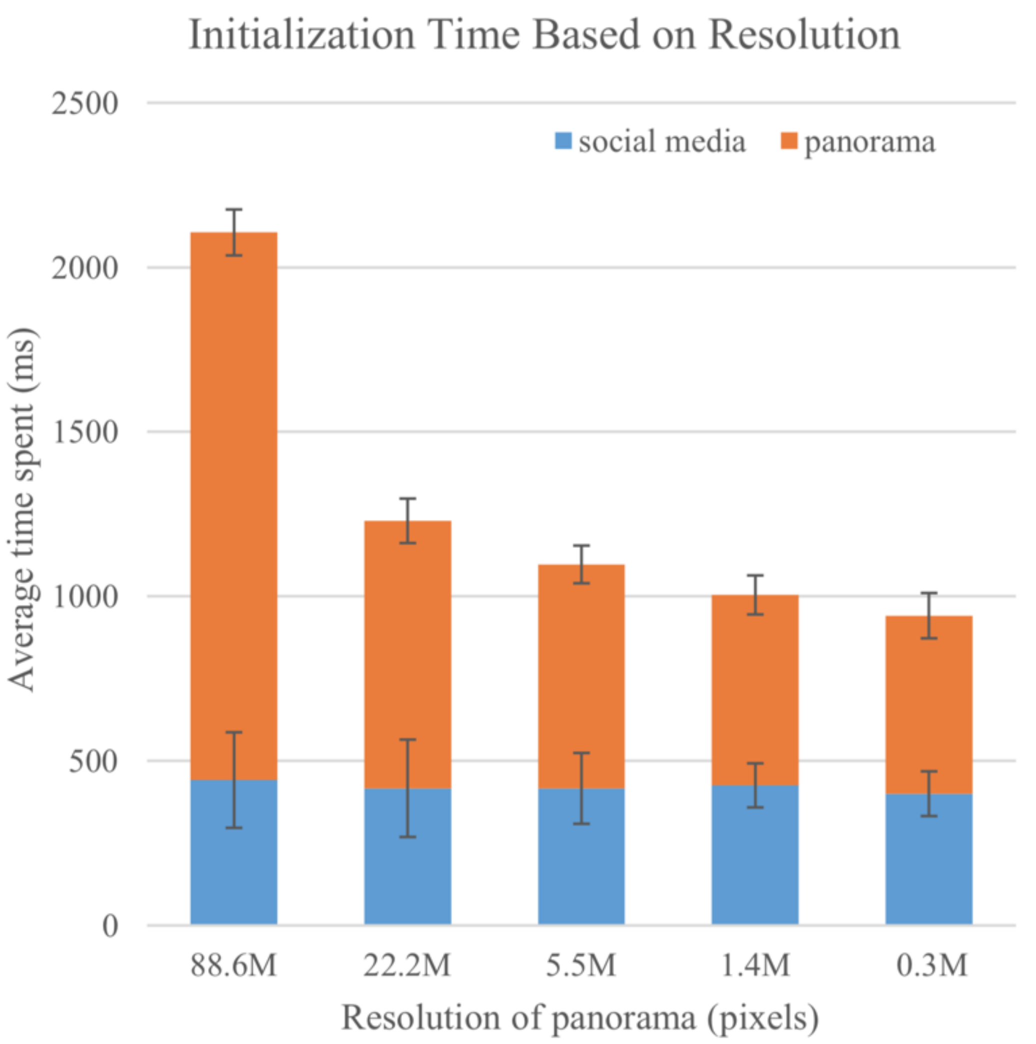
Immersive high-resolution screens

Pixels	Resolution	Number of tiles	File size
88.6M	13312×6656	26×13	$\sim 5M$
22.2M	6656×3328	13×7	$\sim 2M$
5.5M	3328×1664	7×4	$\sim 800K$
1.4M	1664×832	4×2	$\sim 300K$
0.3M	832×416	2×1	$\sim 90K$

Common Consumer-level Displays

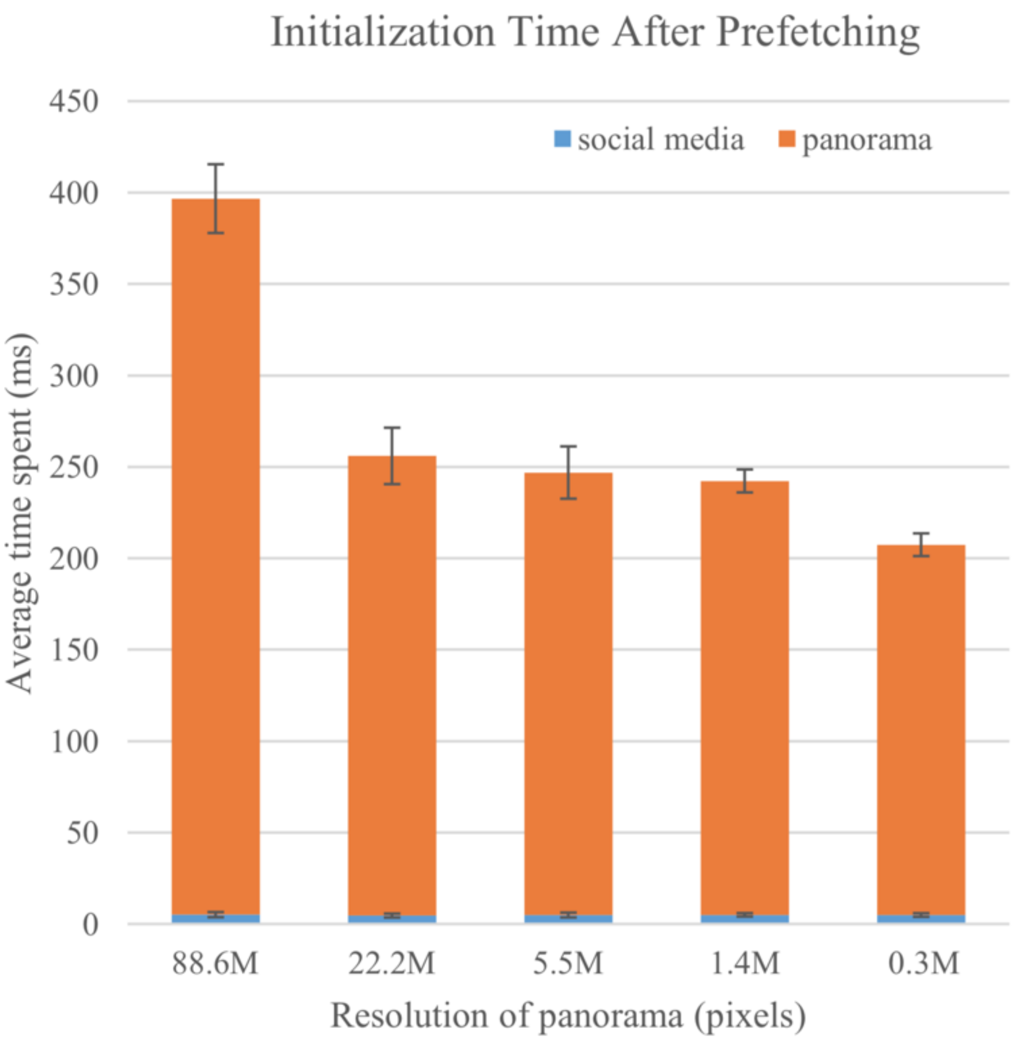
Initialization Time

Panorama takes a while to load



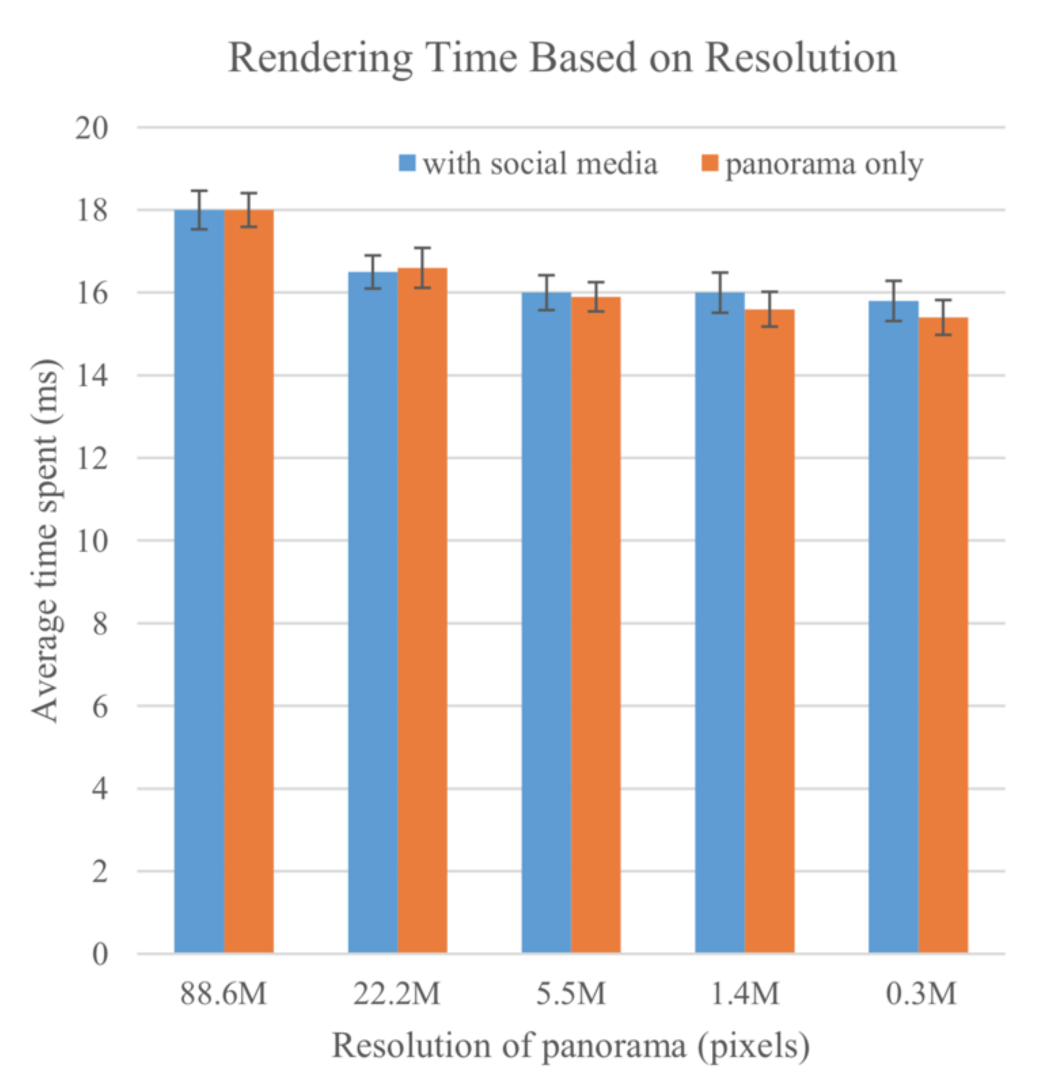
After Prefetching

¾ - ⅘ time reduced



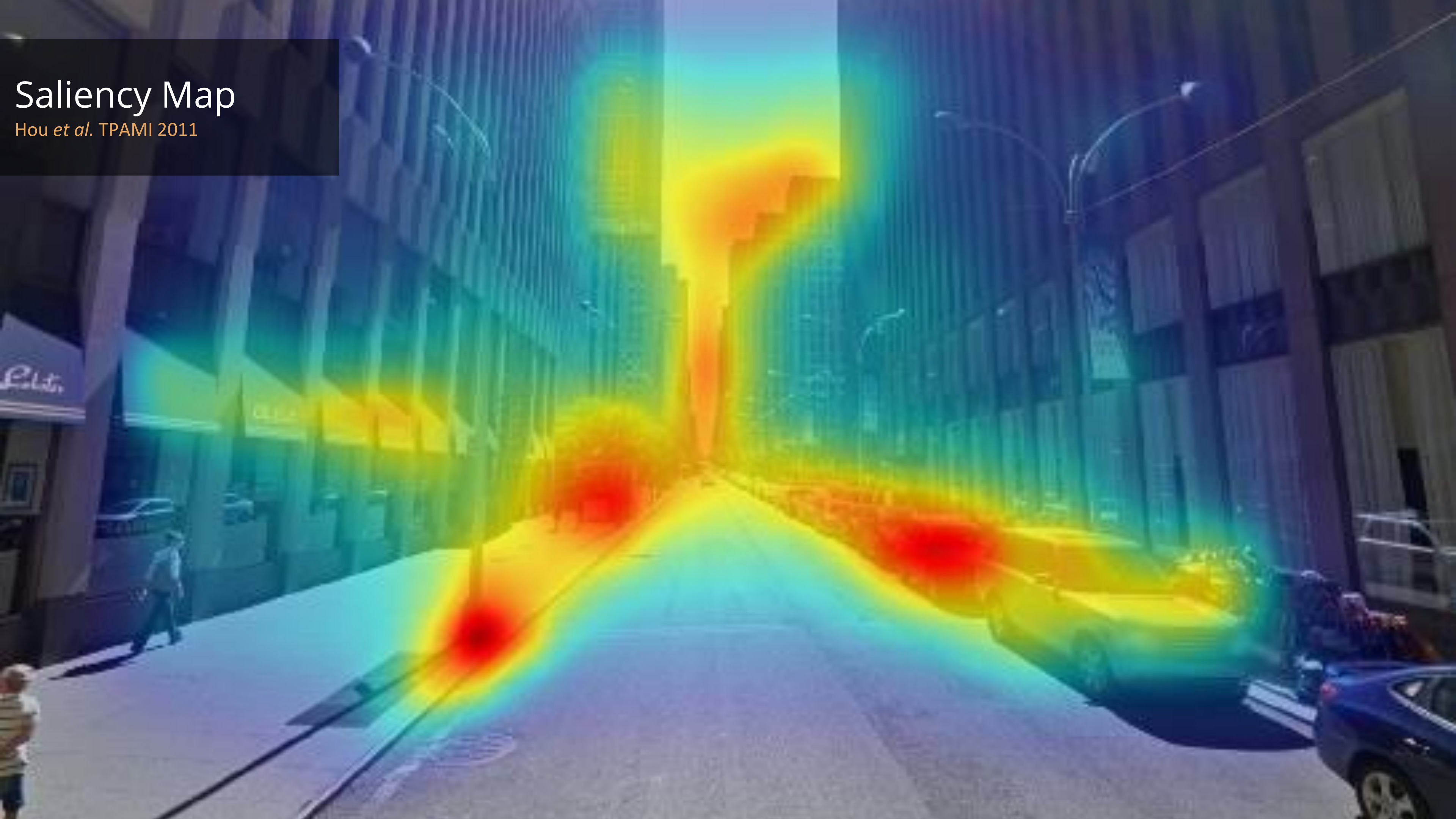
Rendering Time

Almost 58~60 FPS



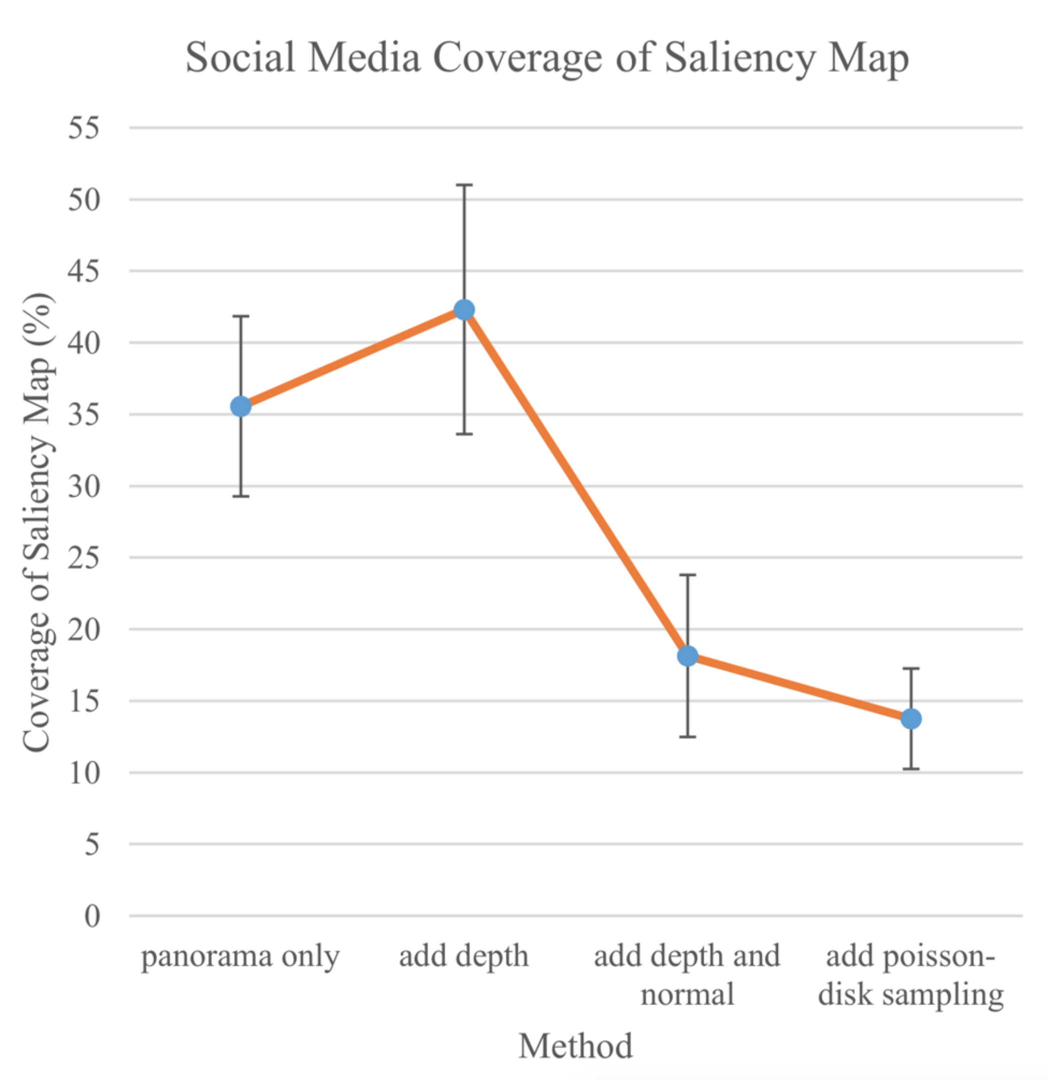
Saliency Map

Hou *et al.* TPAMI 2011



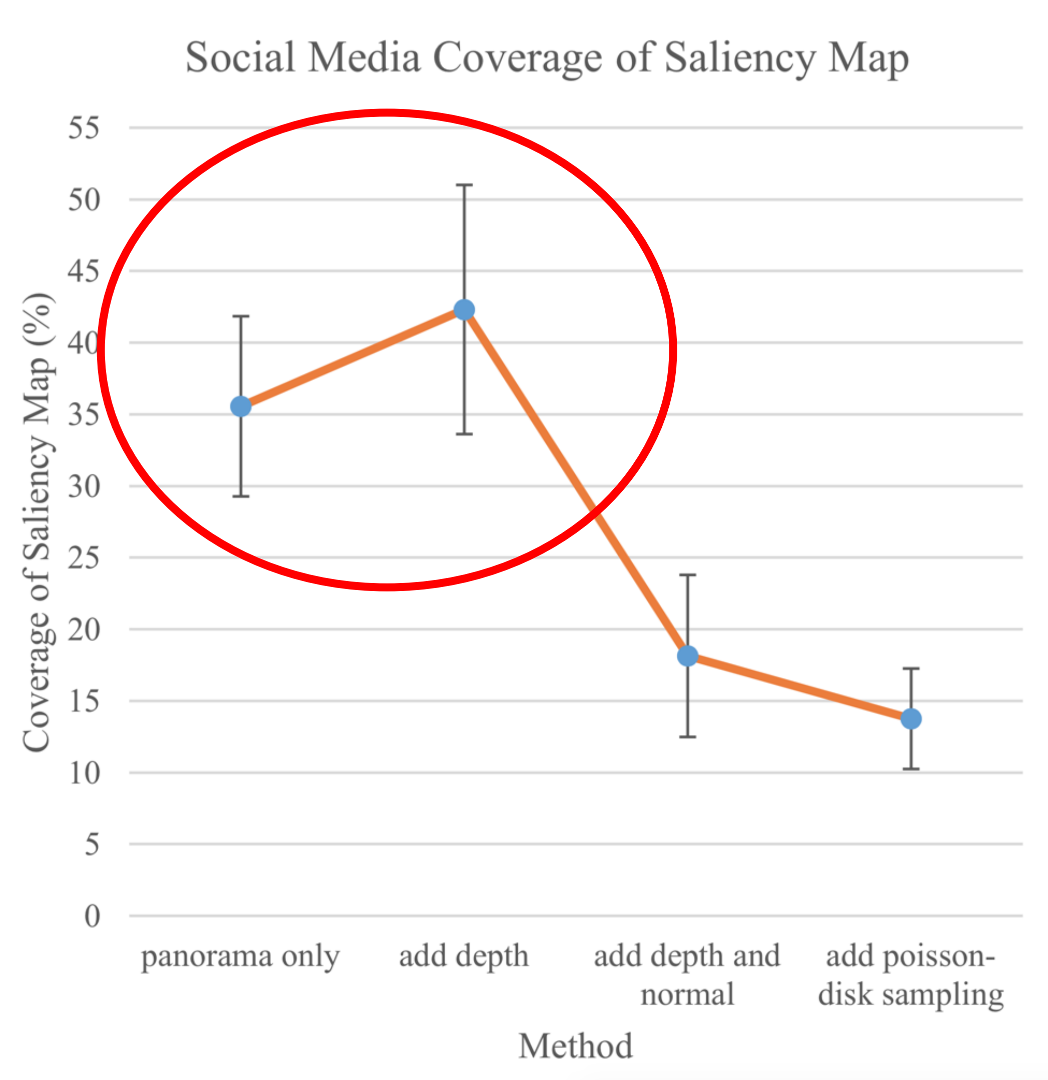
Social Media Coverage

100 panoramas for each algorithm



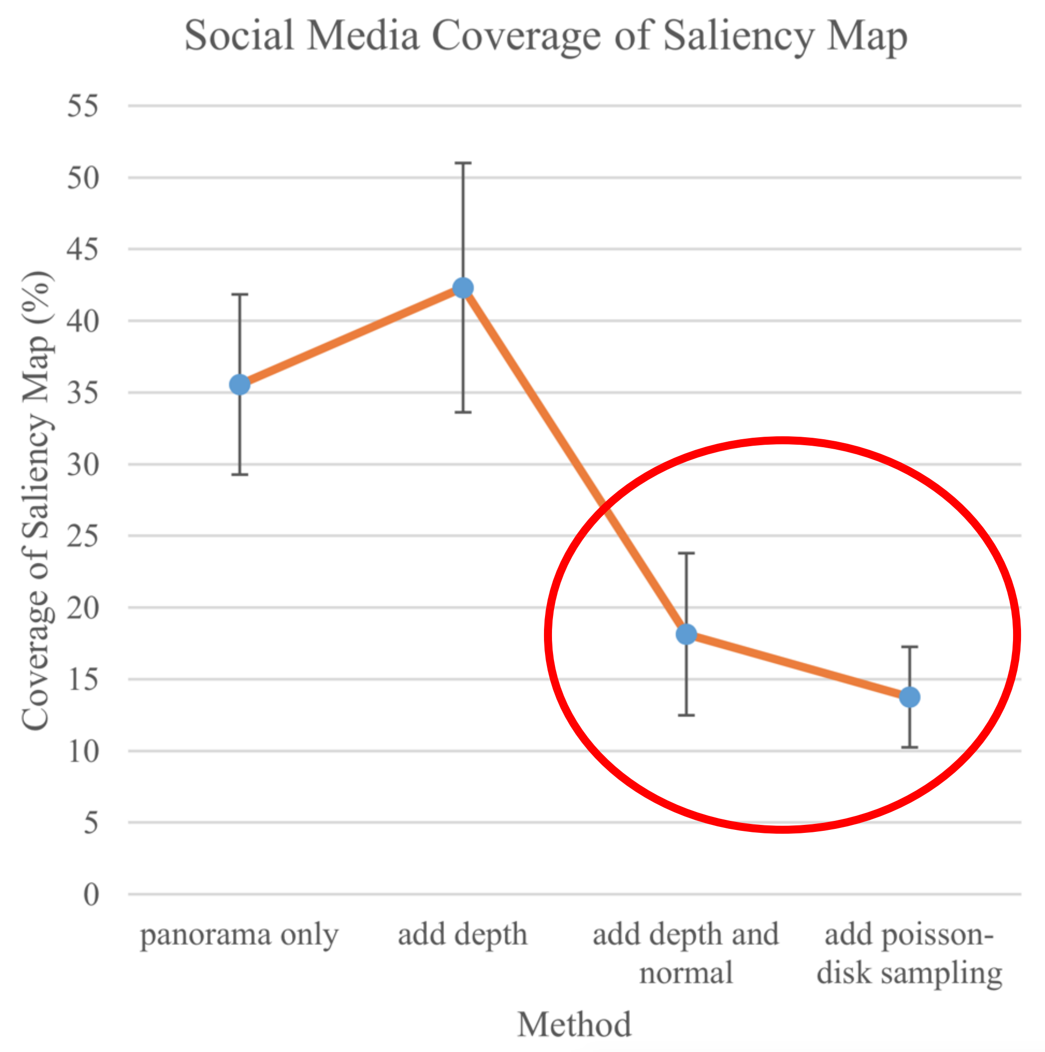
Social Media Coverage

100 panoramas for each algorithm



Social Media Coverage

100 panoramas for each algorithm



Potential Applications?

Stuck in traffic on our way to Cabo with this awesome view

#roadtrip #cabo #view #mexico

Daniela on *Instagram*
July 12, 2014

Application

Immersive story telling

However, we can hardly enjoy the view given the small posted image.

Business Advertising

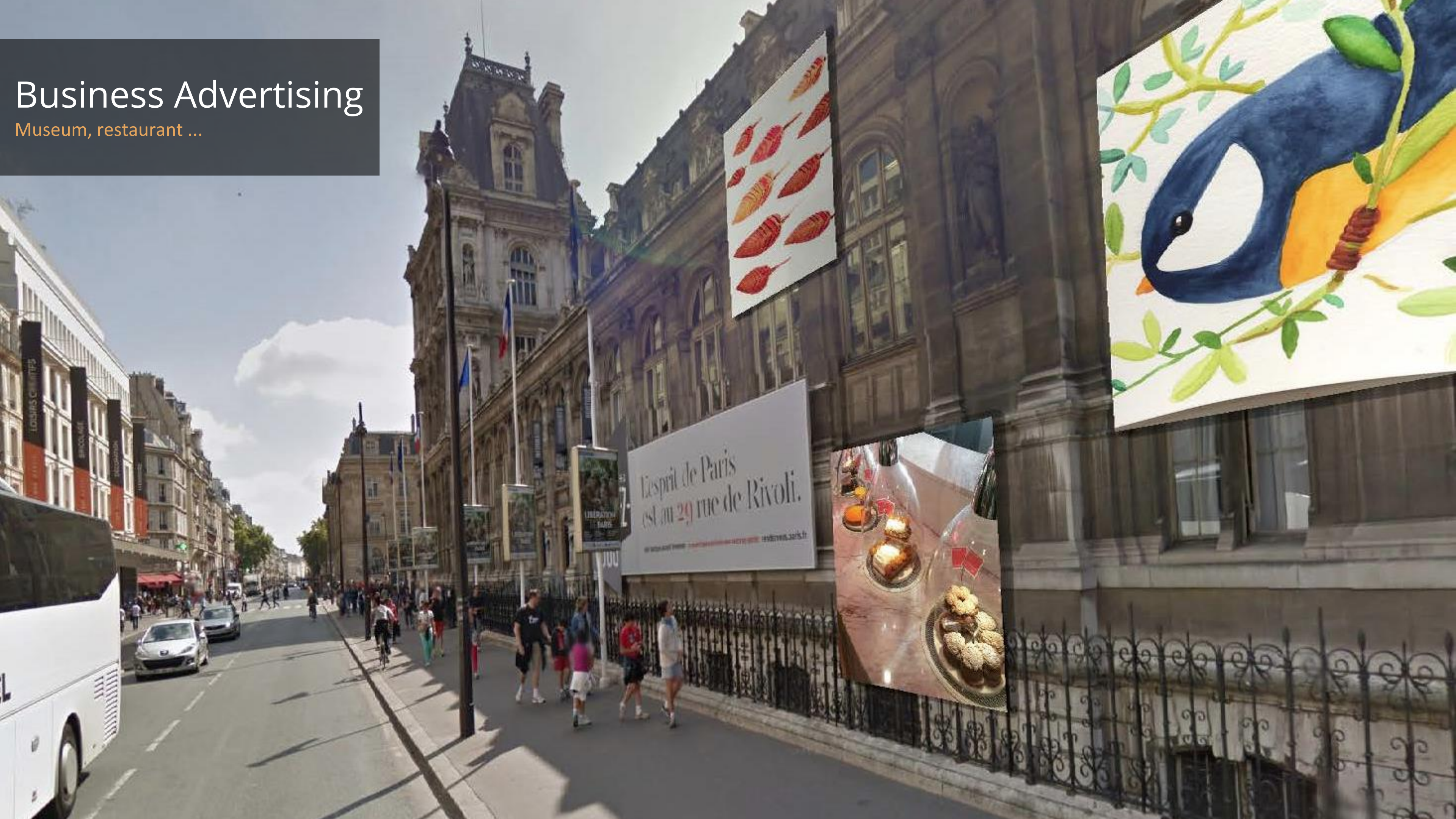
Museum, restaurant, real-estate ...

… dinner started off with *amazing oysters* paired with my favorite Ruinart blanc de blancs champagne

By frankiextah on *Instagram*

Business Advertising

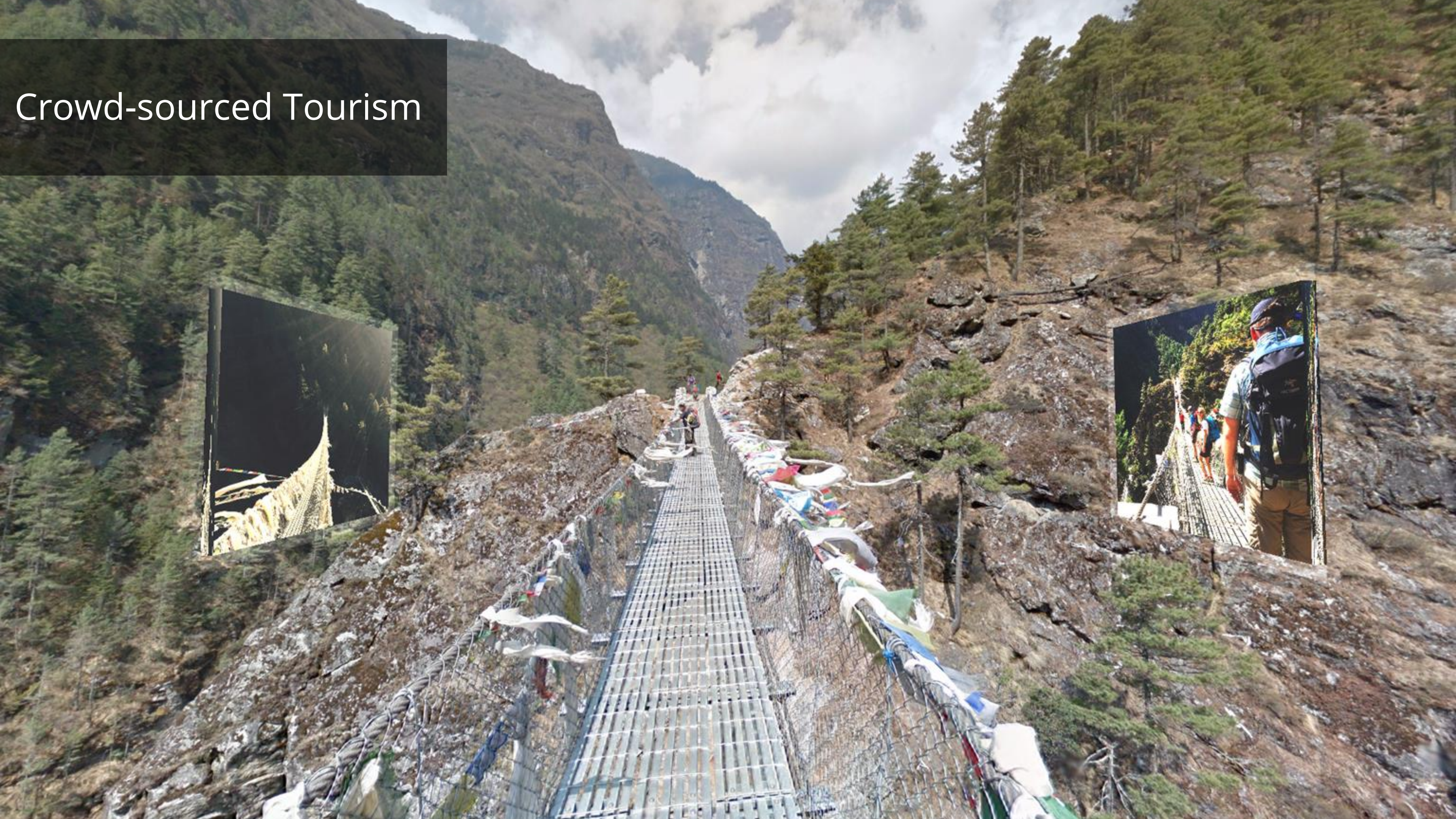
Museum, restaurant ...



Learning Culture

Taierzhuang, Chinese Spring Festivals

Crowd-sourced Tourism




Crowd-sourced Tourism

What if …

Temporal information is used for filtering and rendering?

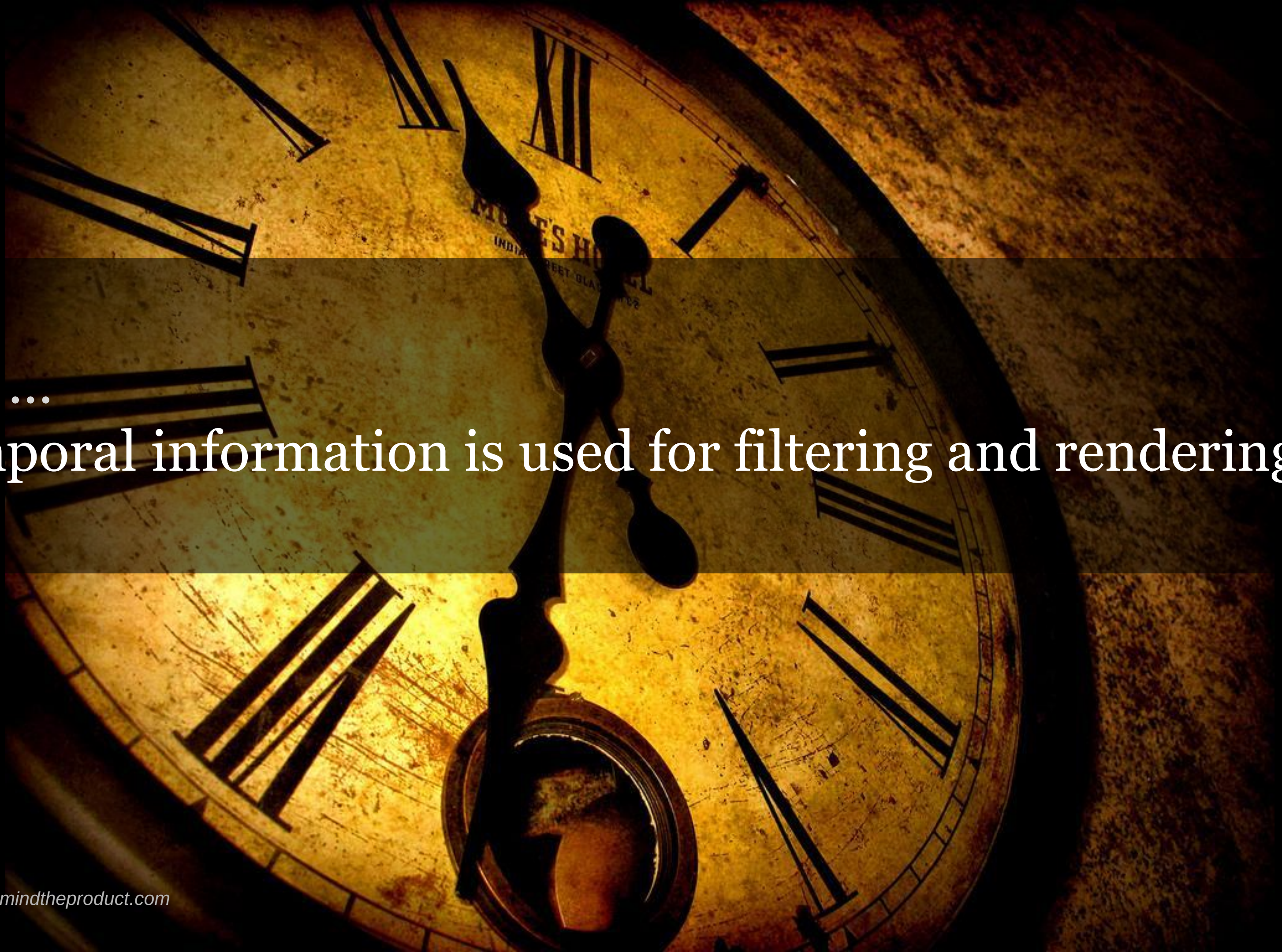
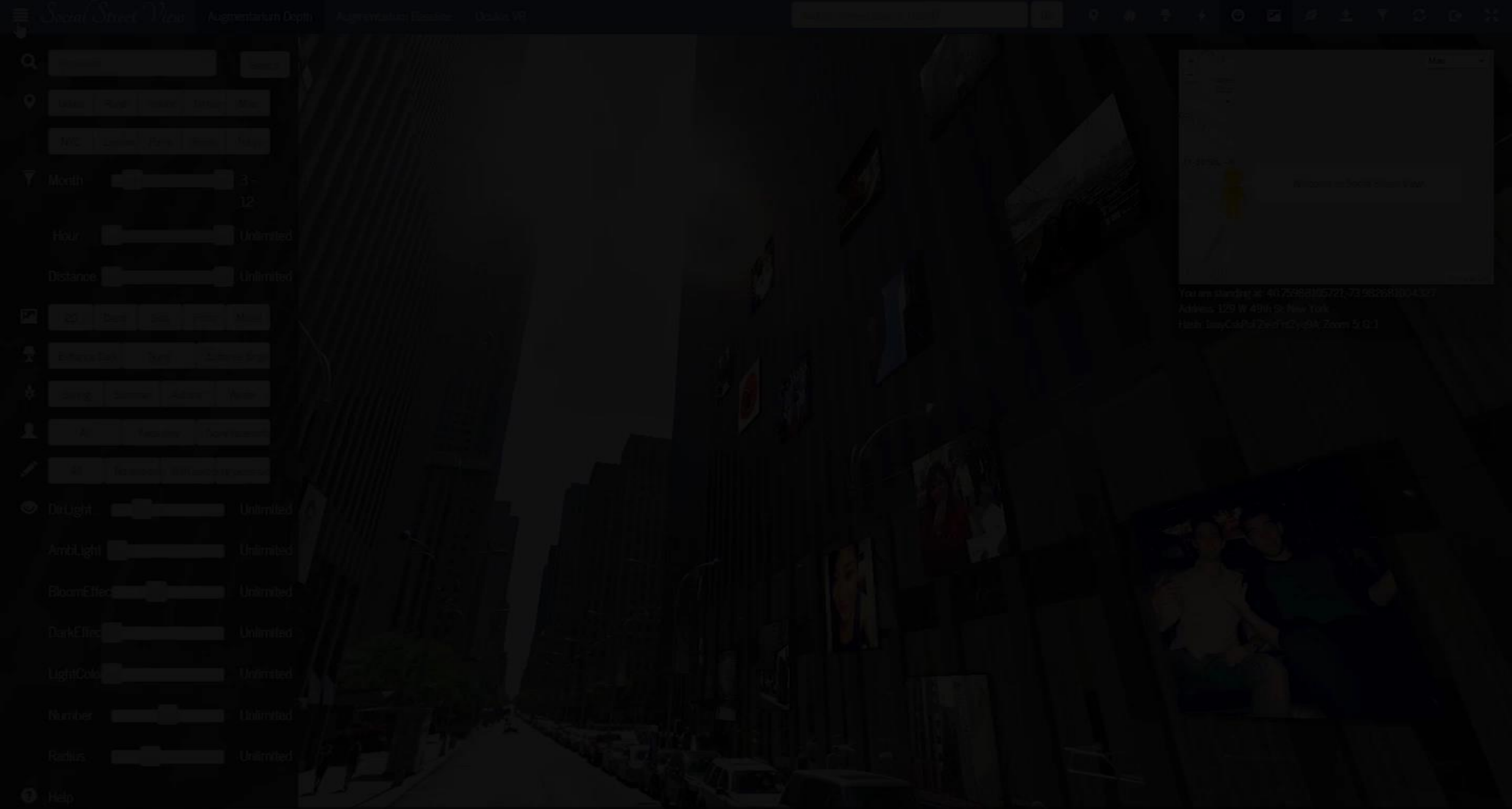


image courtesy: mindtheproduct.com

Visual Enhancement

Spatio-Temporal Augmented Reality

image courtesy: webneel.com



Users can use temporal or distance filters to narrow down their interested queries.

Seasonal Effect

Winter

Seasonal Effect

Winter

Seasonal Effect

Spring

Seasonal Effect

Autumn in Paris

Failure Cases

A Square in London

Future Work

Digital City by 3D Reconstruction, Depth Fusion and Augmented Reality

image courtesy: wallpapervortex.com

Future Work

Incorperate with Open3D in Web3D 2016?

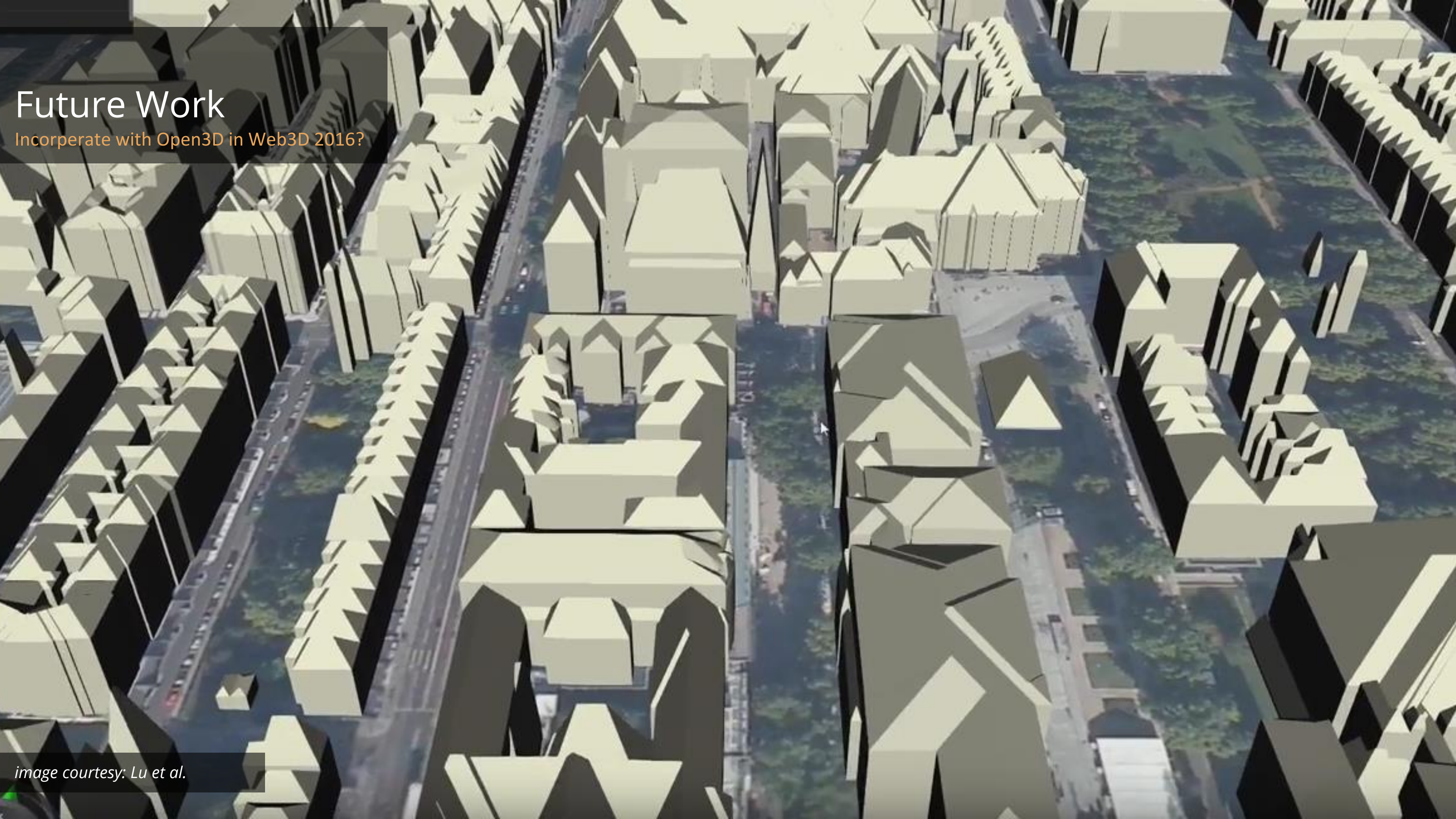
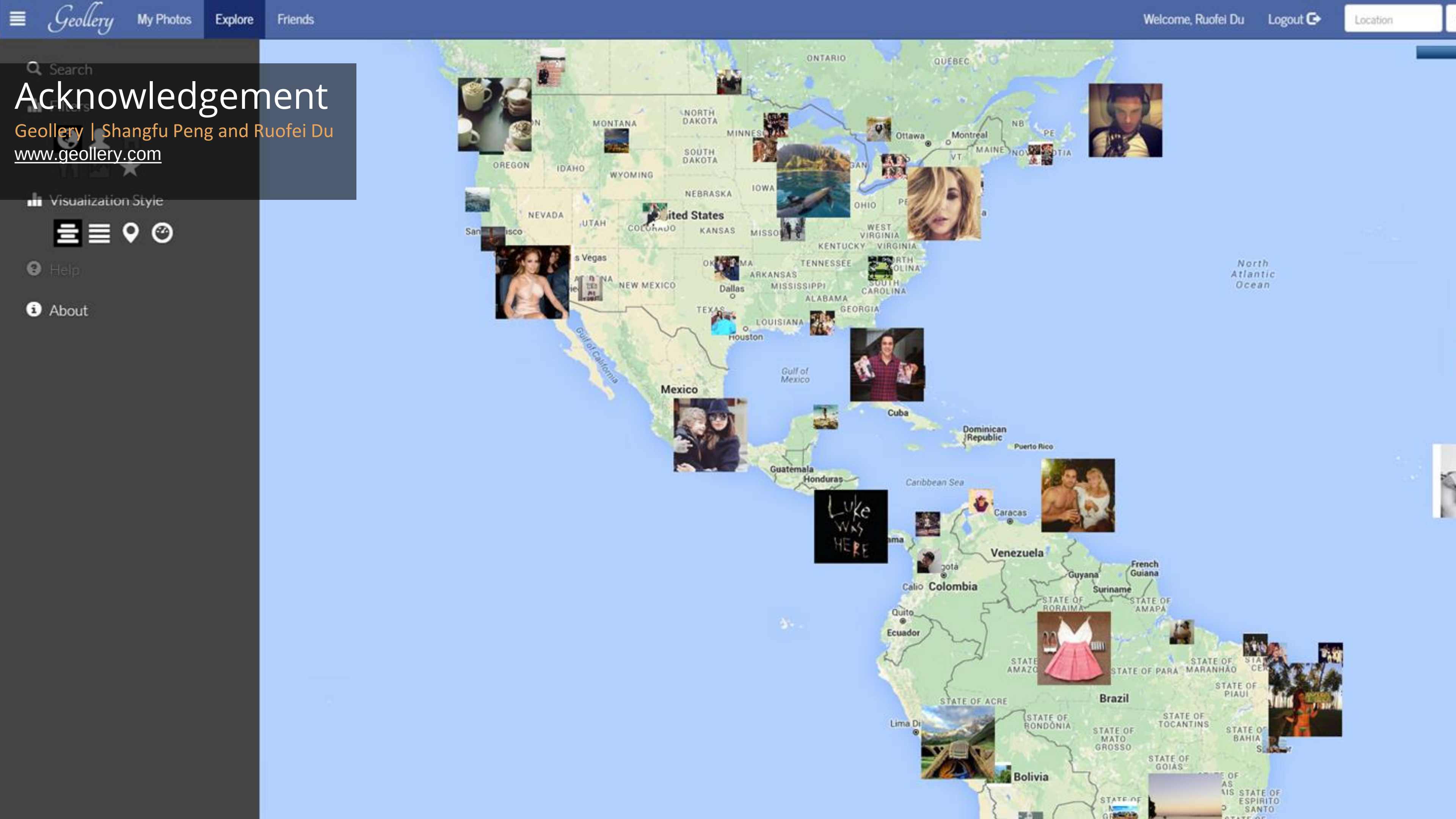


image courtesy: Lu et al.

Acknowledgement

Geollery | Shangfu Peng and Ruofei Du

www.geollery.com



Acknowledgement

Augmentarium Lab | GVIL | UMIACS

Acknowledgement

NSF | Nvidia | MPower | UMIACS

NSF

NVIDIA

UMIACS
University of Maryland Institute for Advanced Computer Studies

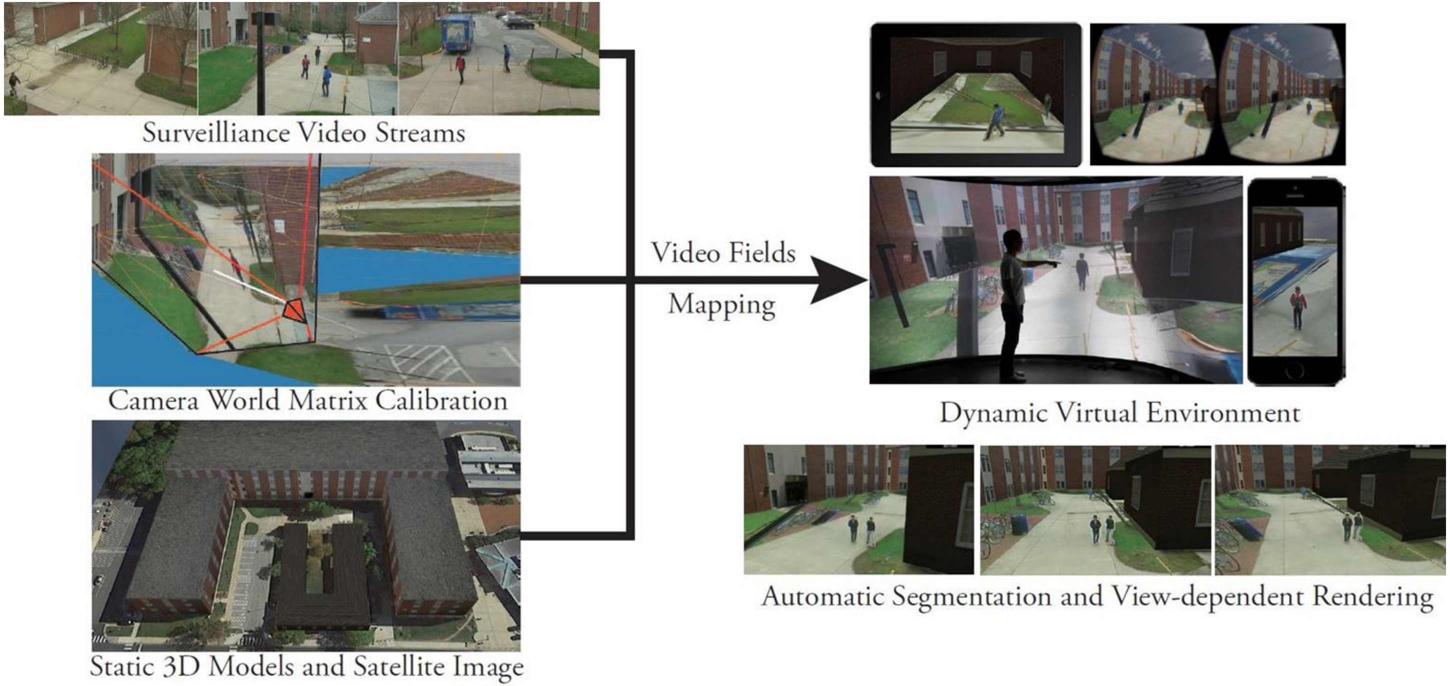
UNIVERSITY OF MARYLAND
MPOWERING THE STATE

UNIVERSITY OF MARYLAND

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Video Fields

Web3D Sunday 9:20 - 9:40 AM



Social Street View

www.SocialStreetView.com

Thank you! Any questions or comments are welcome!

Ruofei Du and Amitabh Varshney

Augmentarium Lab | GVIL | UMIACS

Web3D 2016

Social Street View
Backup Slides

Future Work

Saliency

- Low-level image saliency for layout
- Real-time or post-processed saliency maps

Future Work

User Study

- High-resolution large-area screens
- Head-mounted displays

Future Work

3D Reconstruction and AR

- Real-time geometric fusion
- HoloLens

Future Work

3D Reconstruction and AR

- What is the (experimentally-justifiable) motivation for putting generic images into scenes?
- What is the relevance of which photos are displayed? And how this affects the worth?
- Whether it is possible to try and improve this relevance using vision techniques?

What is ignored in current social media sites?