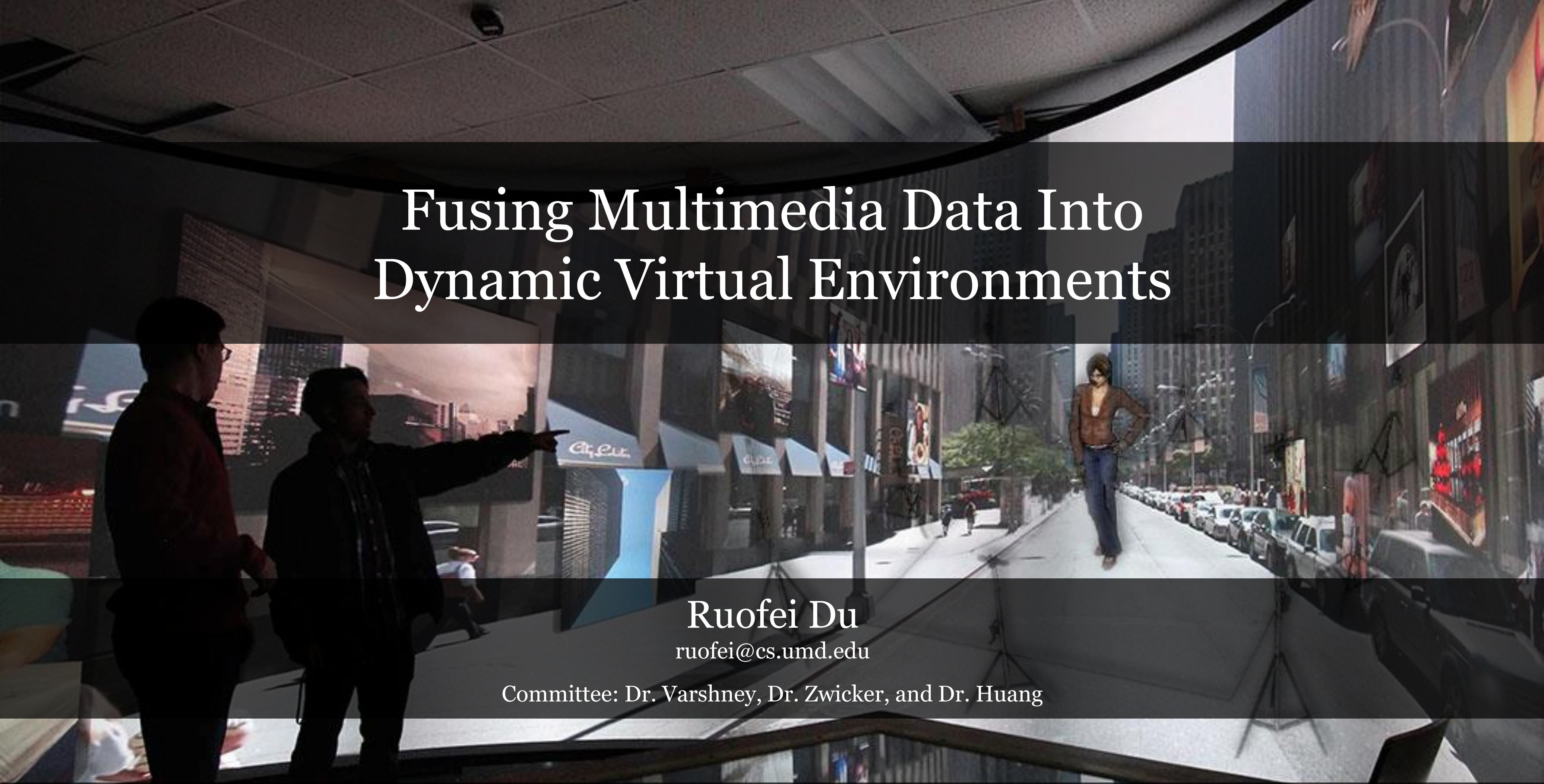


Fusing Multimedia Data Into Dynamic Virtual Environments

Ruofei Du
ruofei@cs.umd.edu

Committee: Dr. Varshney, Dr. Zwicker, and Dr. Huang

THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LABORATORY
AT THE UNIVERSITY OF MARYLAND
GVIL
UMIACS
COMPUTER SCIENCE
UNIVERSITY OF MARYLAND
UNIVERSITY OF MARYLAND

Motivation

Popularity of VR and AR devices

Motivation

Popularity of VR and AR devices

Motivation

Popularity of VR and AR devices

Motivation

Popularity of VR and AR devices

Motivation

Popularity of VR and AR devices

Motivation

Assorted VR and AR applications

Motivation

Assorted VR and AR applications

Motivation

Assorted VR and AR applications

These VR/AR applications have

Huge Data

requirements

These VR/AR applications have

Huge Data

requirements

Where is the
3D data going
to come from?

Motivation

Lack of contents

Motivation

Lack of contents

These VR/AR applications have

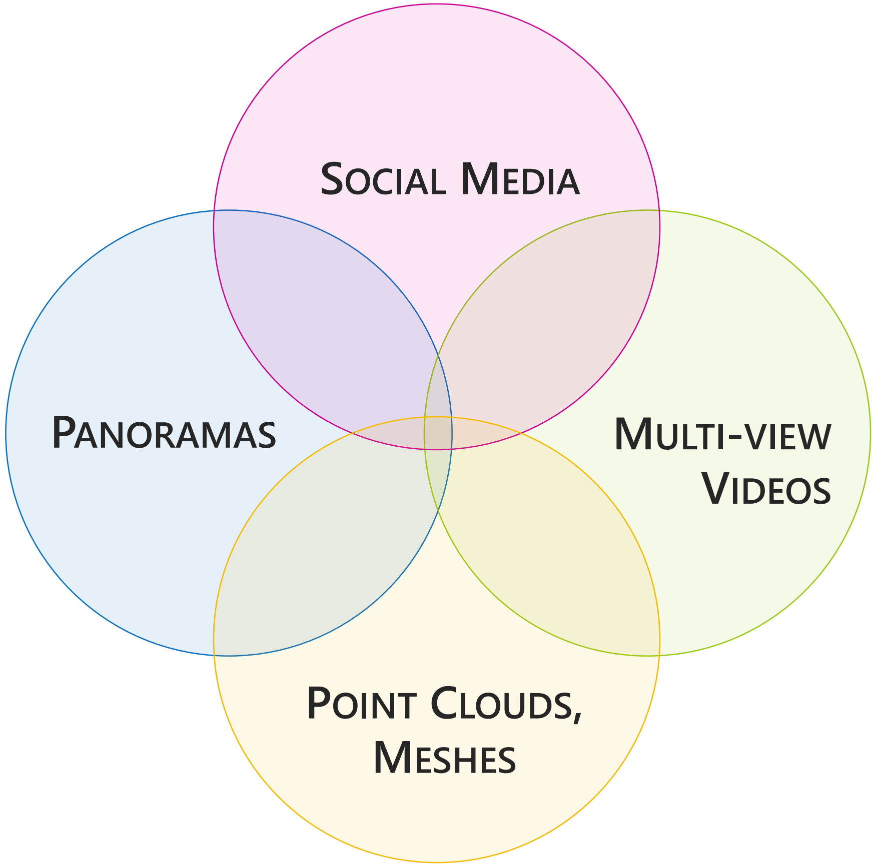
Huge Data

requirements

Automatically fusing multimedia data into virtual environments

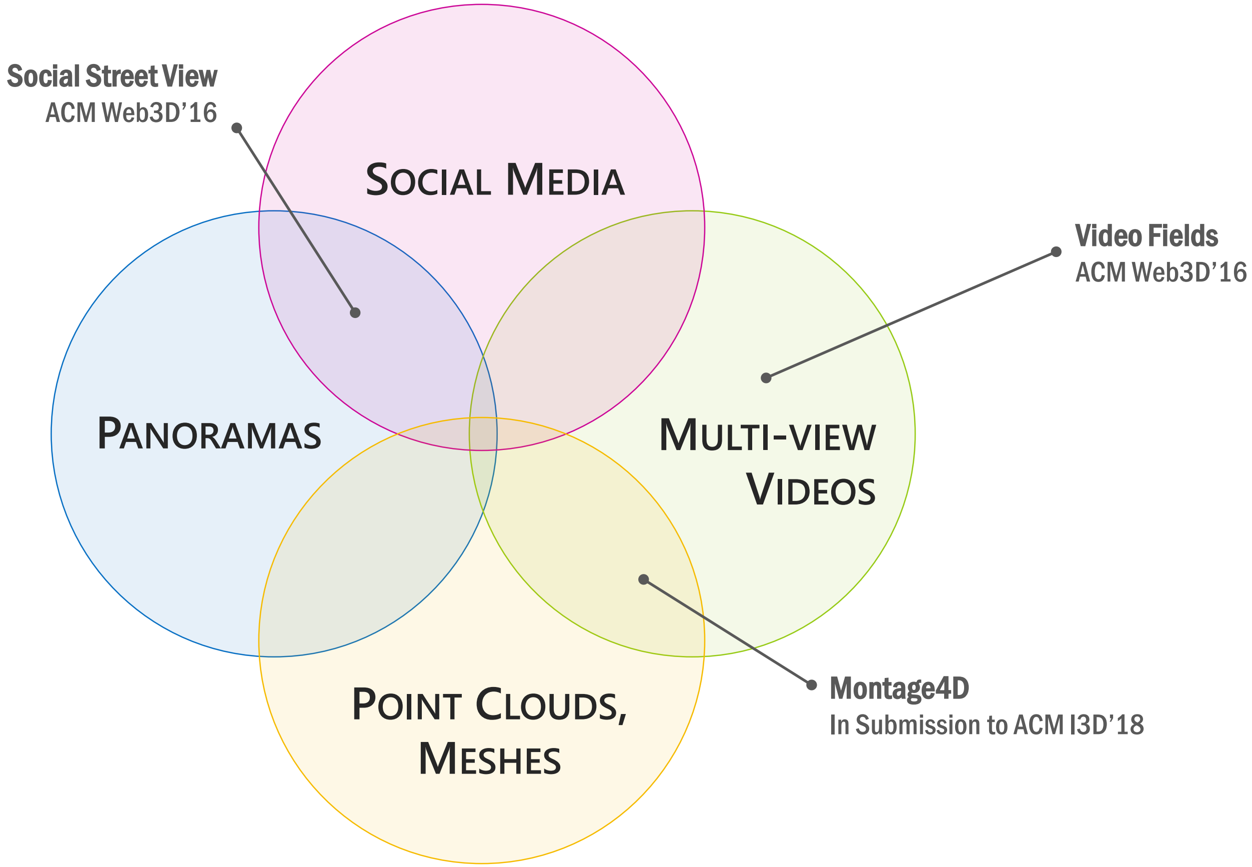
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



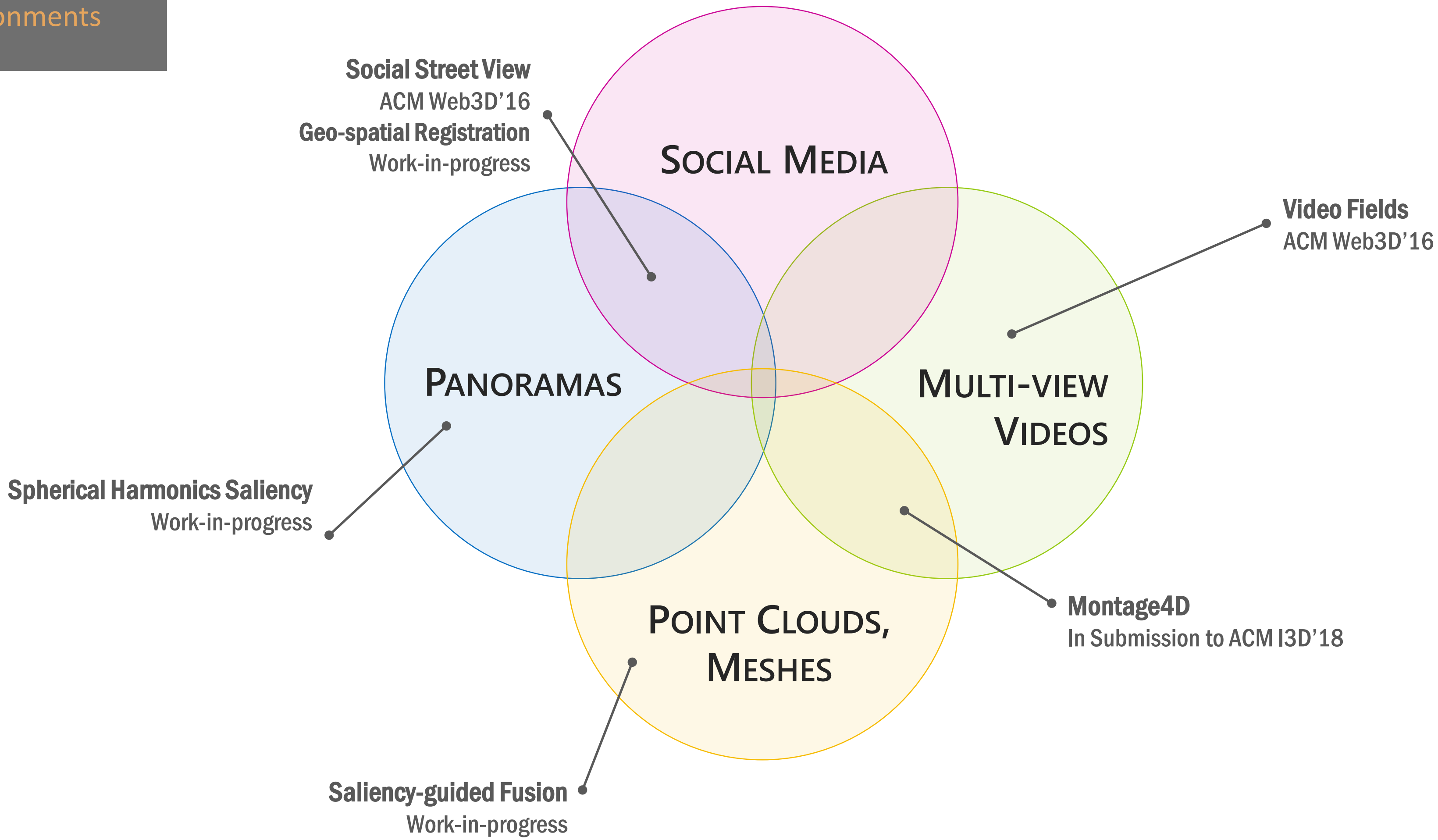
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



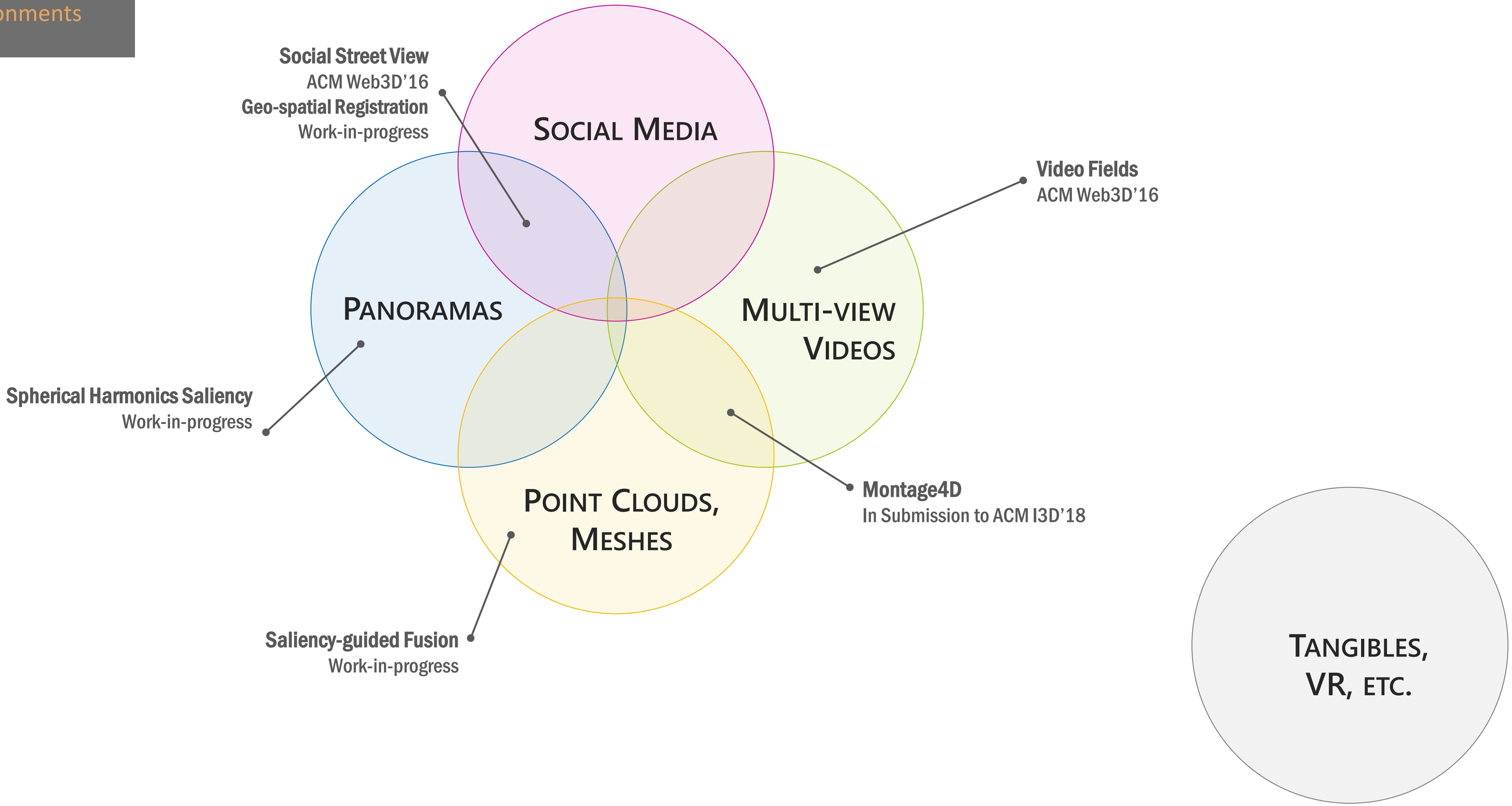
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



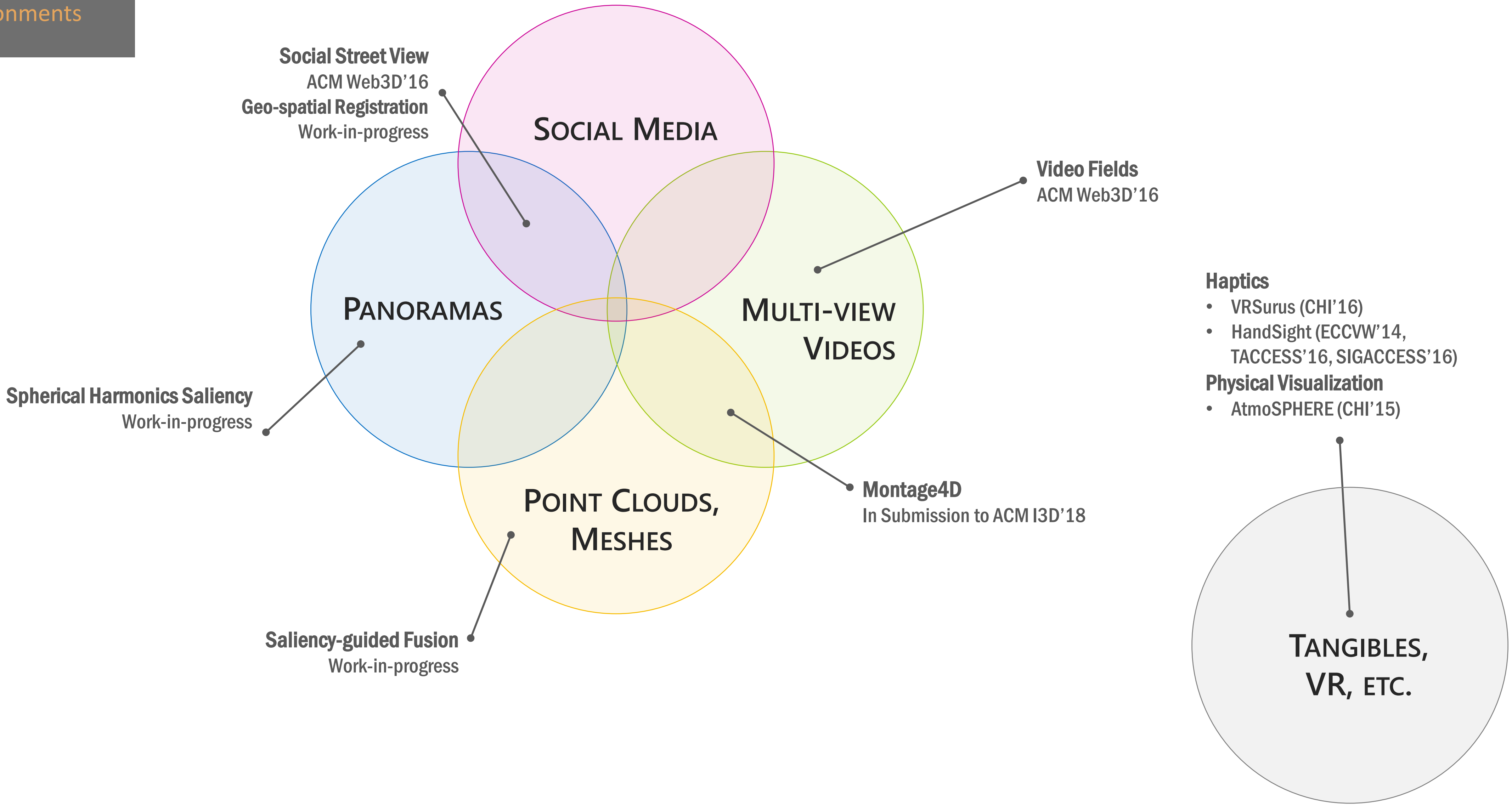
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



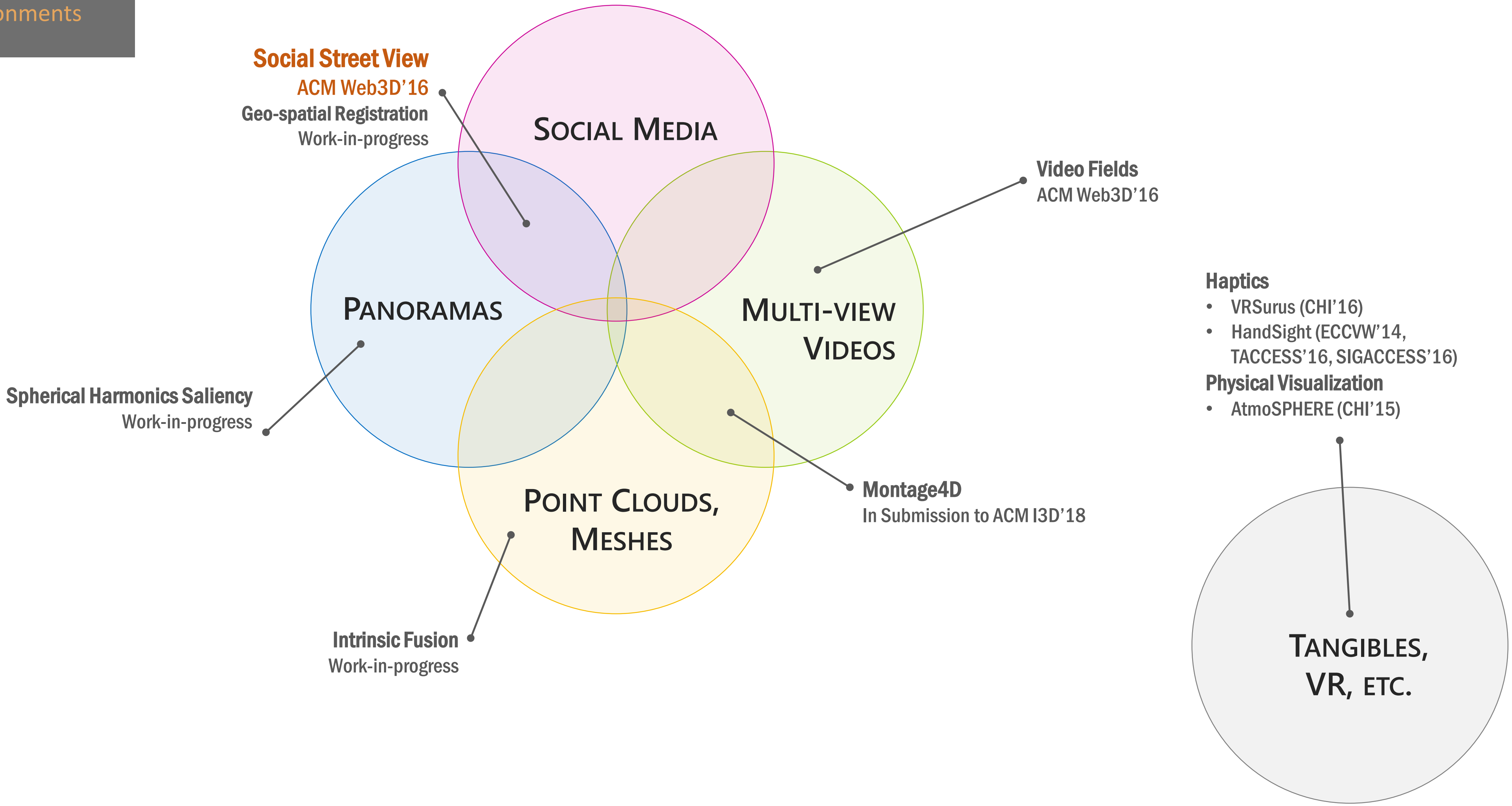
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



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

Ruofei Du and Amitabh Varshney

{ruofei, varshney} @ cs.umd.edu

www.SocialStreetView.com

In Proceedings of the International Conference of Web3D 2016. (Best Paper Award)

THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LABORATORY
AT THE UNIVERSITY OF MARYLAND

GVIL

UMIACS

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

UNIVERSITY OF MARYLAND

Introduction

Social Media

image courtesy: plannedparenthood.org

But why is social media **so popular**?

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 are useful for understanding **spatial relevance, relationship amongst users, sentiment mining**, and **propagation of influence**.

EXIF

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

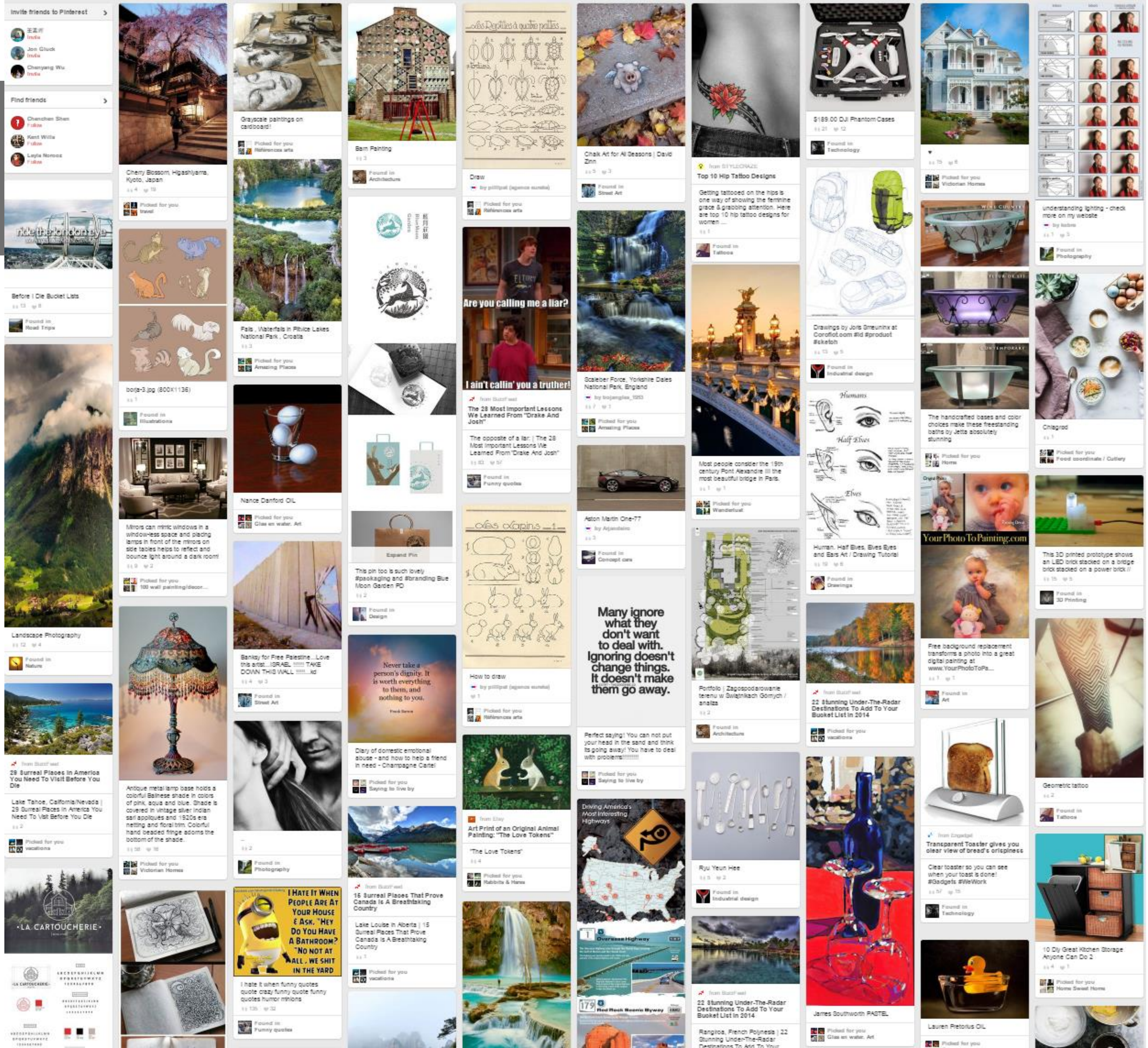


image courtesy: pinterest.com

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

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 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 disord 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.

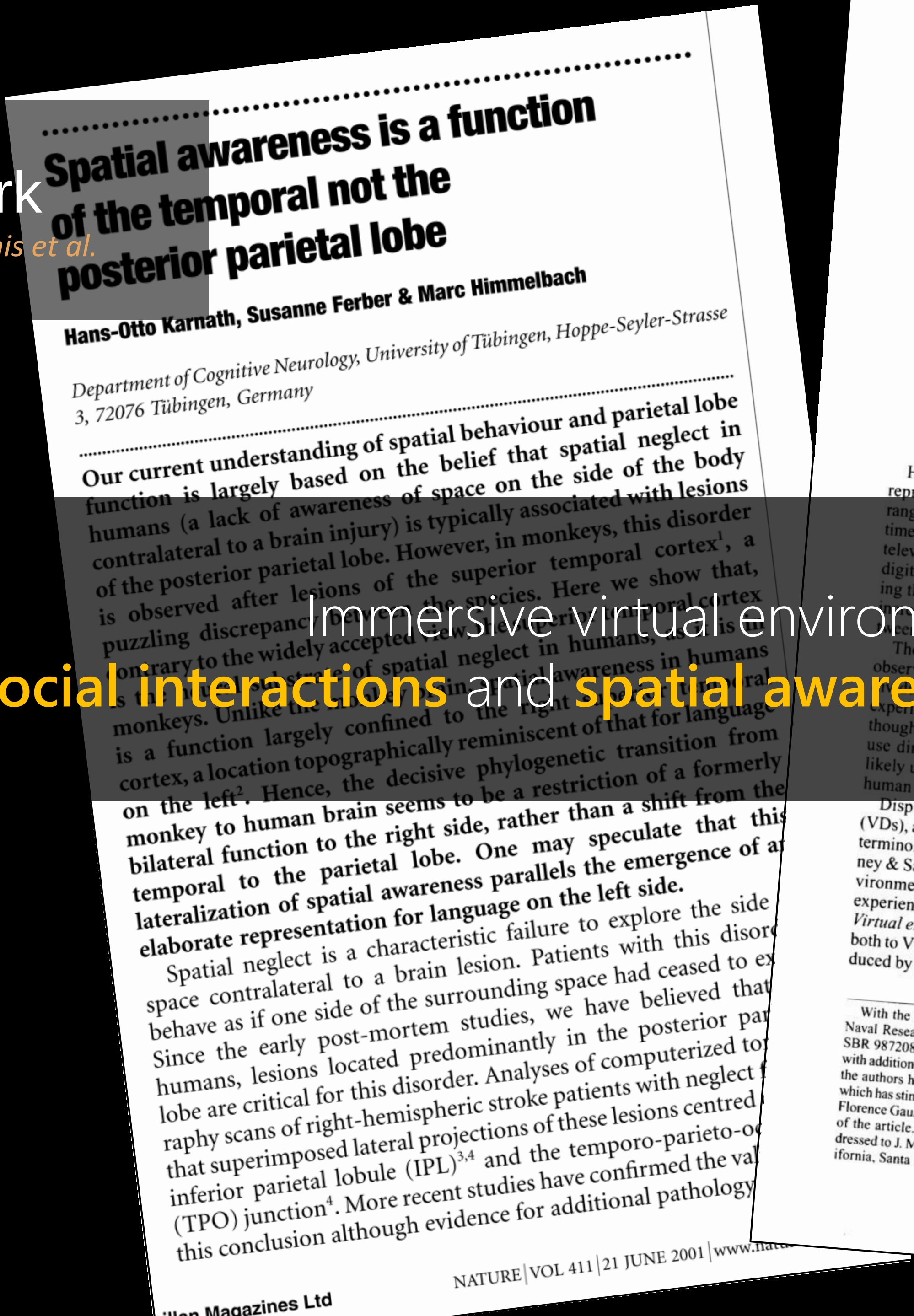
(*Virtual reality* is widely used as an alternative term, but we prefer VE.) An immersive virtual environment (IVE) is one in which the user is perceptually surrounded by the VE. Ivan Sutherland (1965), one of the originators of 3-D computer graphics, was the first person to conceive and build an immersive 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, &

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).

Related Work

Karnath et al. and *Loomis et al.*

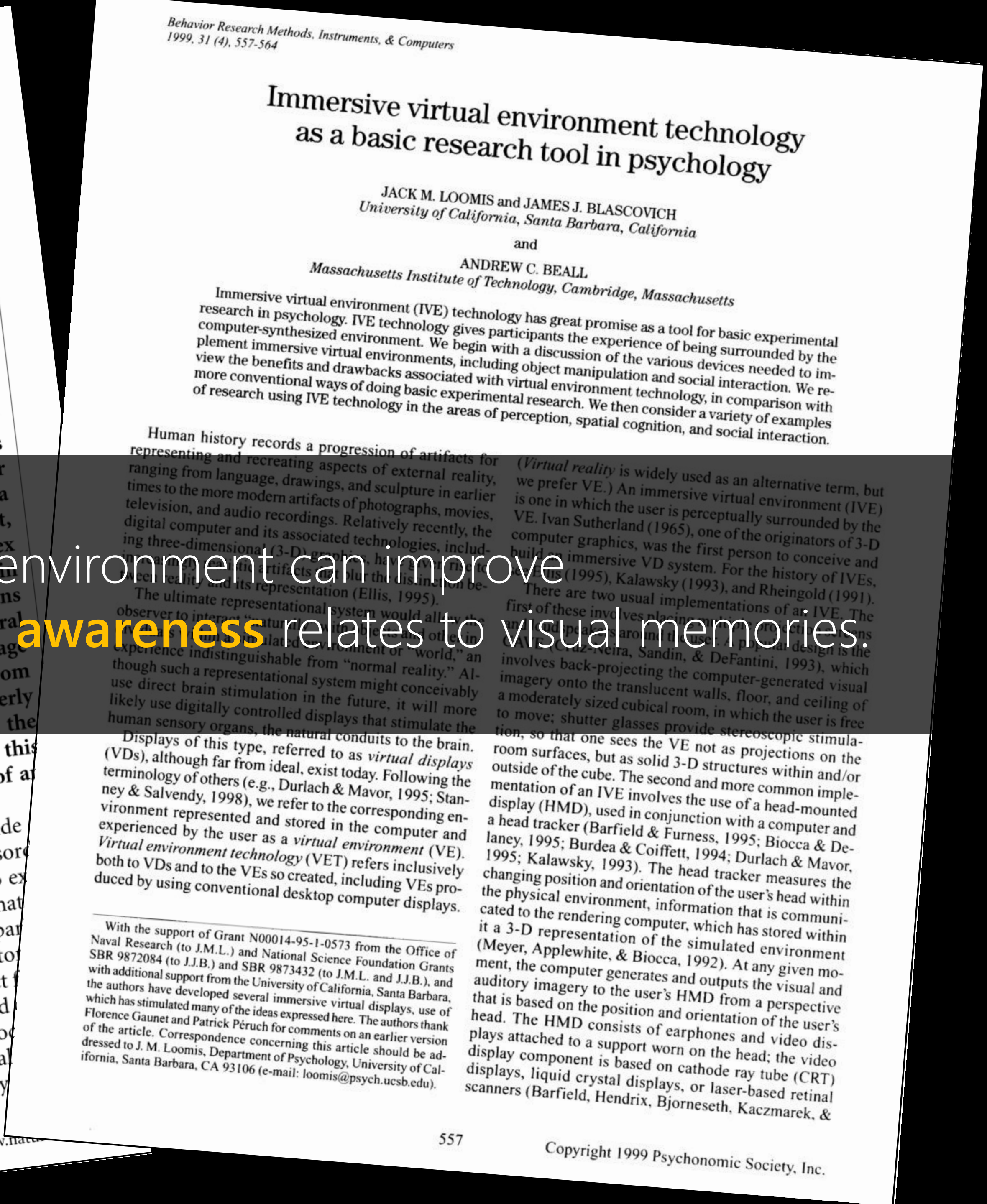


Spatial awareness is a function of the temporal not the posterior parietal lobe

Hans-Otto Karnath, Susanne Ferber & Marc Himmelbach

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

NATURE | VOL 411 | 21 JUNE 2001 |



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.

557

Copyright 1999 Psychonomic Society, Inc.

Immersive virtual environment can improve
social interactions and **spatial awareness** relates to visual memories.

Related Work

Visualization of Geo-tagged Information

Very little work has been carried out in
designing immersive interfaces that interleave
visual navigation of our surroundings
with **social media content**

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

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

*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.

Related Work

TwitterStand: News in Tweets

Sankaranarayanan et al. SIGGIS 2009

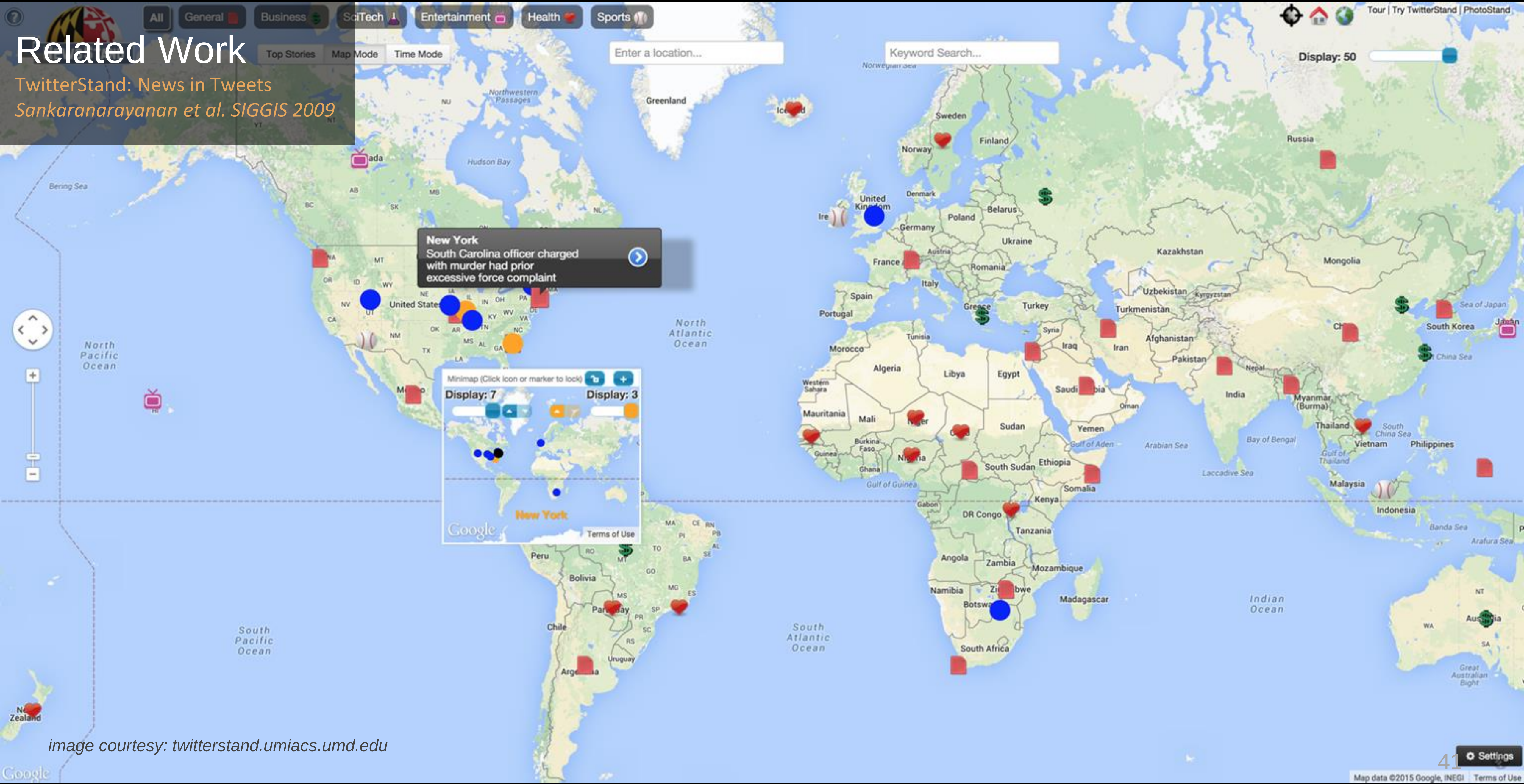
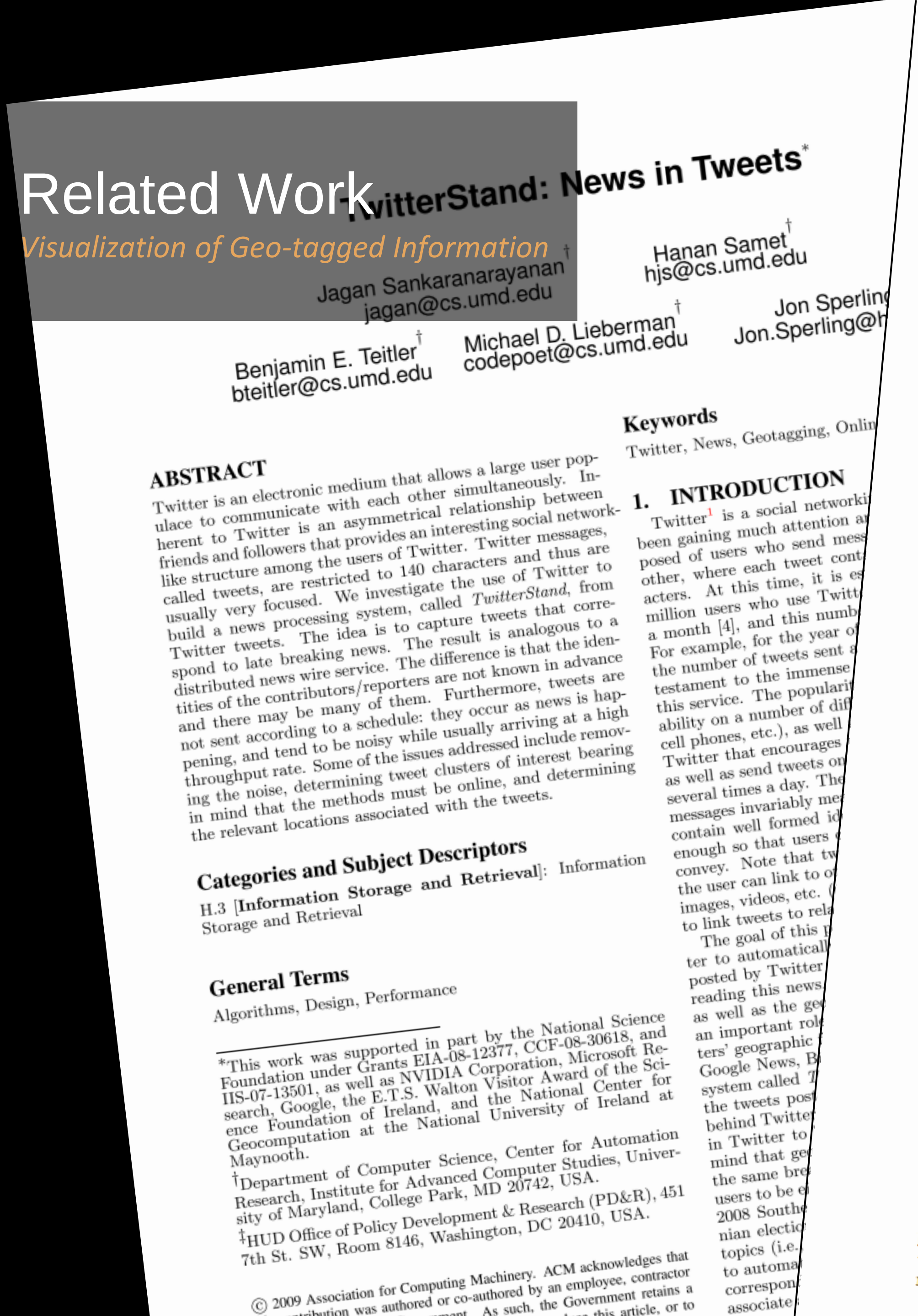


image courtesy: twitterstand.umiacs.umd.edu

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@h

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

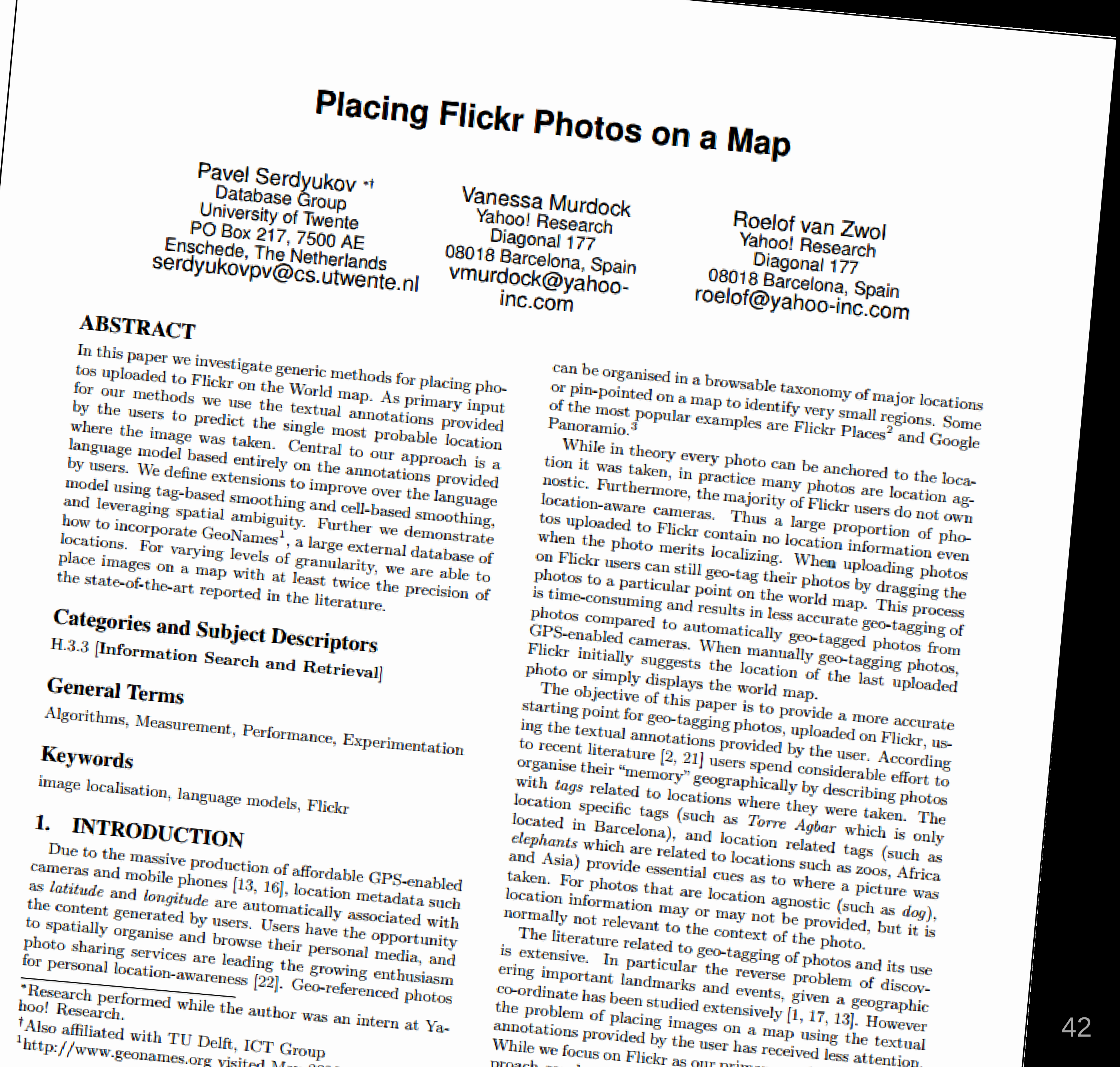
Twitter, News, Geotagging, Onlin

1. INTRODUCTION

Twitter[1] is a social networki been gaining much attention a posed of users who send mess other, where each tweet cont acters. At this time, it is es million users who use Twitt a month [4], and this numb

*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.



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. While we focus on Flickr as our primary

*Research performed while the author was an intern at Yahoo! Research.
†Also affiliated with TU Delft, ICT Group
[1]http://www.geonames.org visited May 2009

Related Work

Flickr - World Map

Serdyukov et al. SIGIR 2009

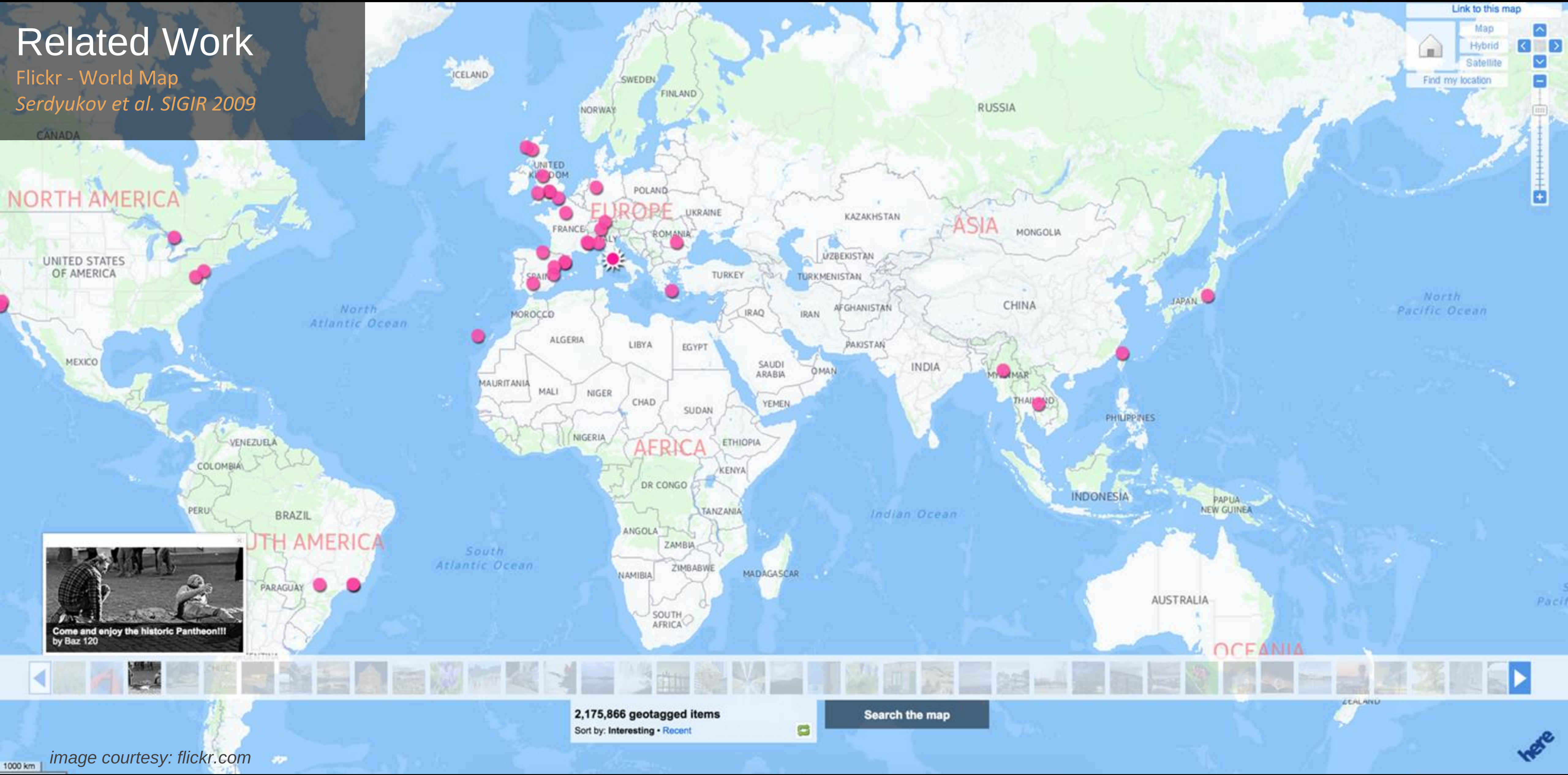
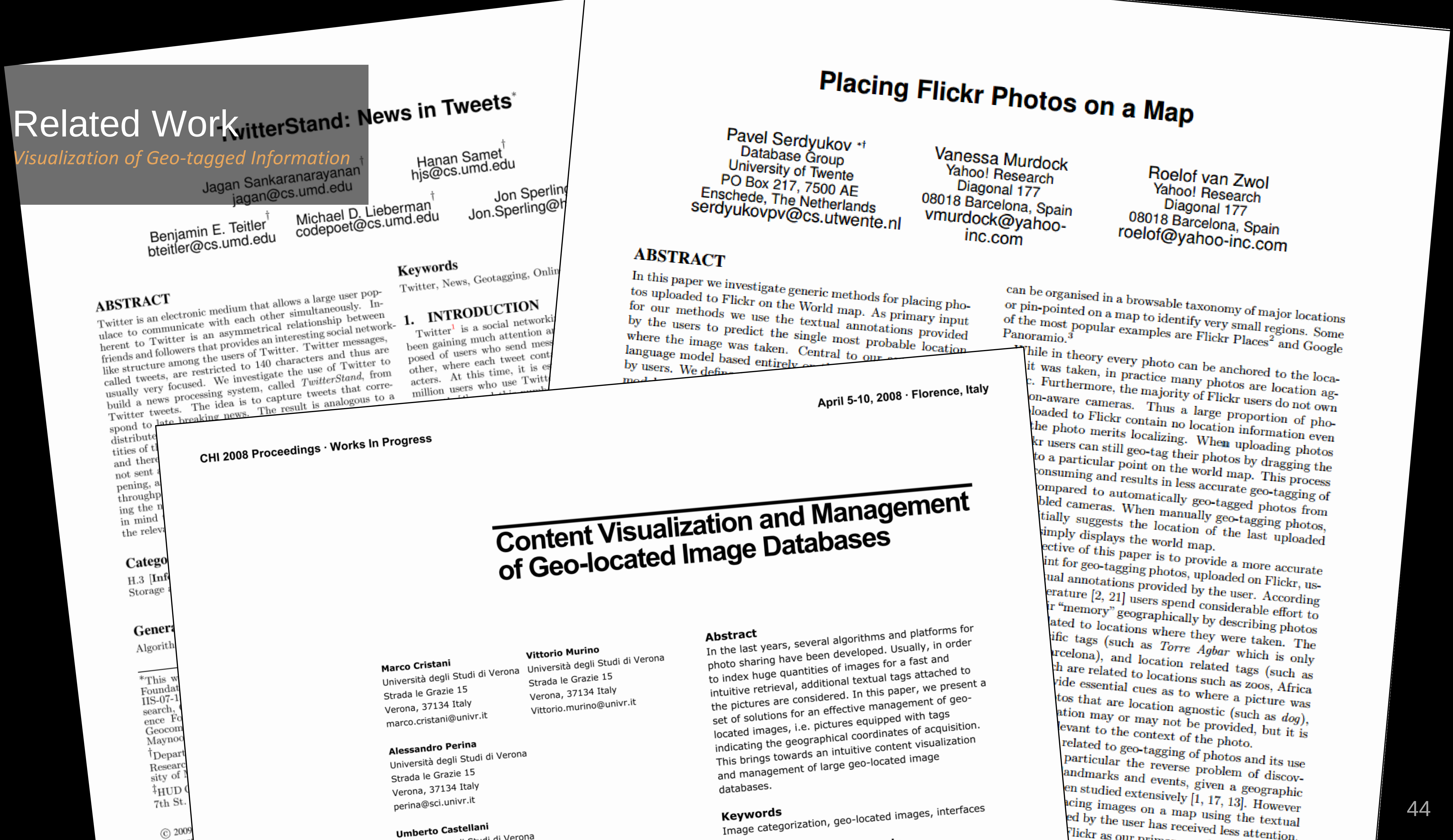


image courtesy: flickr.com

Related Work

Visualization of Geo-tagged Information



Related Work

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

image courtesy: panoramio.com

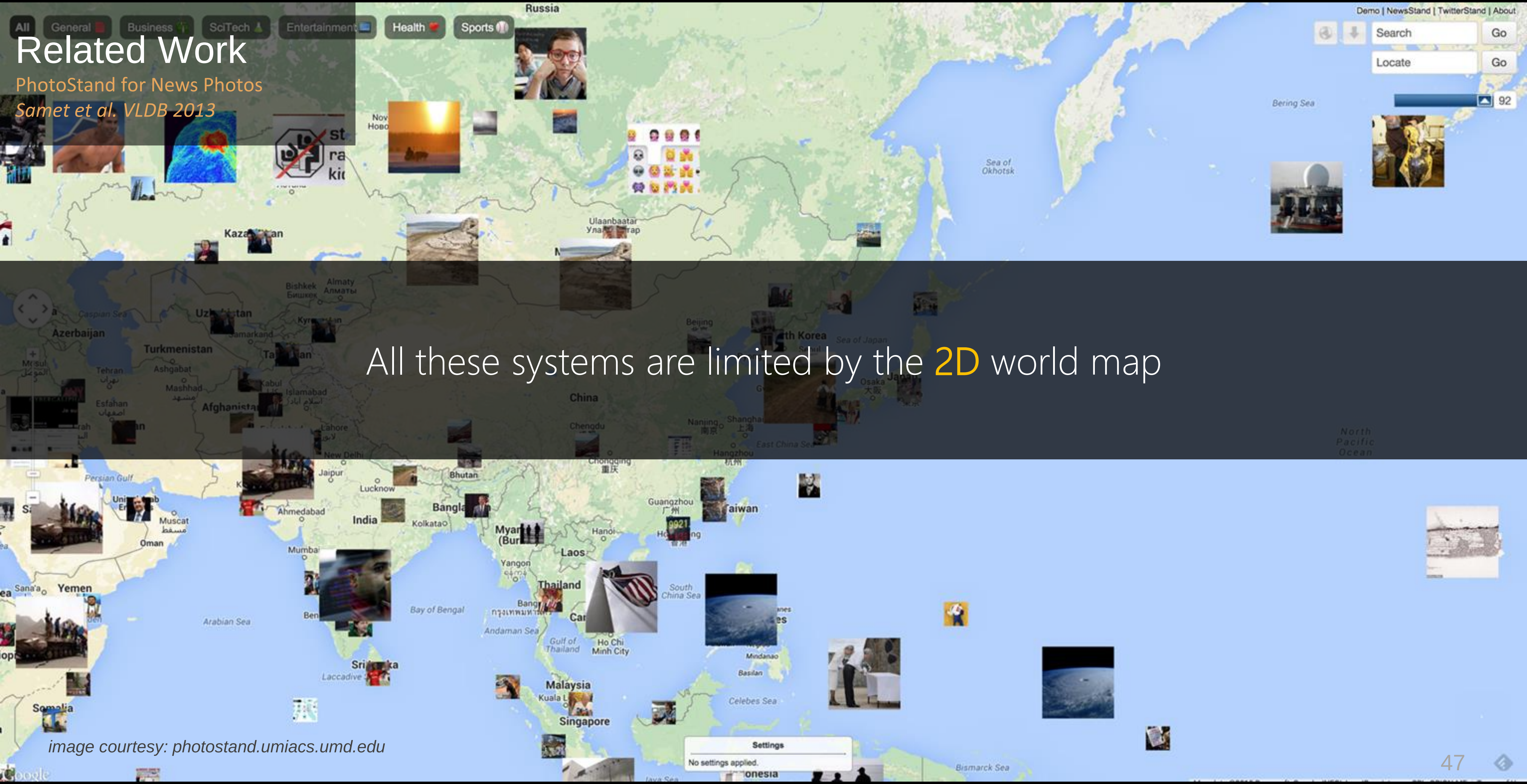
Related Work

Visualization of Geo-tagged Information

Related Work

PhotoStand for News Photos

Samet et al. VLDB 2013



All these systems are limited by the the 2D world map

image courtesy: photostand.umiacs.umd.edu

Related Work

Visualization of Geo-tagged Information (cont.)

Previous research also advances server 3D solutions.

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 objects, or the viewing specification may be explicitly ...

Registers user-annotated text and images to a particular point in 3D space.

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 properties 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

Both the 3D model and the annotation are predefined for rendering such a scene.

Related Work

Visualization of Geo-tagged Information (cont.)

UIST 2001 (ACM Symp. on User In

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 exampl properties that are simulation and which objects may be anno flexible.

We introduce algo extents to represent efficient approximat the projections of vi the unoccupied spa

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

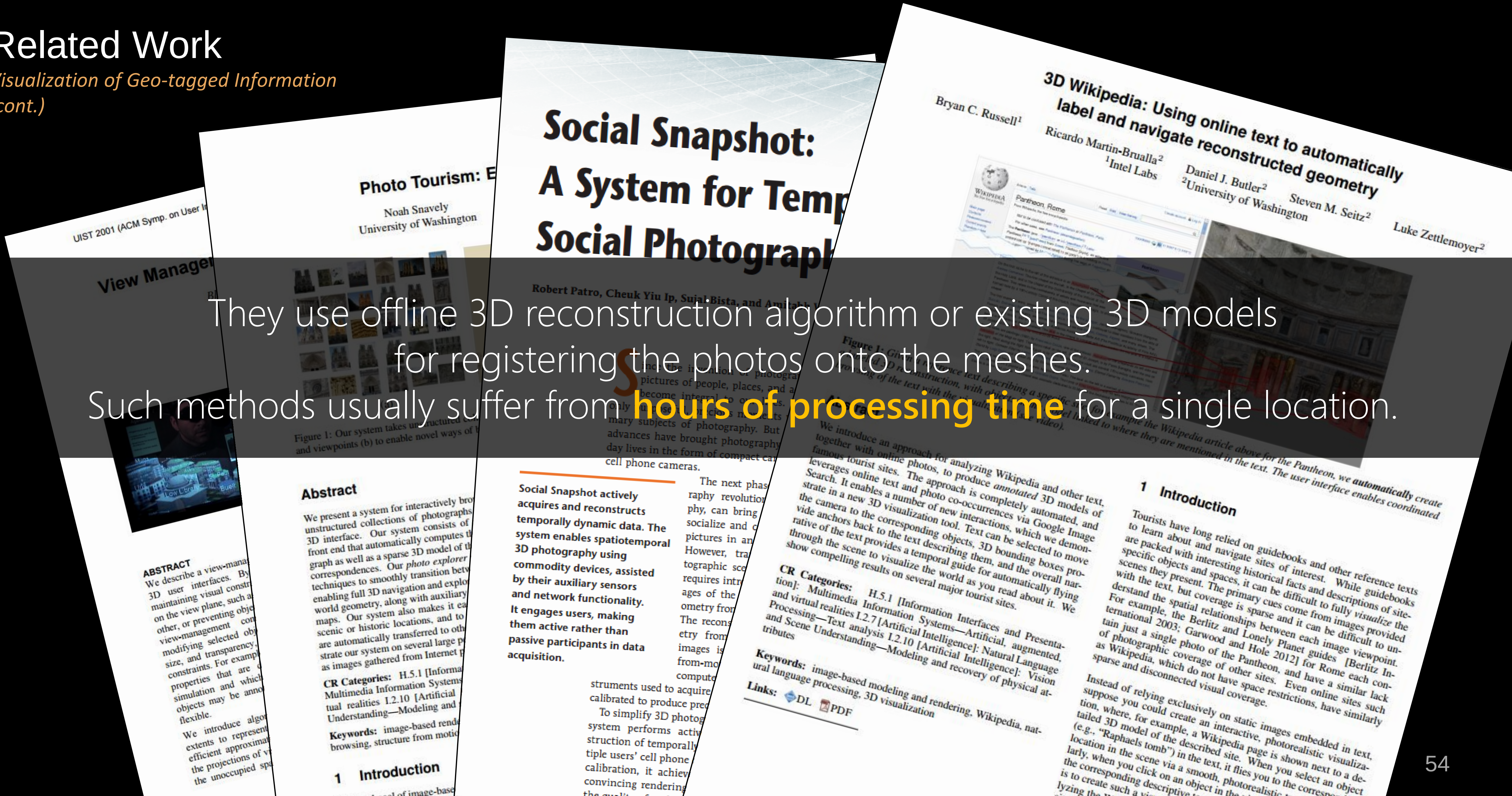
Visualization of Geo-tagged Information (cont.)

Related Work

Visualization of Geo-tagged Information (cont.)

Related Work

Visualization of Geo-tagged Information (cont.)



They use offline 3D reconstruction algorithm or existing 3D models for registering the photos onto the meshes.
Such methods usually suffer from **hours of processing time** for a single location.

So what about **our approach**?

Social Street View

Social Street View

The first 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.

Algorithm

Adding depth & normal map & maximal Poisson-disk sampling

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

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?

Architecture

Social Street View System Flowchart

Street View Cars - Cameras, LIDARs and GPS

Image courtesy from Google Street View

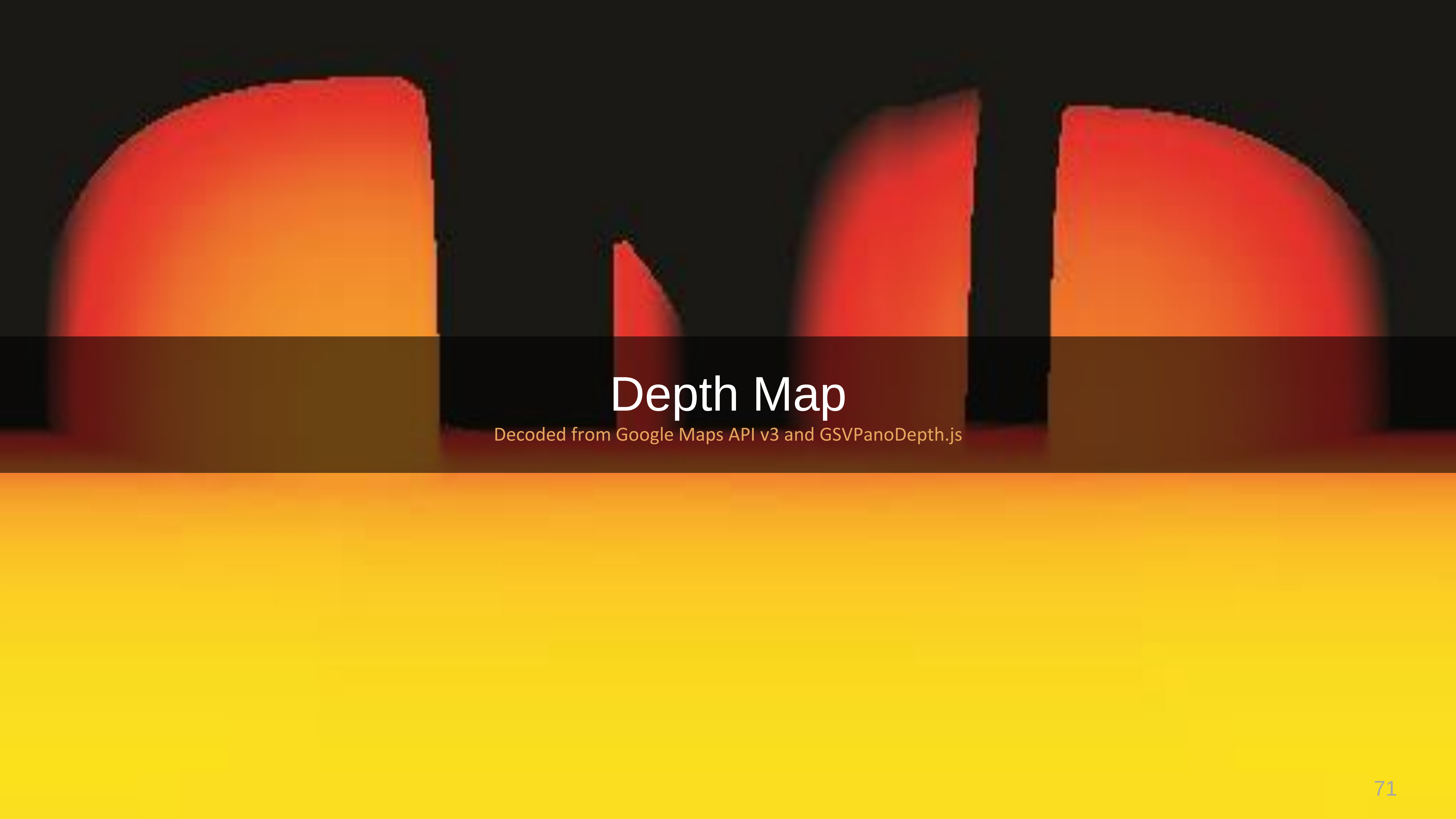
image courtesy: Google

Tiles of Panoramic Images

Image courtesy from Google Street View

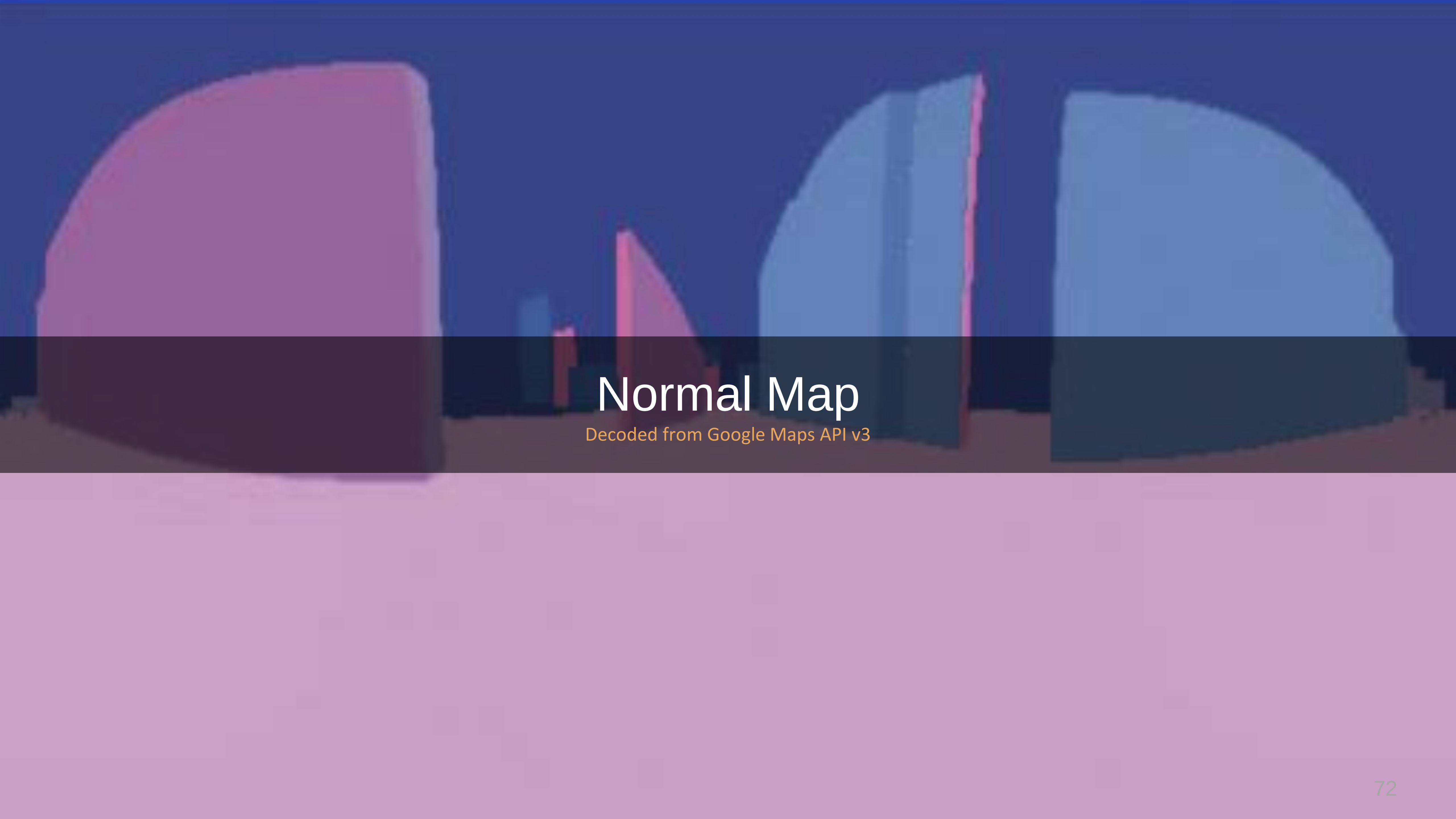
image courtesy: Google

Panorama



Depth Map

Decoded from Google Maps API v3 and GSVPanoDepth.js

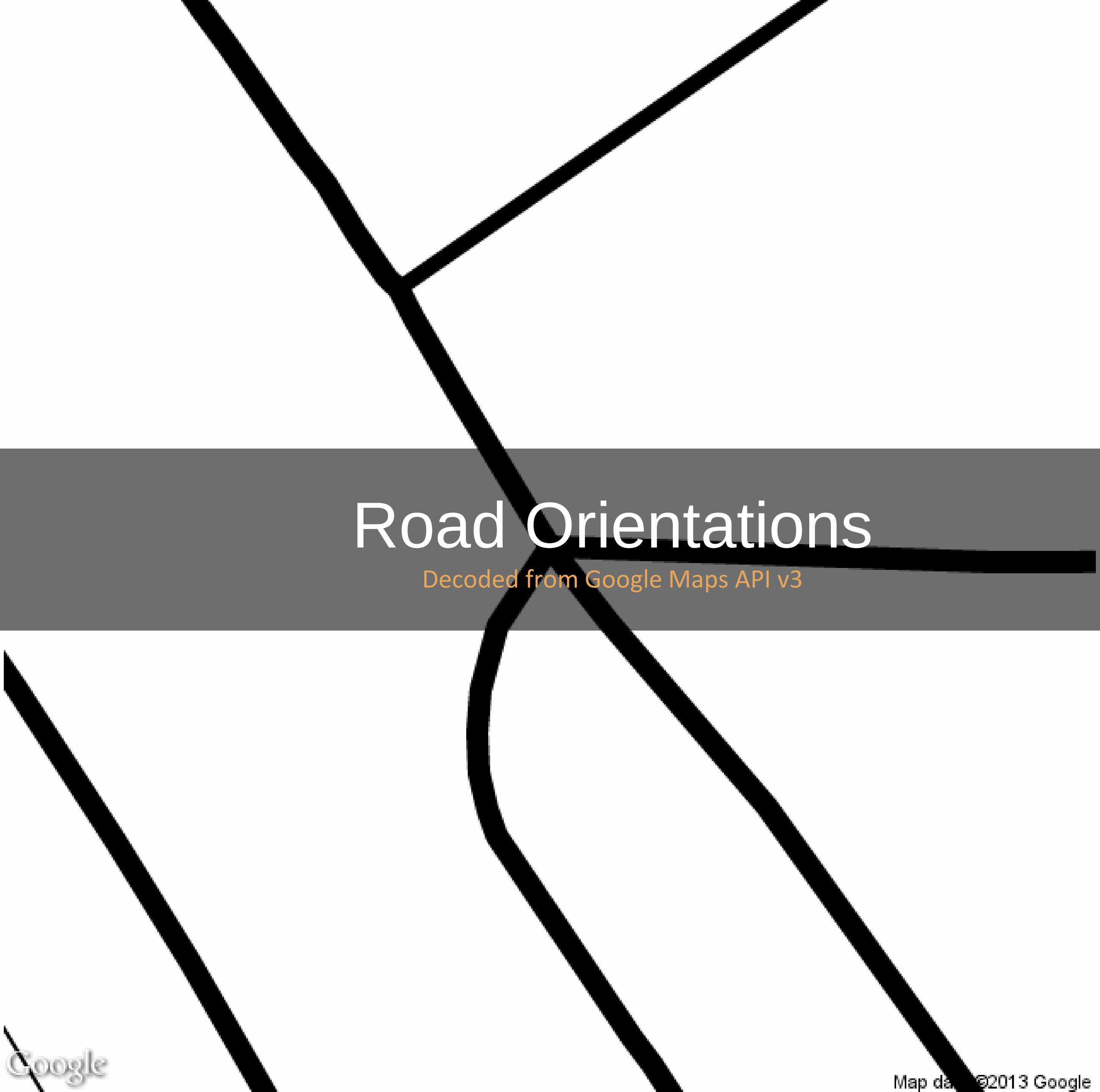


Normal Map

Decoded from Google Maps API v3

Road Orientations

Decoded from Google Maps API v3



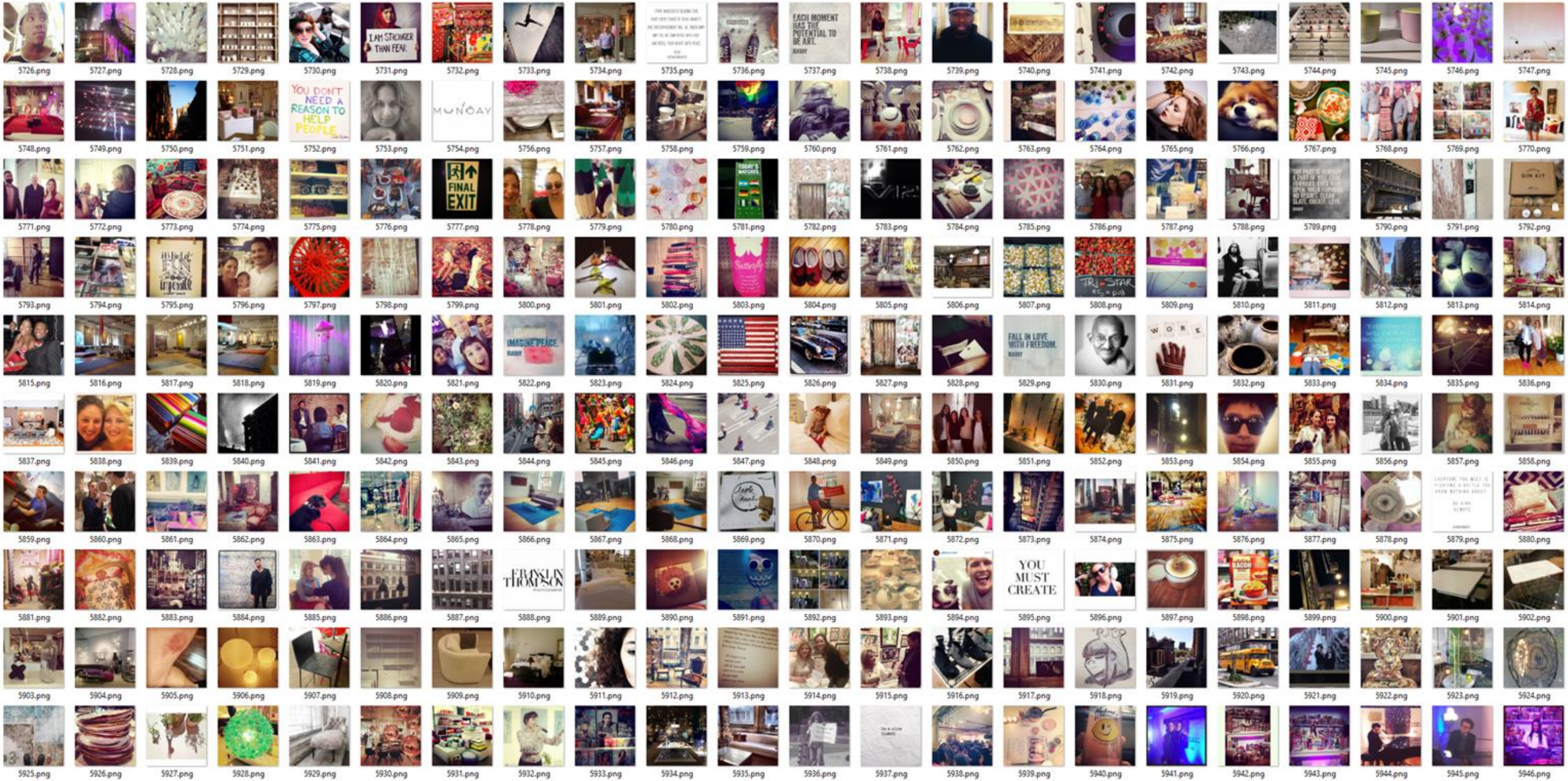
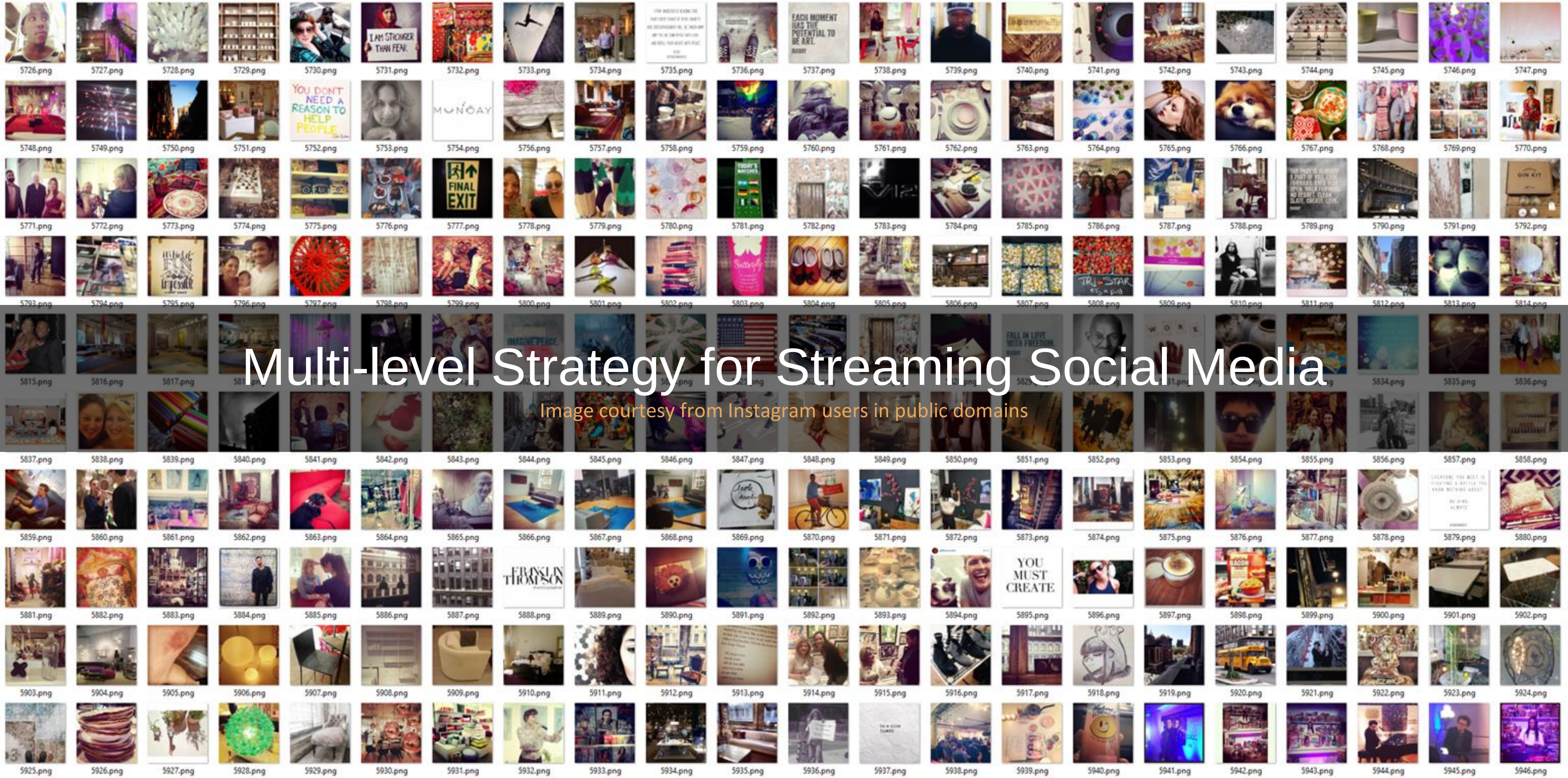


image courtesy: Instagram



Multi-level Strategy for Streaming Social Media

Image courtesy from Instagram users in public domains

image courtesy: Instagram

Data Mining
Multi-level Strategy

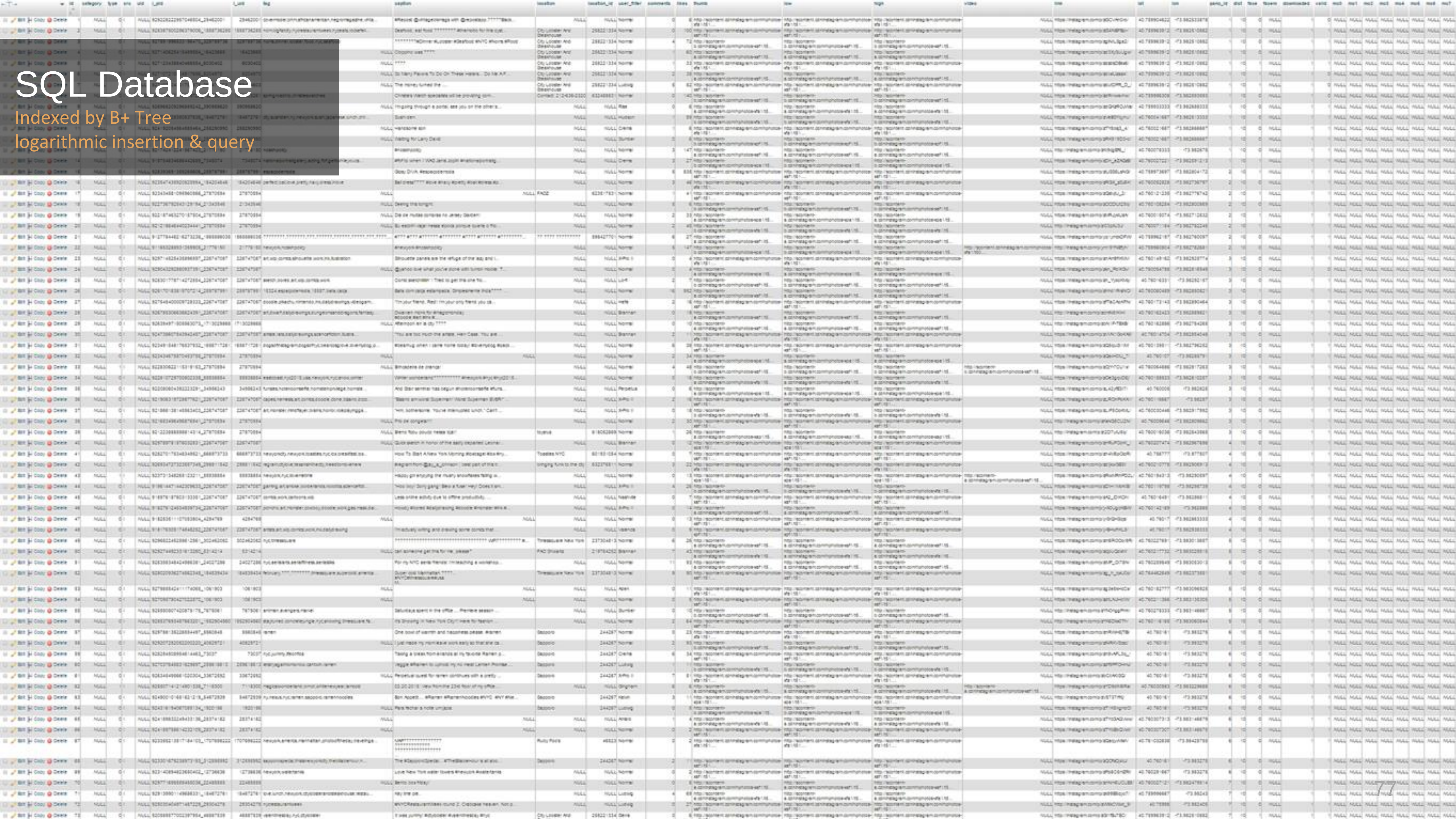
1024 METERS

… …

16 METERS

8 METERS

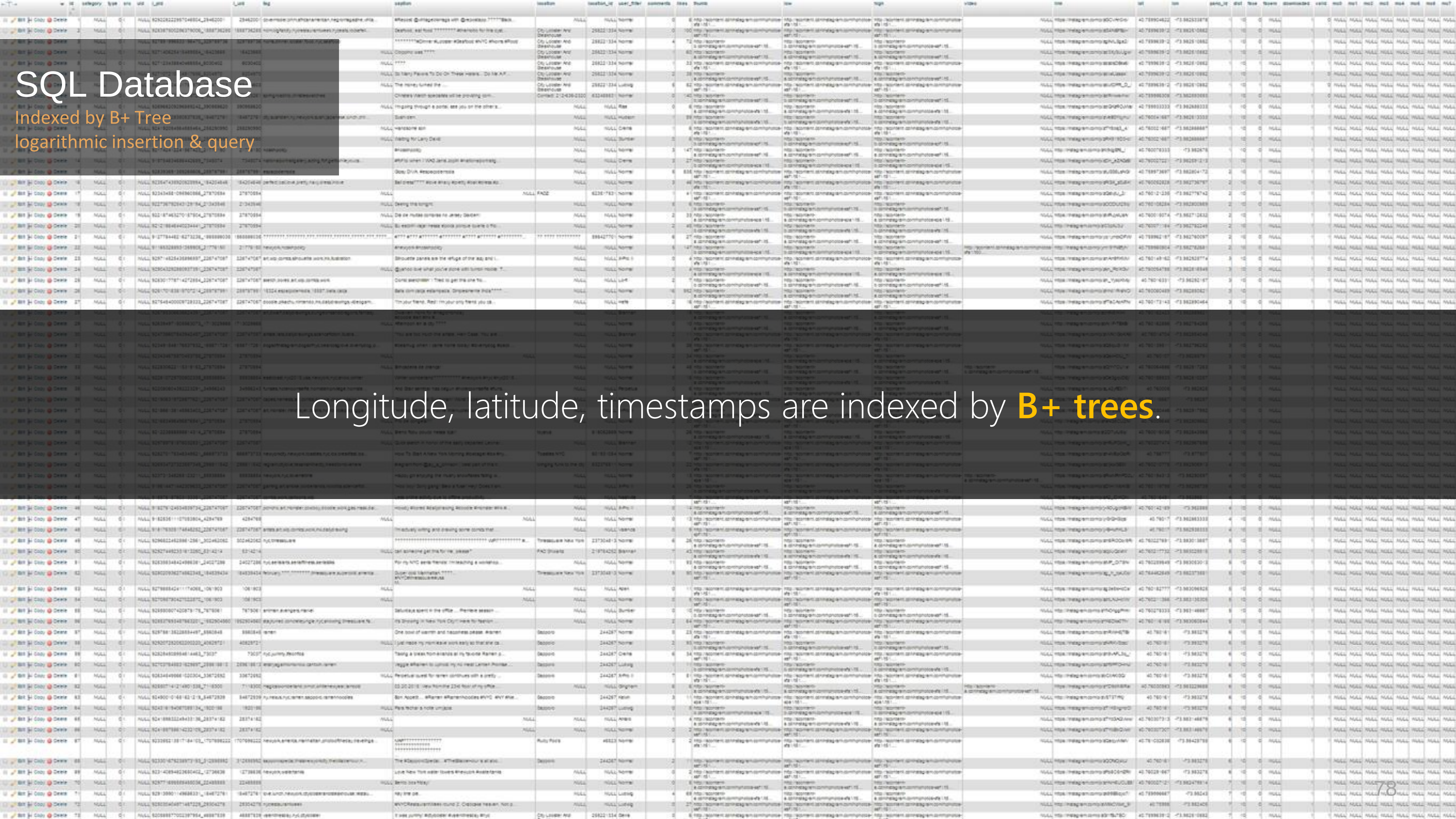
image courtesy: Instagram



SQL Database

Indexed by B+ Tree

logarithmic insertion & query



SQL Database

Indexed by B+ Tree
logarithmic insertion & query

Longitude, latitude, timestamps are indexed by **B+ trees**.

Haversine Formula

Andrew, 1805

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

The distance between social media and street view panorama

Haversine formula, which defines the distance on the surface of a sphere.

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

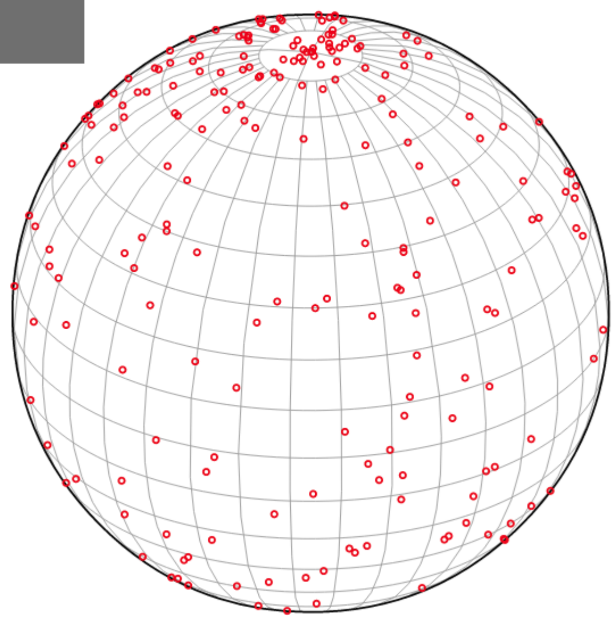
$$d_{ij} = R \cdot \beta_{ij}$$

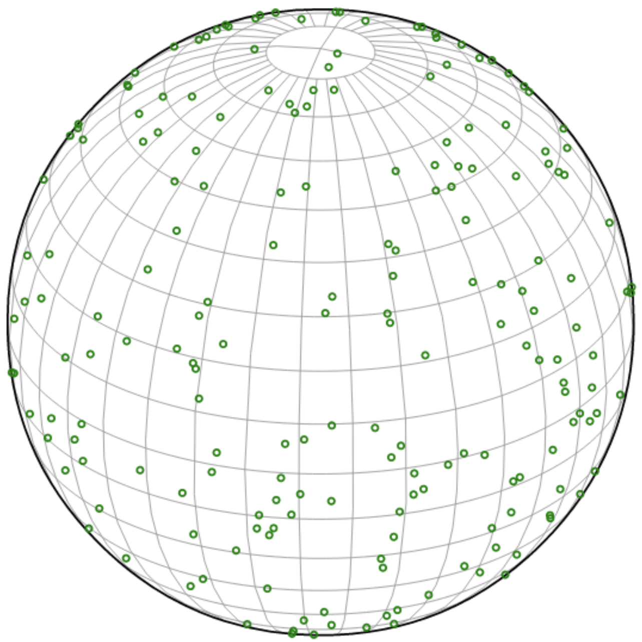
Social Street View
Augmentarium Depth
Augmentarium Baseline
Oculus VR
Address, Geo-location or HashID
Go
Keywords
Search
Urban
Rural
Indoor
Terrain
Misc
NYC
London
Paris
Rome
Tokyo
Month
1 - 9
Hour
3 - 20
Distance
0 - 64
2D
Depth
Side
Front
Model
Enhance Dark
None
Enhance Bright
Spring
Summer
Autumn
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
W 50th St
M&M'S World
Radio City Apartments
Welcome to Social Street View!
Council-Shopping Center
Cort Theatre
W 48th St
Map data ©2015 Google | Terms of Use | Report a map error
Address: 129 W 49th St, New York
SUSHIDEN

How can we **render and layout** the social media?

Baseline

Random Uniform Sampling



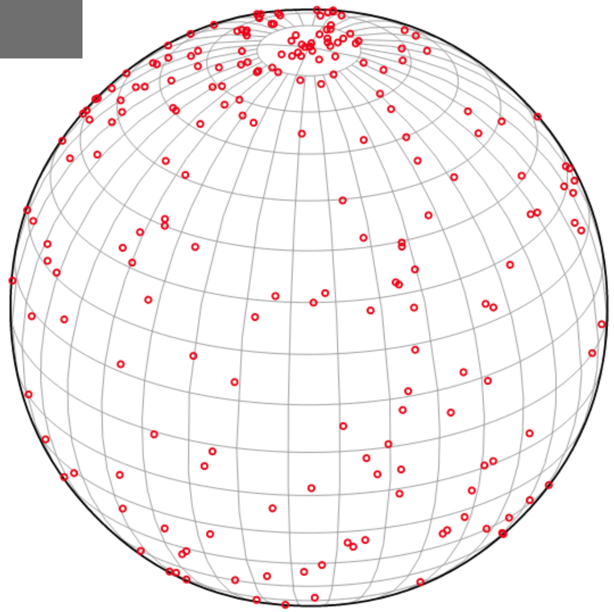


$$\varphi_i = (u_i - \frac{1}{2})\pi,\ \lambda_i = (2v_i - 1)\pi$$

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

Baseline

Random Uniform Sampling



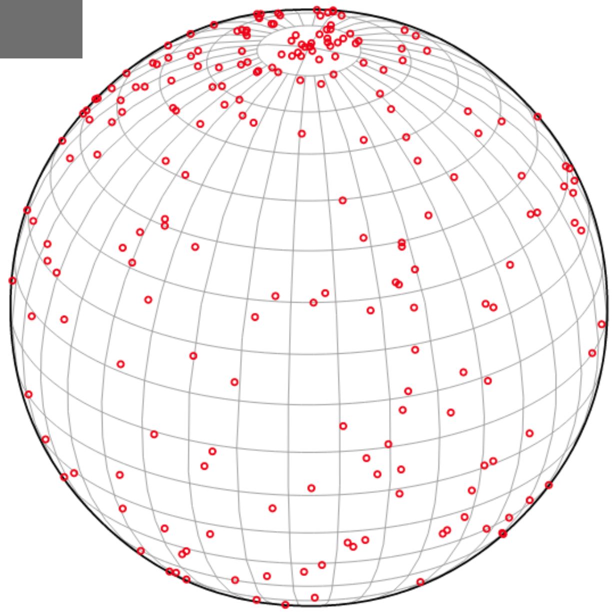
Without uniform random sampling

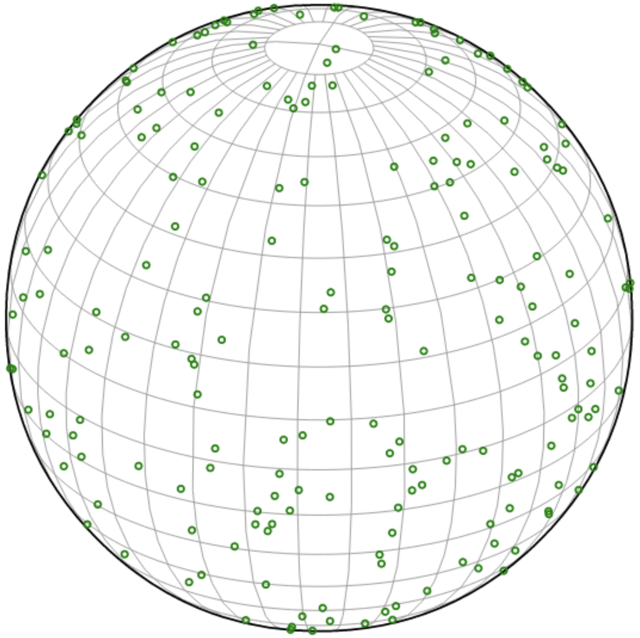
Without uniform sampling

Accumulation

Baseline

Random Uniform Sampling

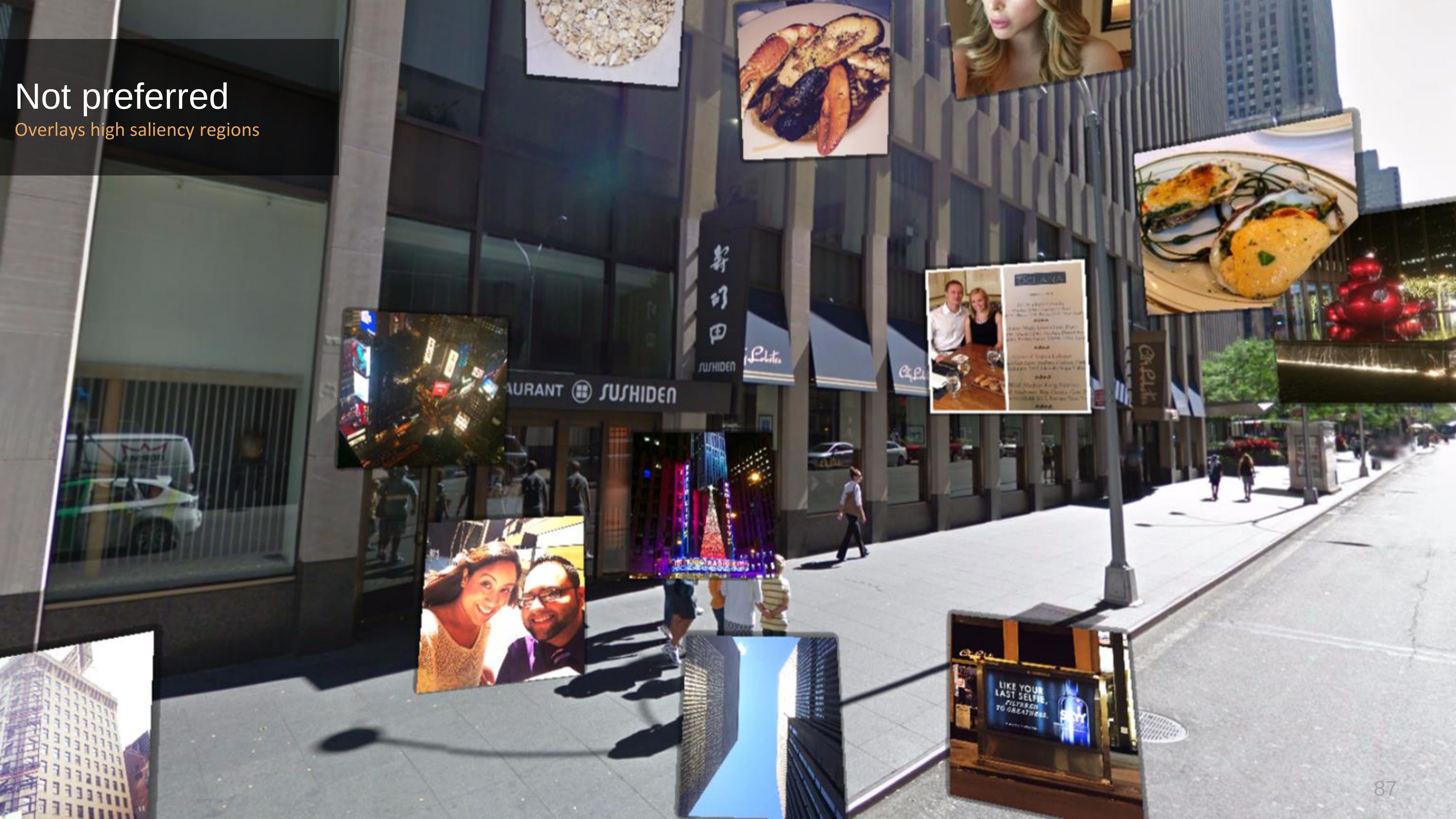




With uniform random sampling

Not preferred

Overlays high saliency regions



Add depth map

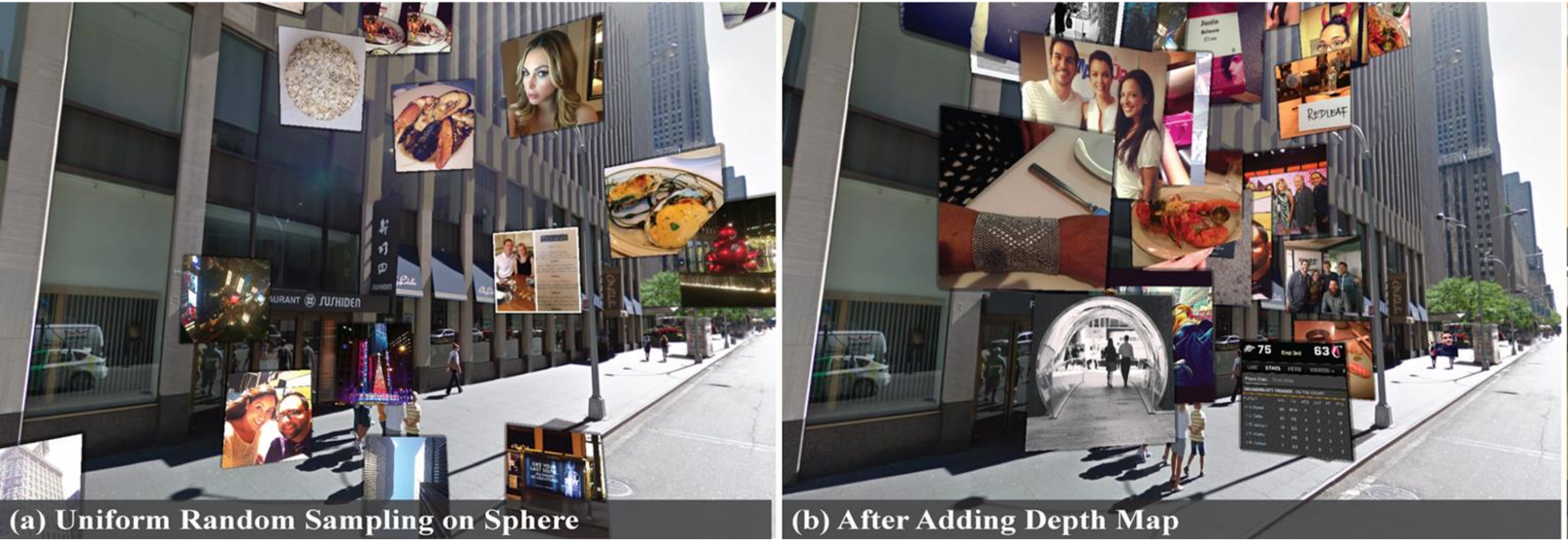
Remove sky and ground (most)

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

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

Add depth map

Remove sky and ground (most)



(a) Uniform Random Sampling on Sphere (b) After Adding Depth Map

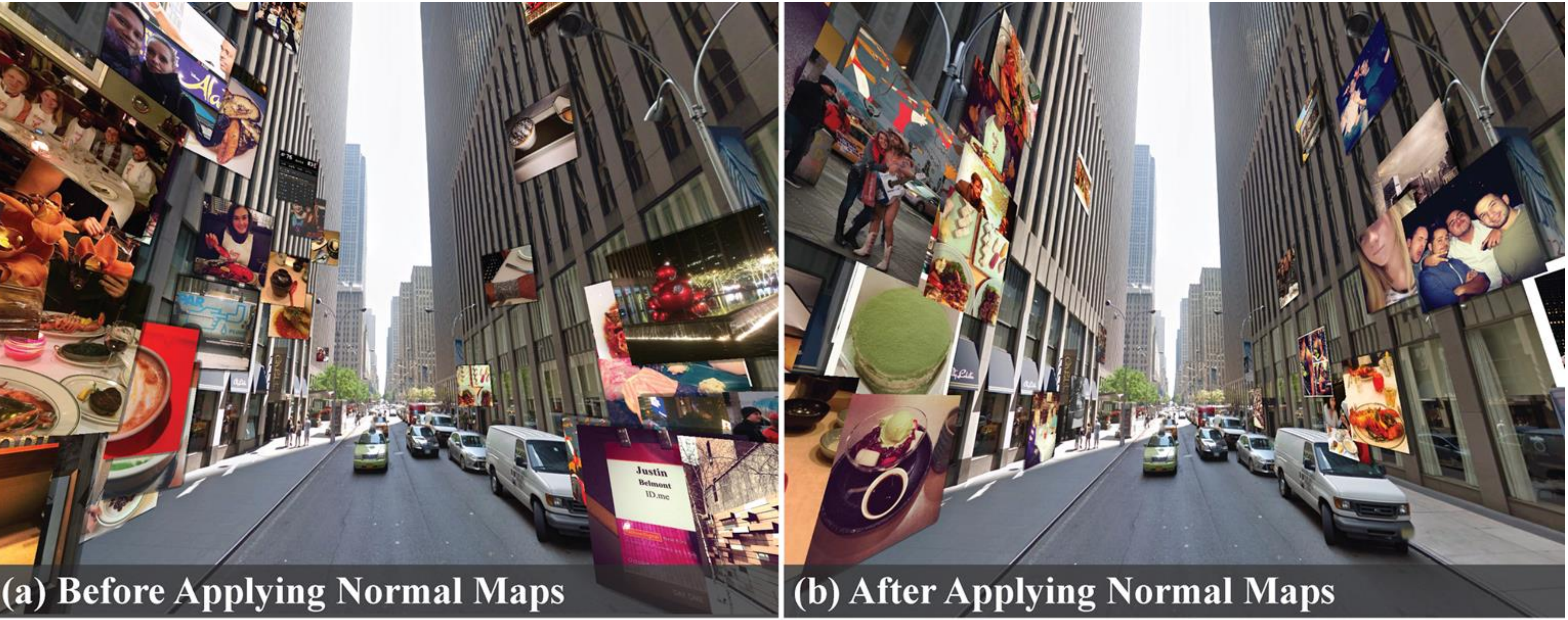
Can we ensure each image **aligns with** the building geometries?

Add normal map

Remove all ground, align images

Define the ground region:

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


Justin
Belmont
ID.me
(a) Before Applying Normal Maps
(b) After Applying Normal Maps

Can we further reduce **visual clutter** and **occlusion**?

Maximal Poisson-disk Sampling

Gamito et al. Remove visual clutter and occlusion

$$\forall \tilde{p}_i \in \tilde{P}, \tilde{P} \subseteq T, \forall S \subseteq \Omega : Pr(q_i \in S) = \int_S di \quad (10)$$

The Poisson-disk distribution is **uniform**.

$$\forall \tilde{p}_i, \tilde{p}_j \in \tilde{P}, \tilde{p}_i \neq \tilde{p}_j : \|\tilde{p}_i - \tilde{p}_j\| \geq r \quad (11)$$

Minimum distance between each pair of social media is **greater than R**.

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

Terminates the sampling when it reaches the **maximal coverage**.

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 $\tilde{P}' \subseteq \tilde{R}$ 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 $R \leftarrow \emptyset$

Pixel-Pie **by Ip. et al.** uses GPU depth-testing feature for efficient sampling.

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 $i \leftarrow 1 \ldots \min(N, |P|)$ **do**

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

  [illegible] 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 in regions which are not *Sky* nor *Ground*, and with a limitation of depth values;

Place social media as **billboards upon** the building geometries;

Cast soft shadows as if lighting were 45 degree to the normal vectors.

Sampling Comparison

Remove visual clutter and occlusion

(a) Before Maximum Poisson-disk Sampling

(b) After Maximum Poisson-disk Sampling

Algorithm

Adding depth & normal map &
maximal Poisson-disk sampling

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

This algorithm works well in **dense urban areas** such as the Manhattan District, What if there are **no buildings** in the scene?

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:
$I = I_1 \dots I_M, M \geq K \cdot |O|$.

Set $I \leftarrow \emptyset$;

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

  S[illegible]dius 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 $k \leftarrow 1 \dots K$ **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 $\mathbf{q}' \leftarrow \tilde{\mathbf{q}} - \mathbf{t}$;

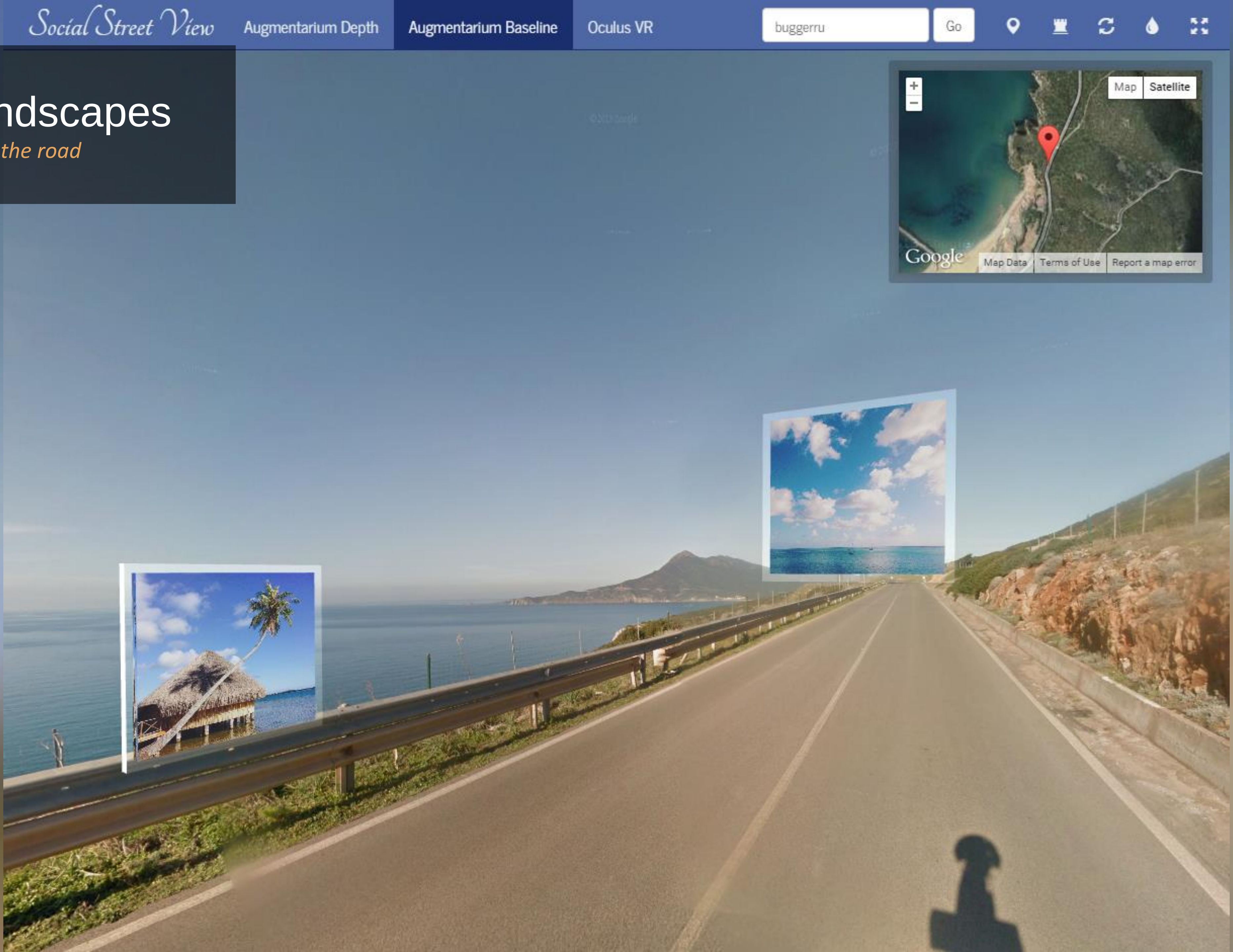
  end

end

We use only the **orientations of the road**.

Scenic Landscapes

Using orientation of the road



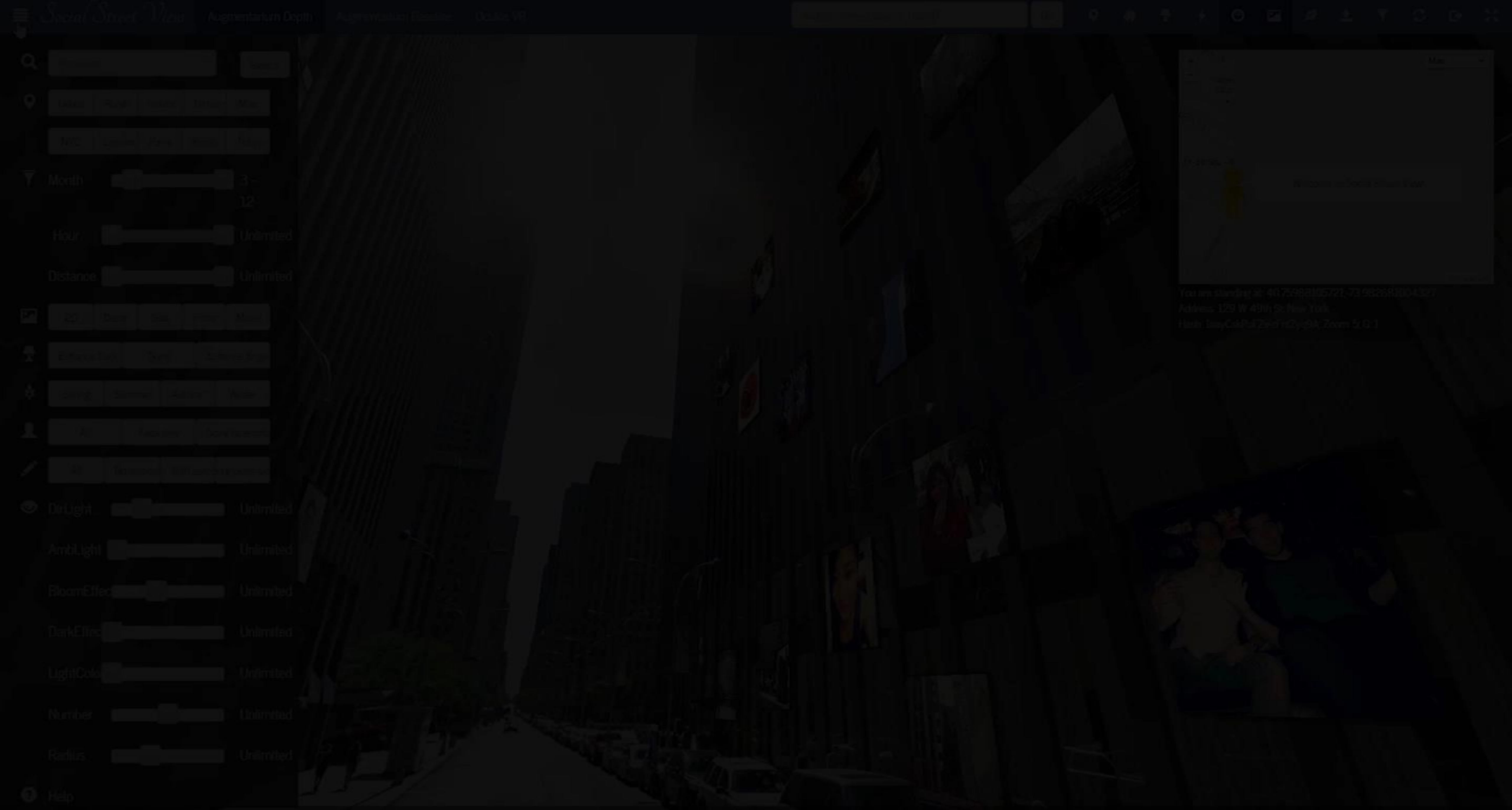
Scenic Landscapes

Using orientation of the road

What if …

Temporal information is used for filtering and rendering?

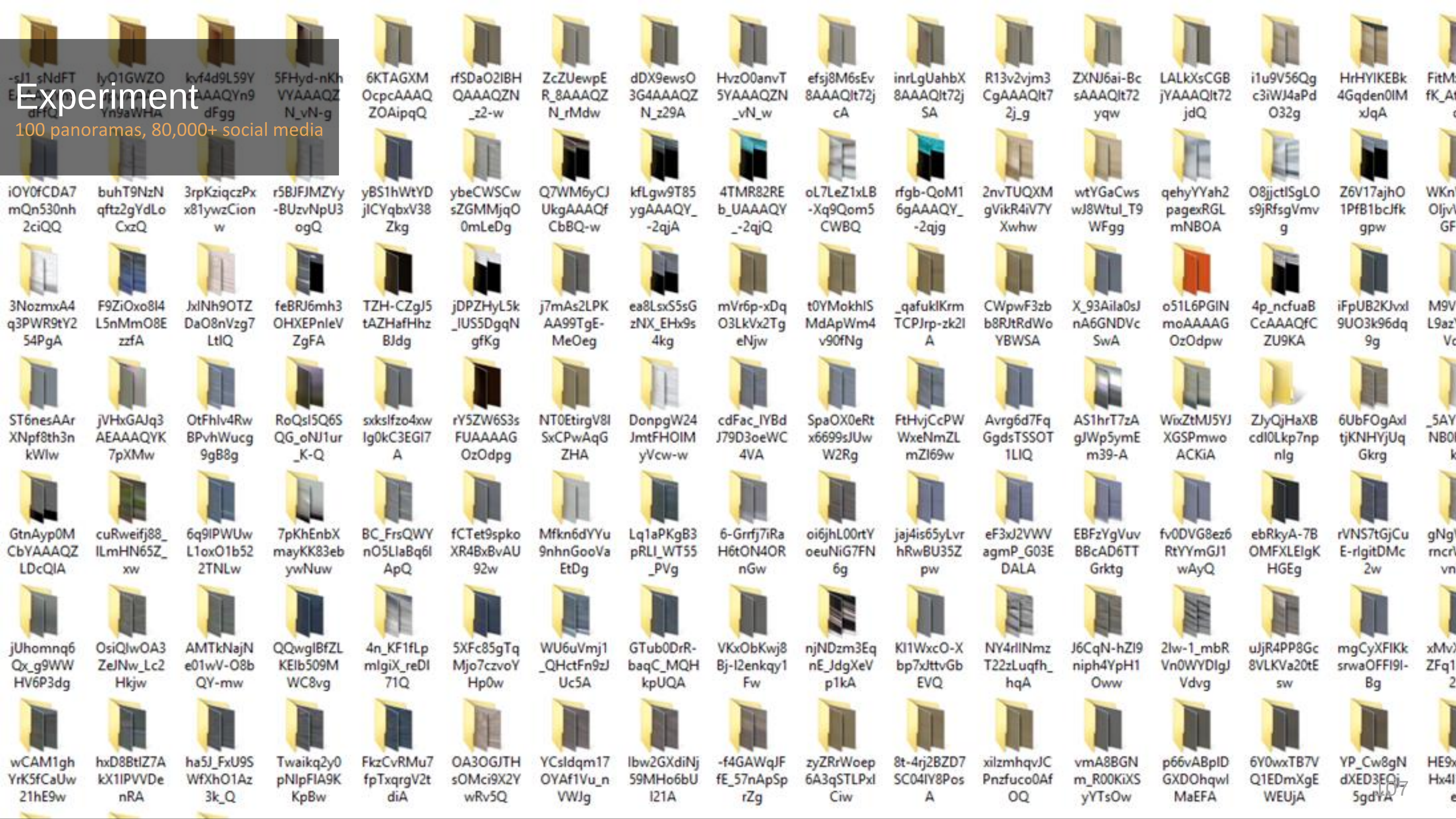
image courtesy: mindtheproduct.com



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

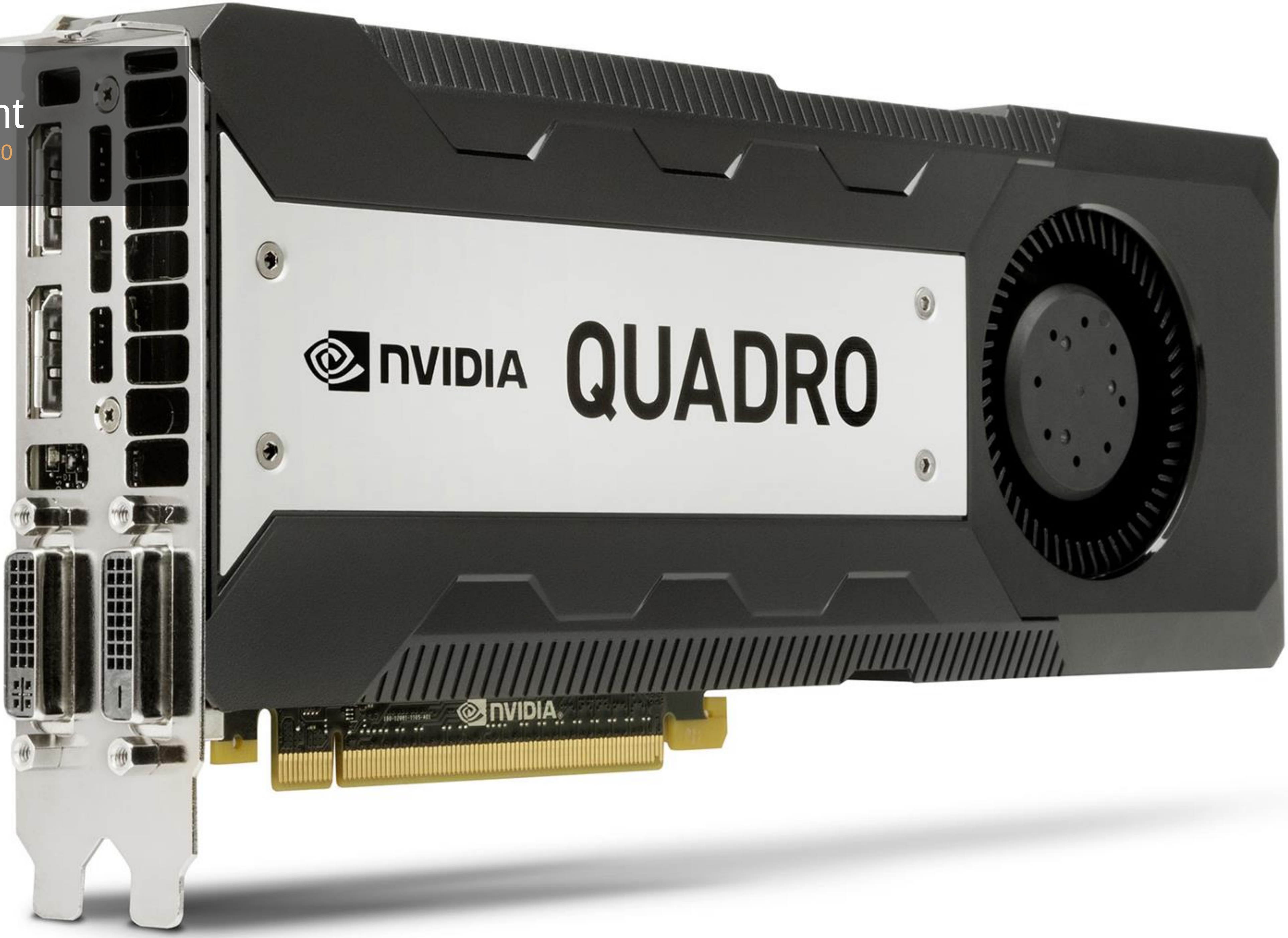
Experiment

100 panoramas, 80,000+ social media



Experiment

Nvidia Quadro K6000



Experiment

Conditions of panoramic images

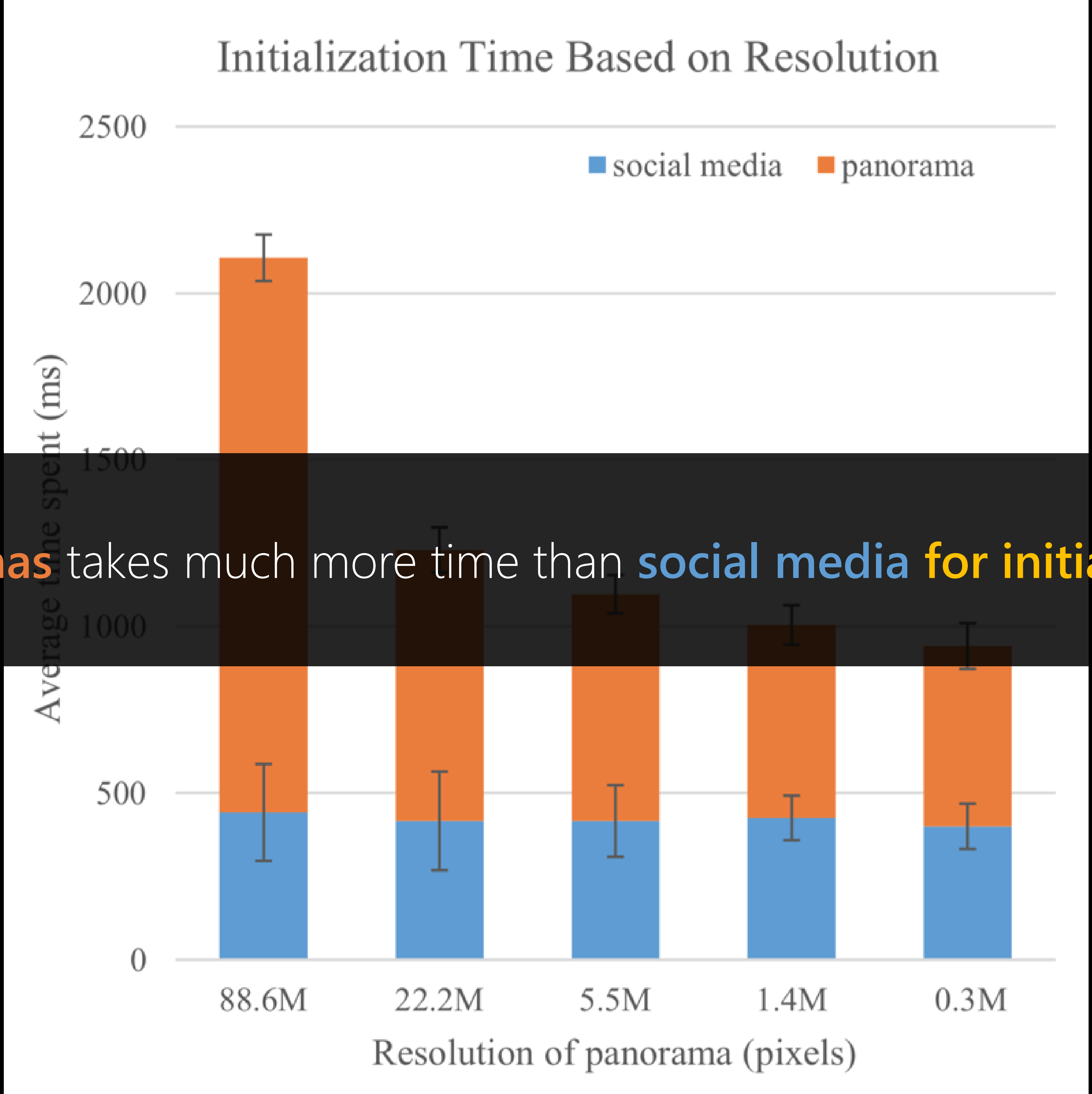
Immersive high-resolution screens

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

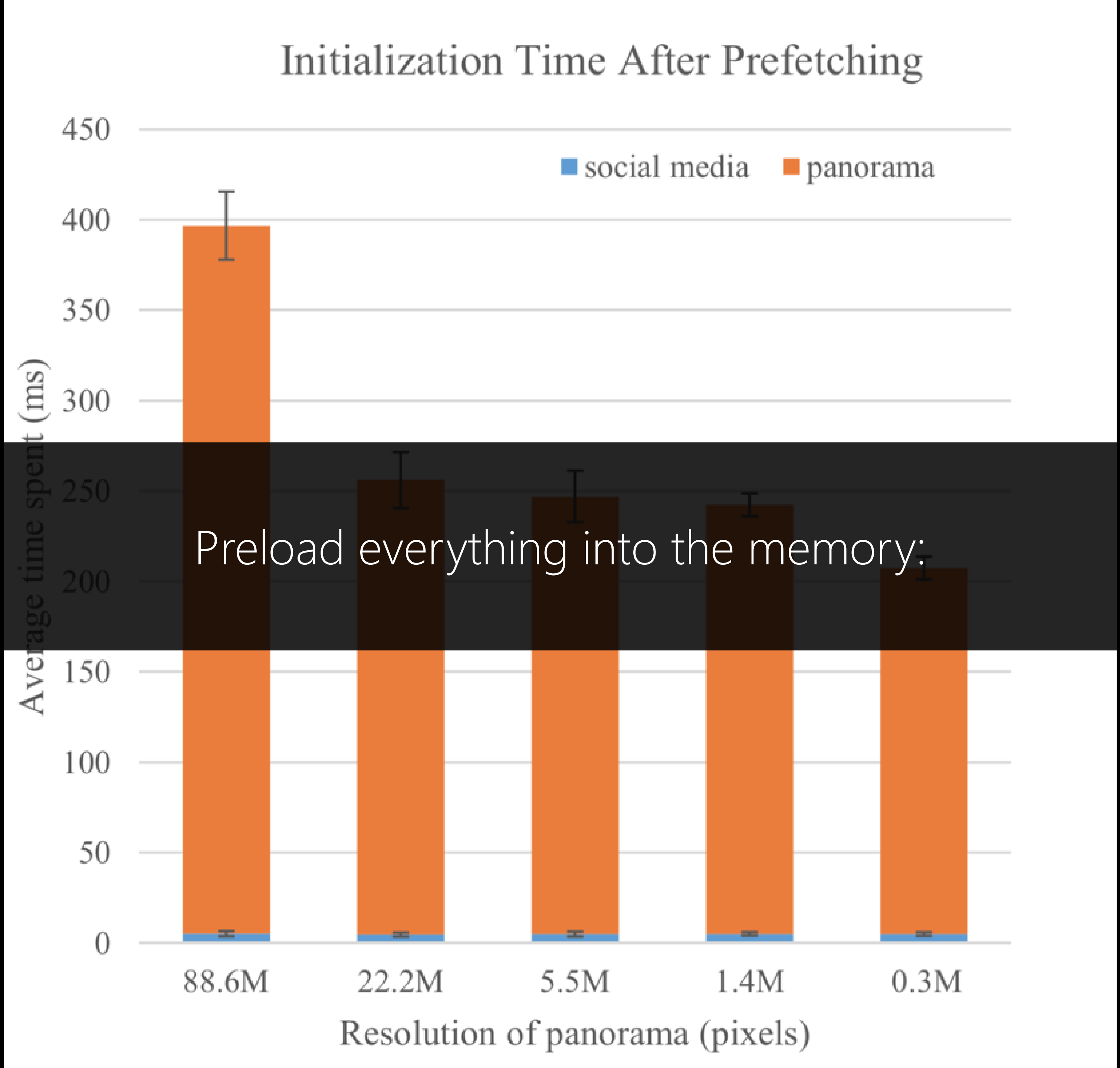
Panorama takes a while to load



Panoramas takes much more time than social media for initialization.

After Prefetching

¾ - ⅘ time reduced

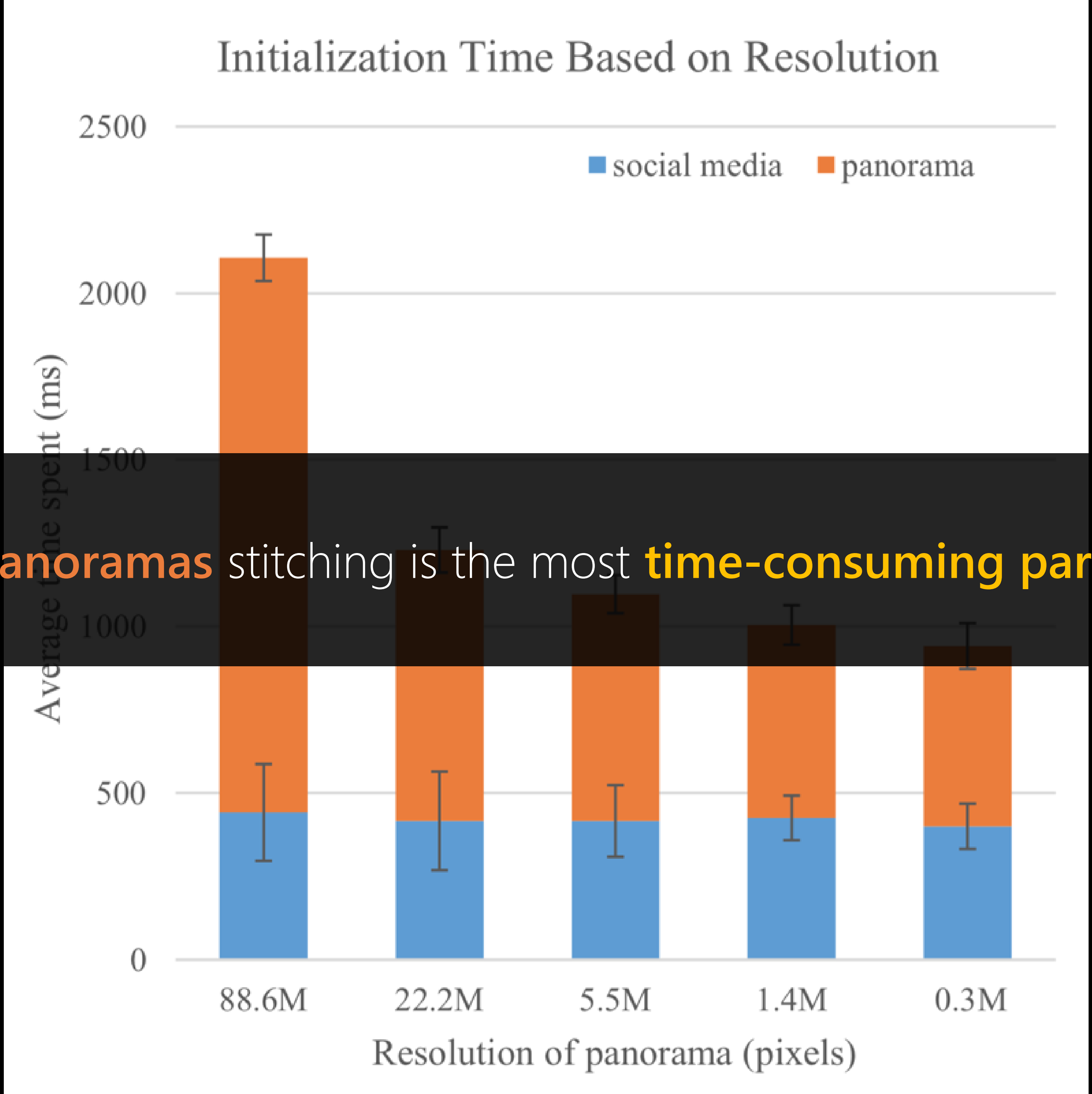


Preload everything into the memory:

~250ms

Initialization Time

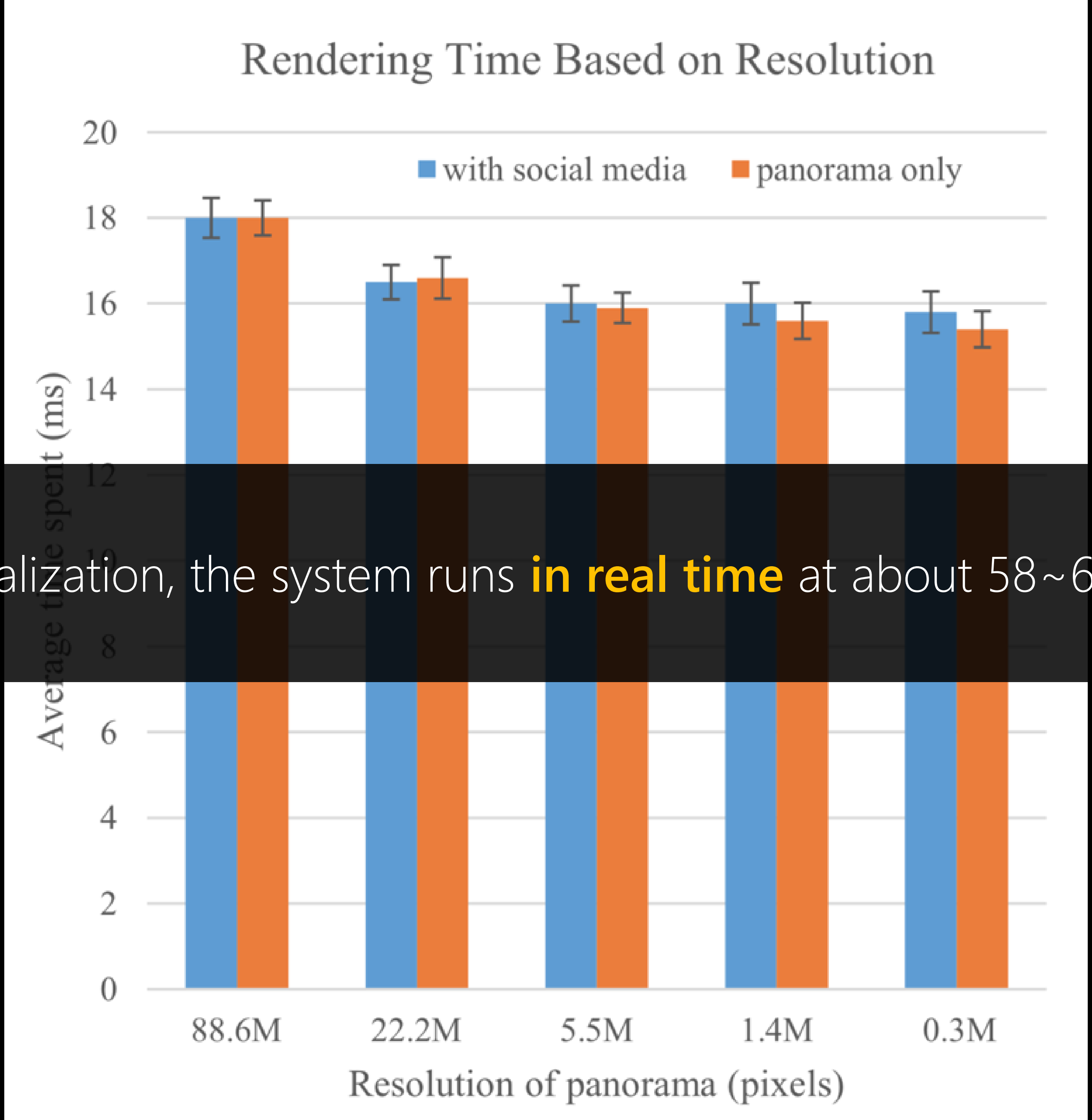
Panorama takes a while to load



Panoramas stitching is the most **time-consuming part**.

Rendering Time

Almost 58~60 FPS



After initialization, the system runs **in real time** at about 58~60 FPS.

How effective is the Maximal Poisson-disk Sampling for reducing the visual clutter?

Saliency Map

Hou *et al.* TPAMI 2011

How much social media **covers** the salient regions of the panorama?

Saliency Map

Hou *et al.* TPAMI 2011

Saliency maps use color, intensity, and orientation contrasts to estimate **where humans will look at**.

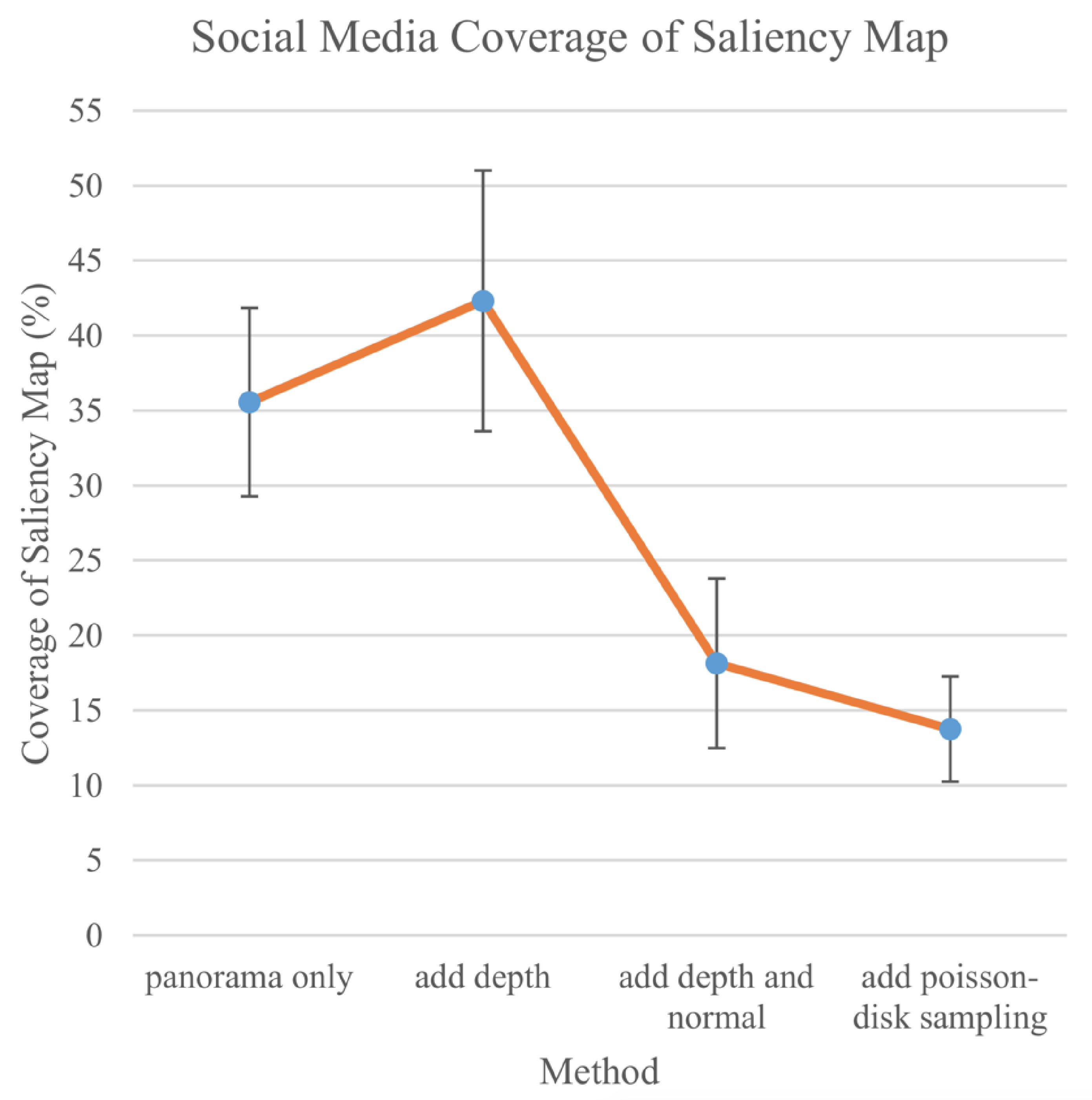
Saliency Map

Hou *et al.* TPAMI 2011

We prefer the social media to cover **less high-saliency regions**.

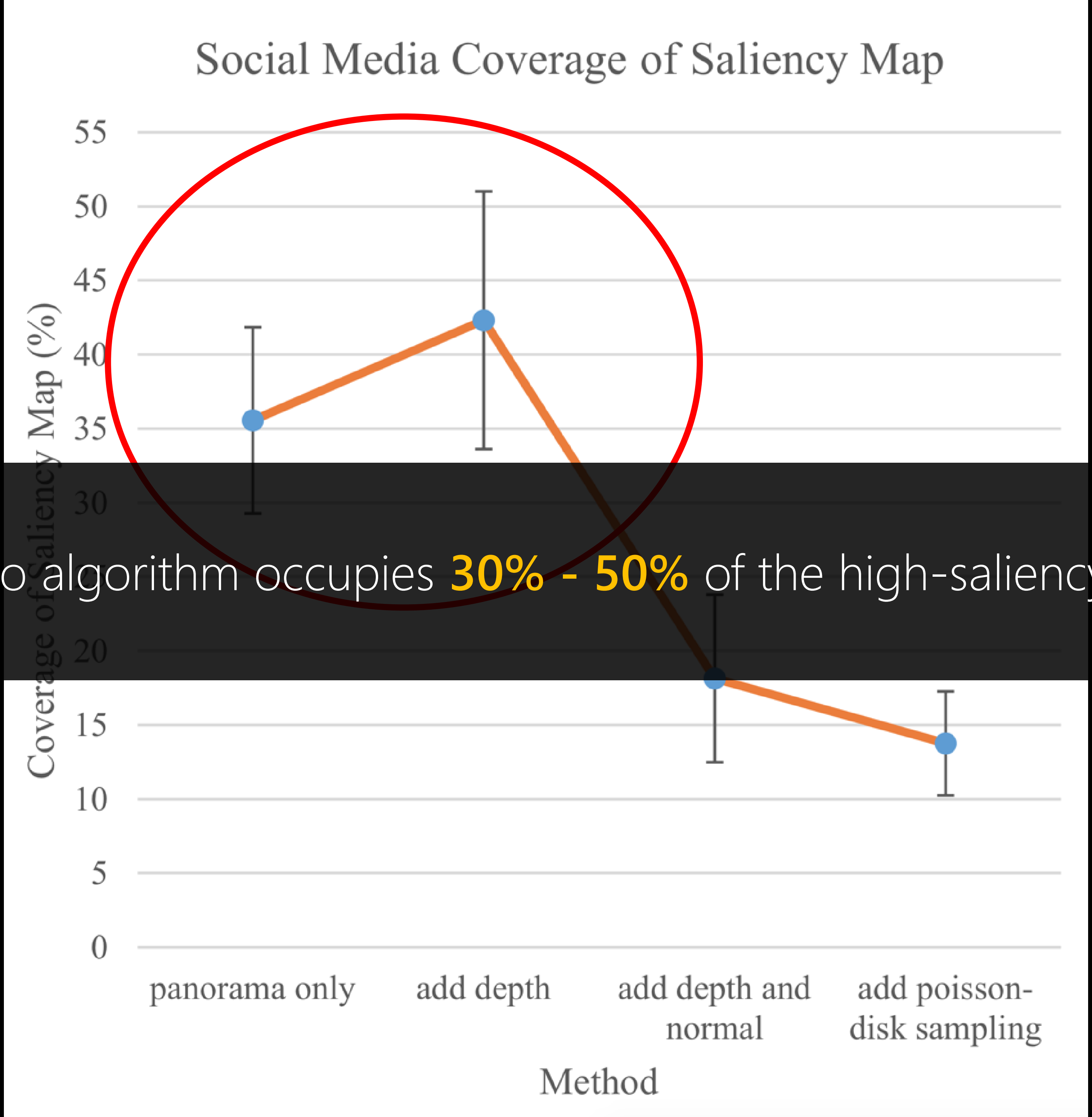
Social Media Coverage

100 panoramas for each algorithm



Social Media Coverage

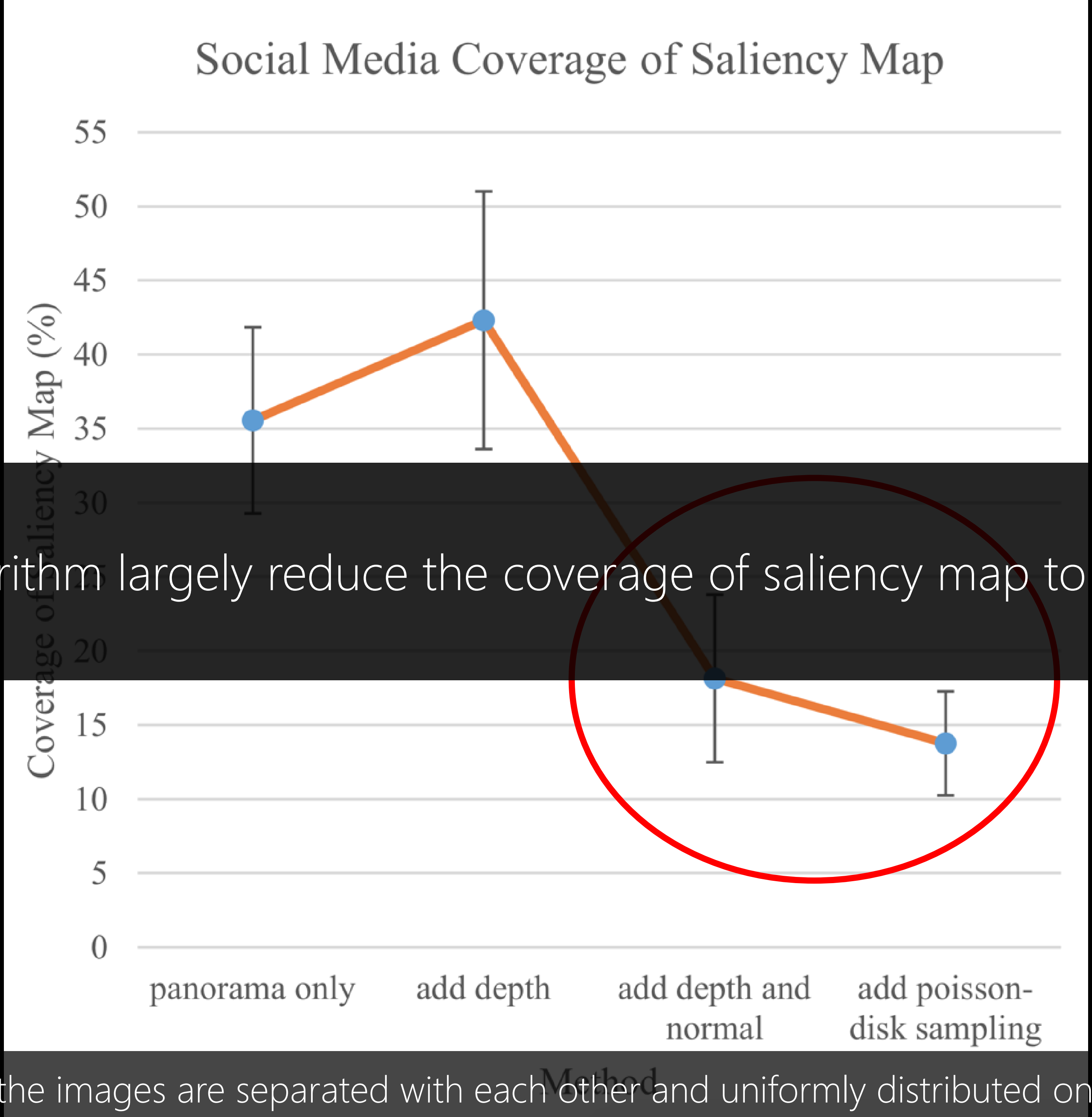
100 panoramas for each algorithm



The first two algorithm occupies **30% - 50%** of the high-saliency regions.

Social Media Coverage

100 panoramas for each algorithm



The last two algorithm largely reduce the coverage of saliency map to **less than 20%.**

In the last algorithm, all the images are separated with each other and uniformly distributed on the building surfaces.

What could be the **potential applications** of Social Street View?

“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*

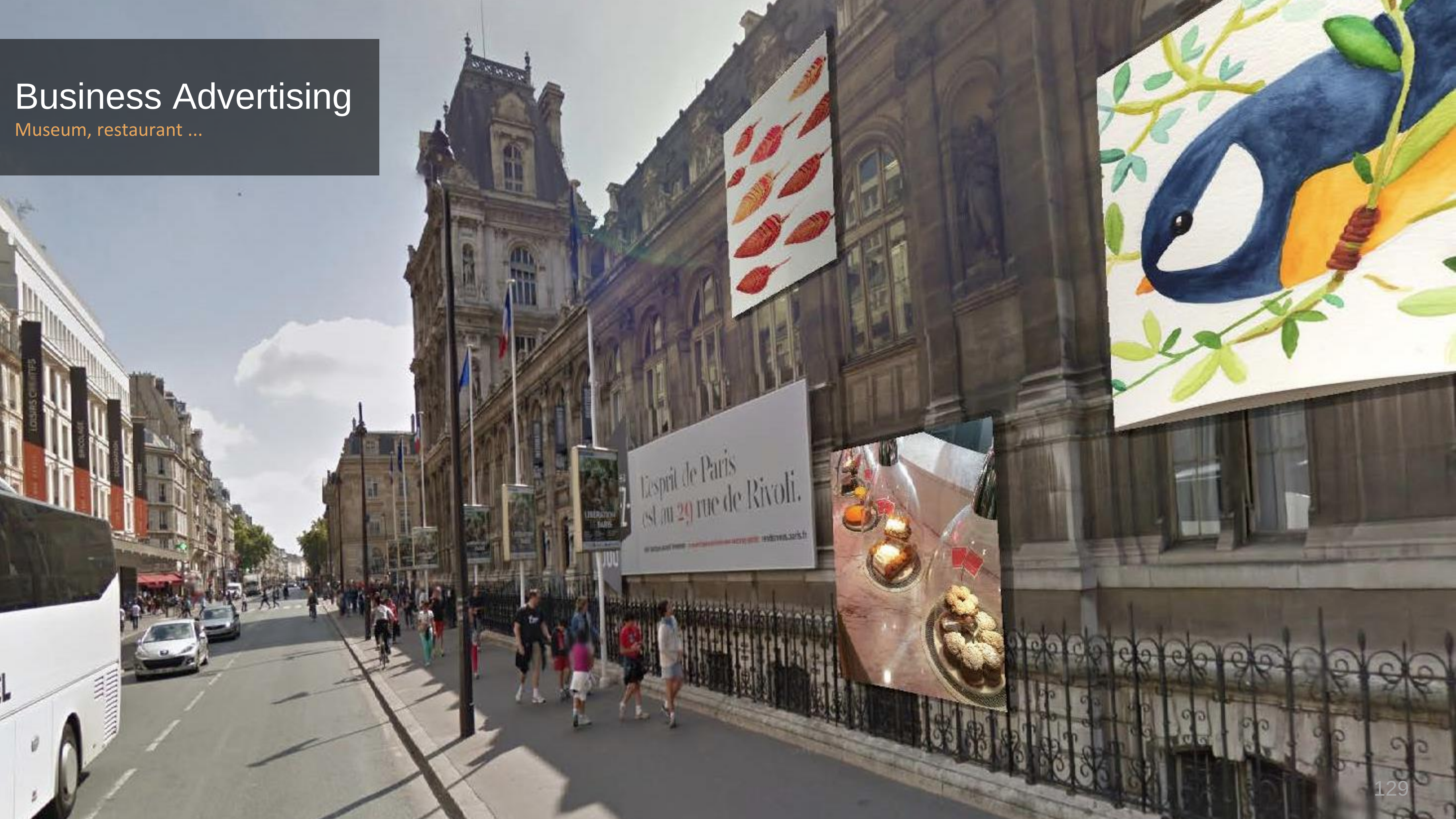
Business Advertising

Museum, restaurant, real-estate ...

Social Street View enhances future consumers' **visual memories** and makes it easier for them to seek the "amazing oysters" around this place.

Business Advertising

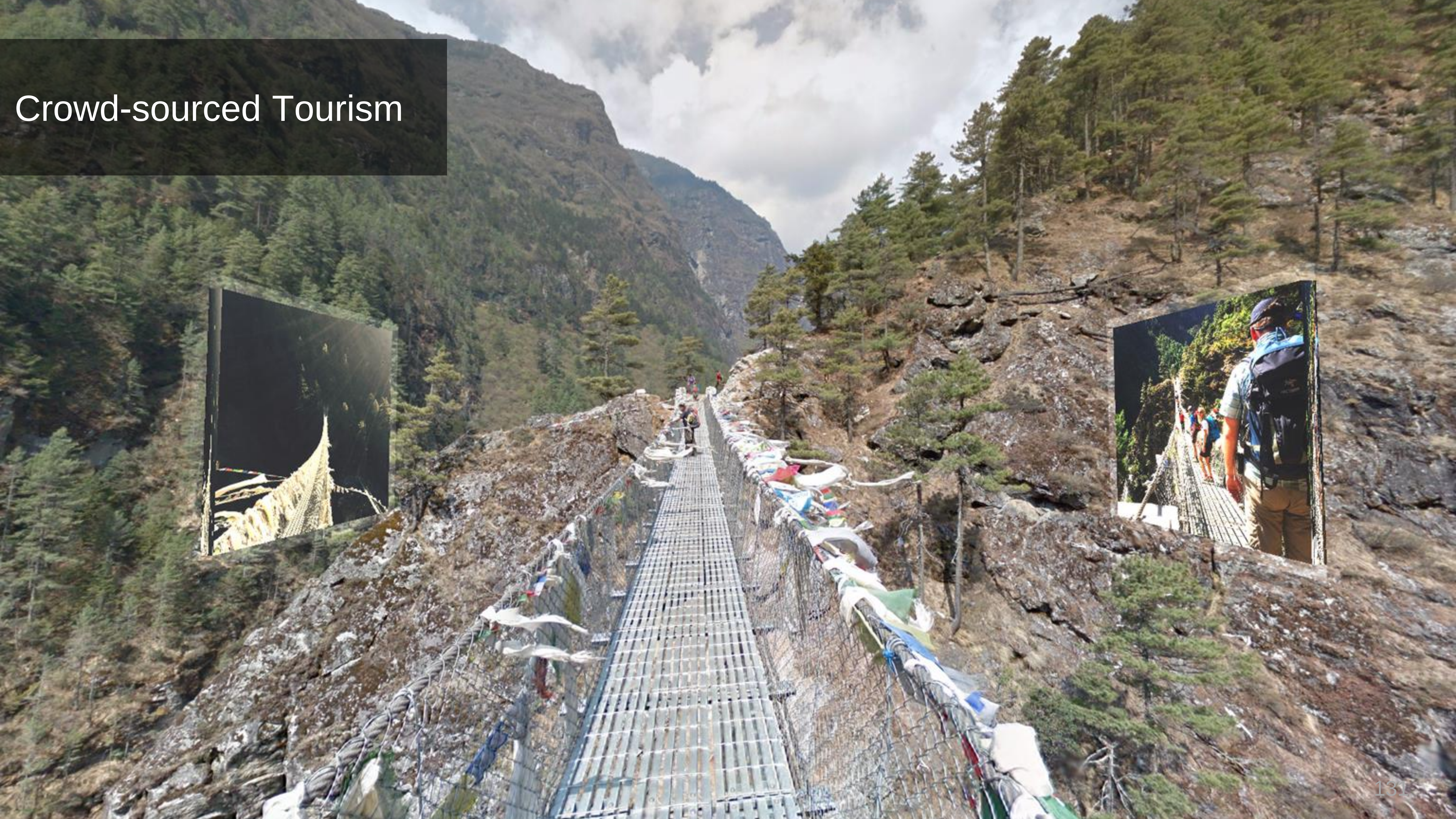
Museum, restaurant ...

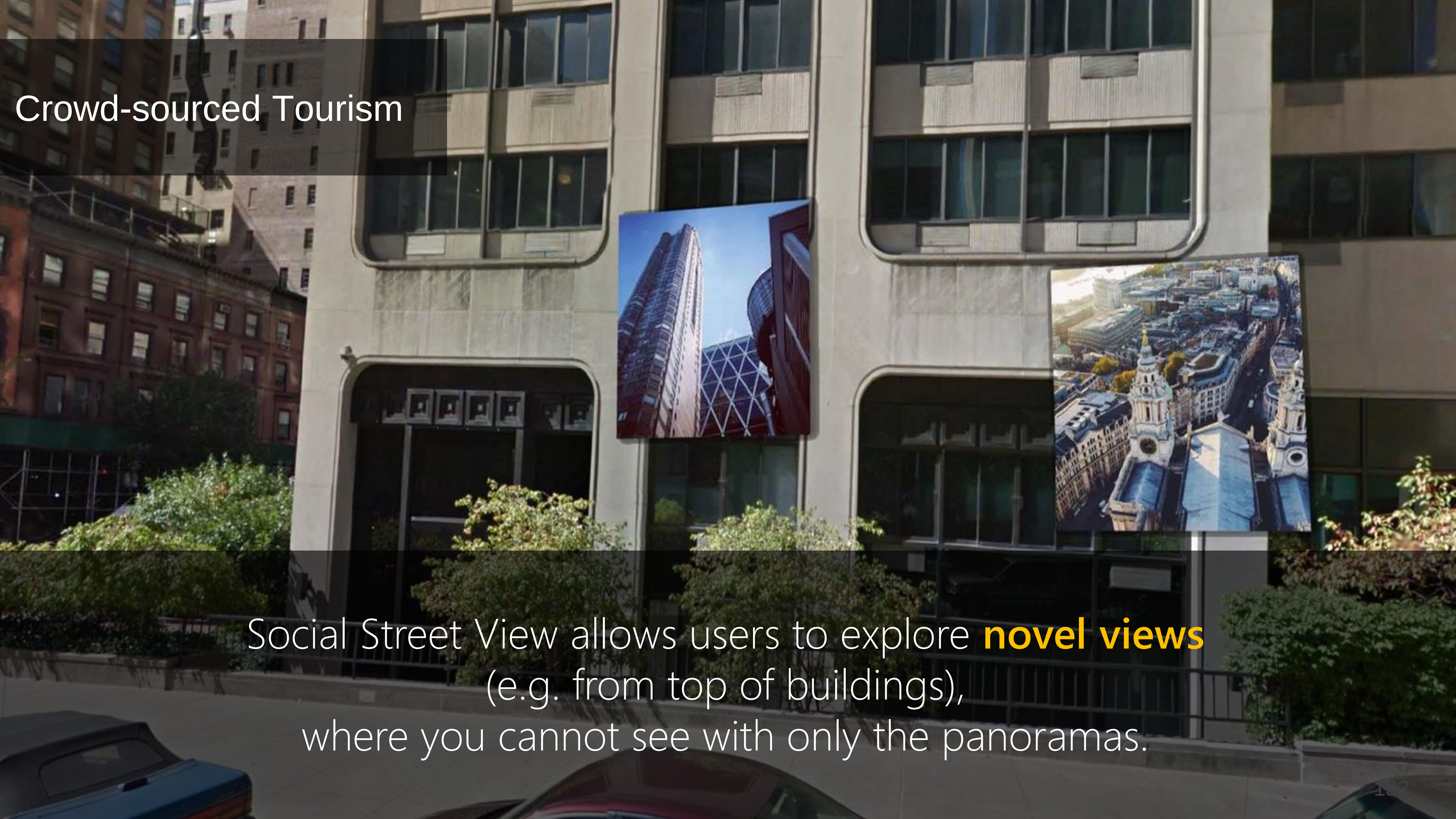


Learning Culture

Taierzhuang, Chinese Spring Festivals

Crowd-sourced Tourism





Crowd-sourced Tourism

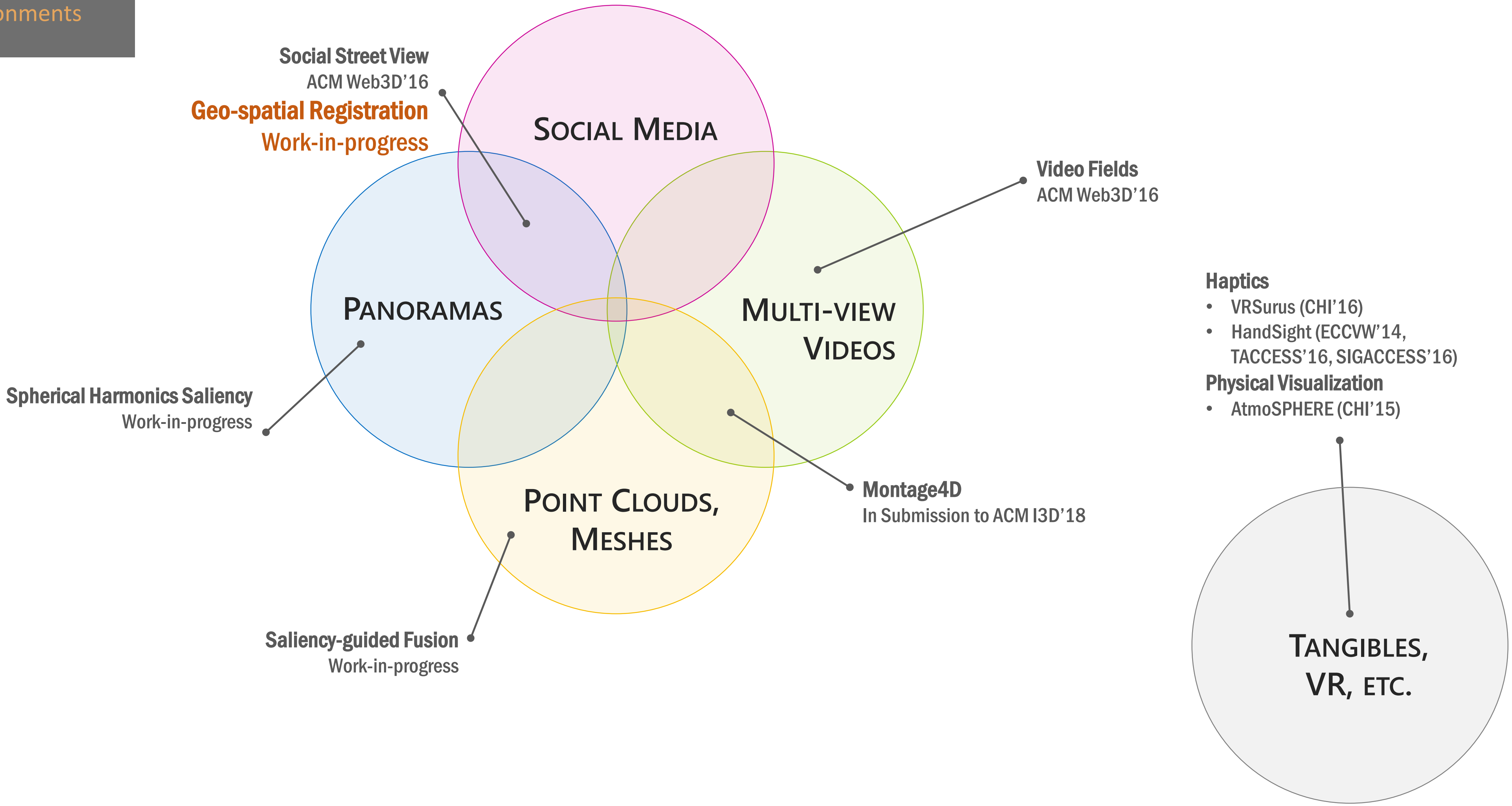
Social Street View allows users to explore **novel views**
(e.g. from top of buildings),
where you cannot see with only the panoramas.

In conclusion, **Social Street View** provides a novel solution for automatically fusing geo-tagged social media in an immersive 3D environments.


We envision **social media** and **panoramic videos** are
significant parts of the **big data** for VR and AR applications
UNIVERSITY OF MARYLAND
UMIACS

Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



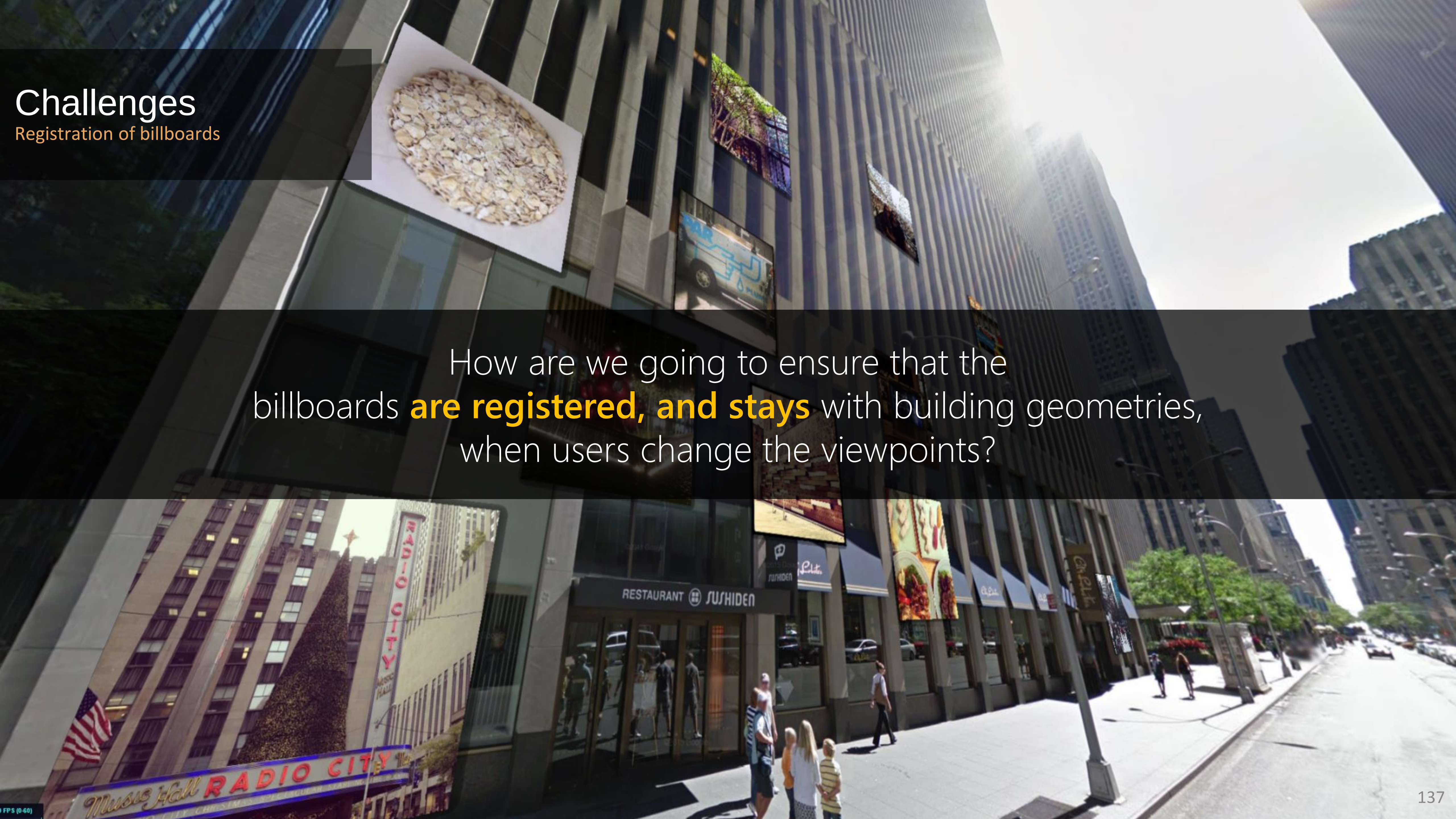
Challenges

Registration of billboards

The social media layout is
re-generated, every time,
when users walk around the street views.

Challenges

Registration of billboards



How are we going to ensure that the billboards **are registered, and stays** with building geometries, when users change the viewpoints?

Proposed Work

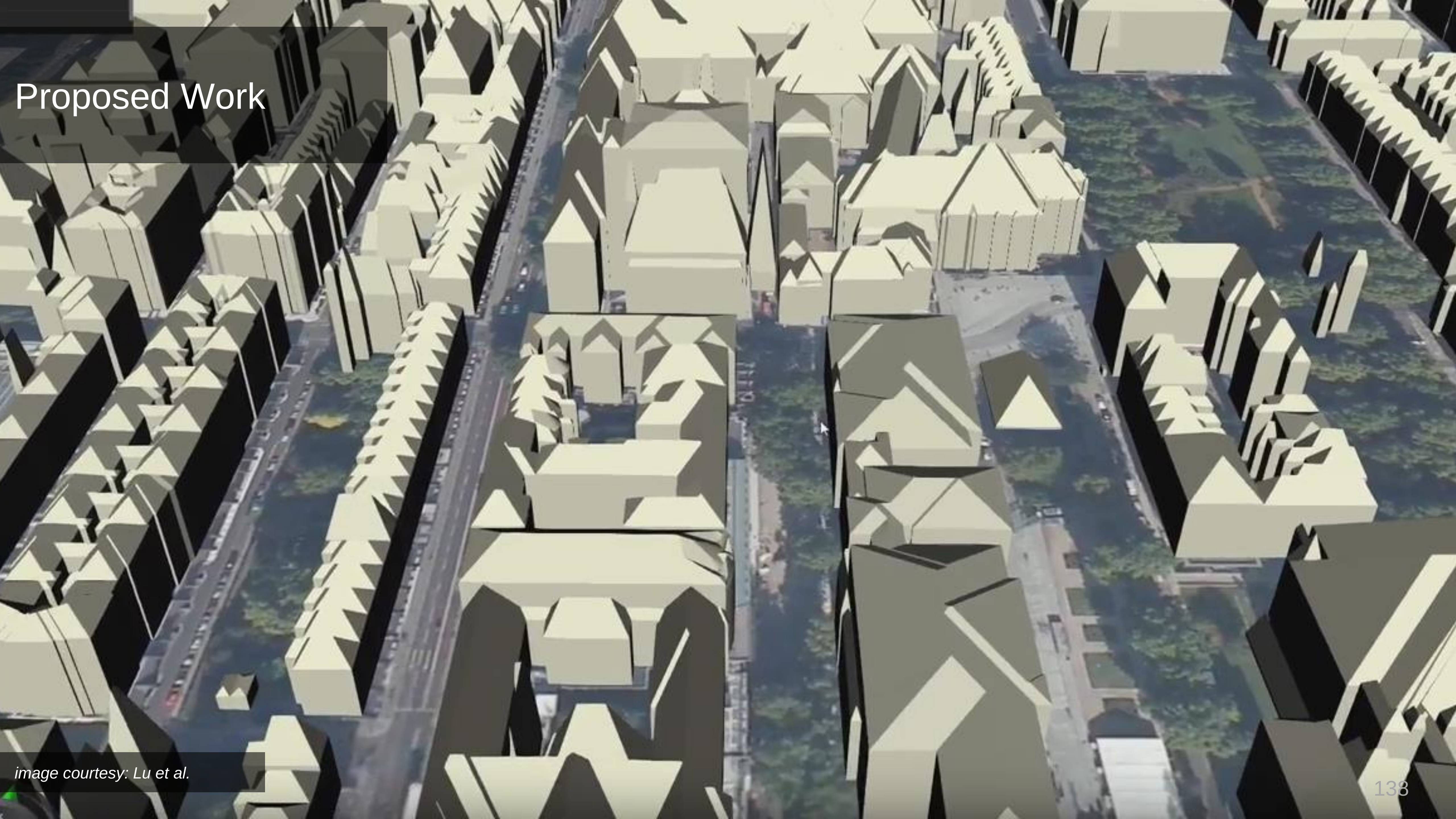


image courtesy: Lu et al.

Proposed Work

Estimate simple 3D building geometries from the depth and normal maps.

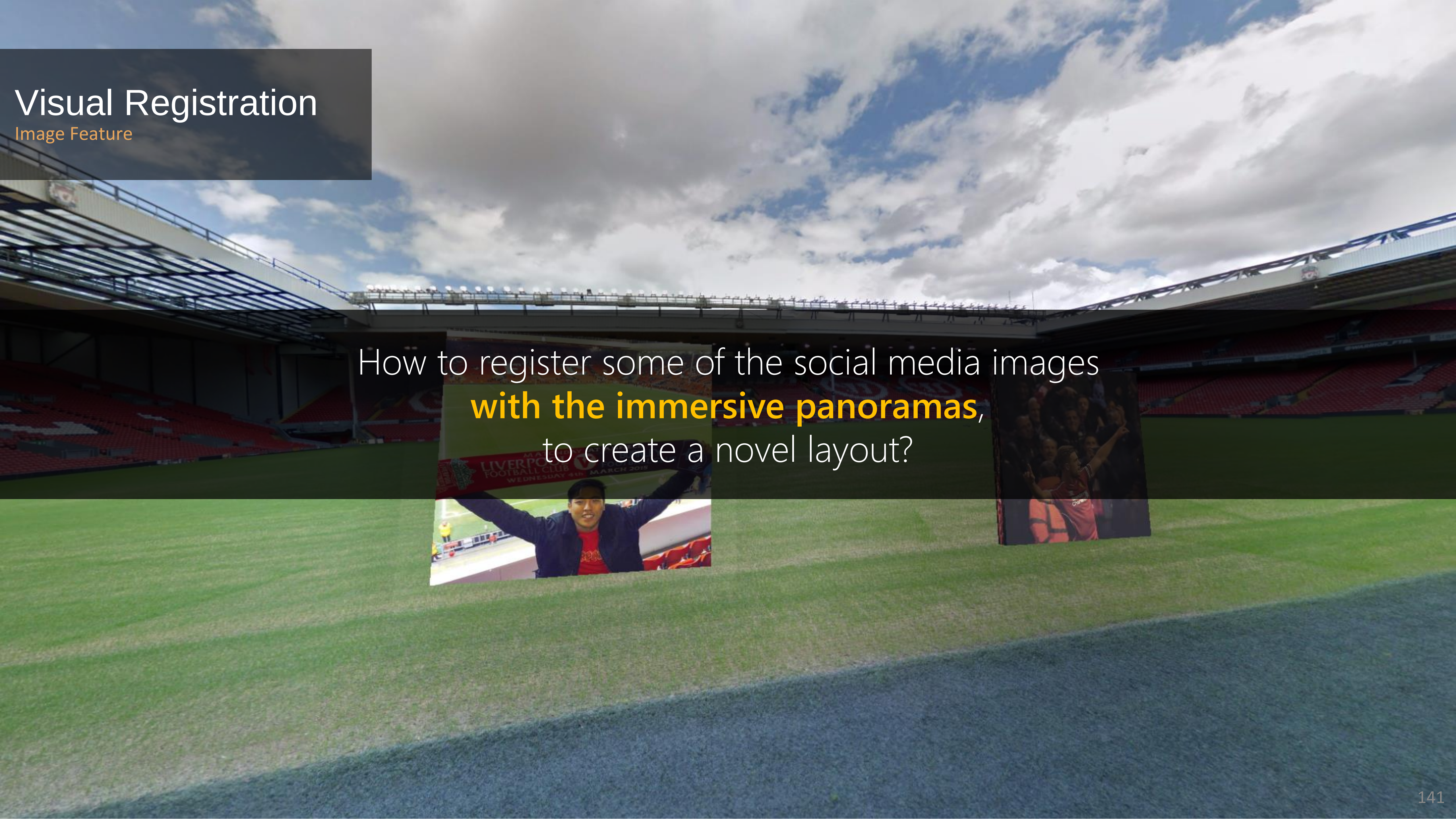
image courtesy: Lu et al.

Visual Registration

Image Feature

Visual Registration

Image Feature



How to register some of the social media images
with the immersive panoramas,
to create a novel layout?

Storytelling

Baja California Sur, Mexico

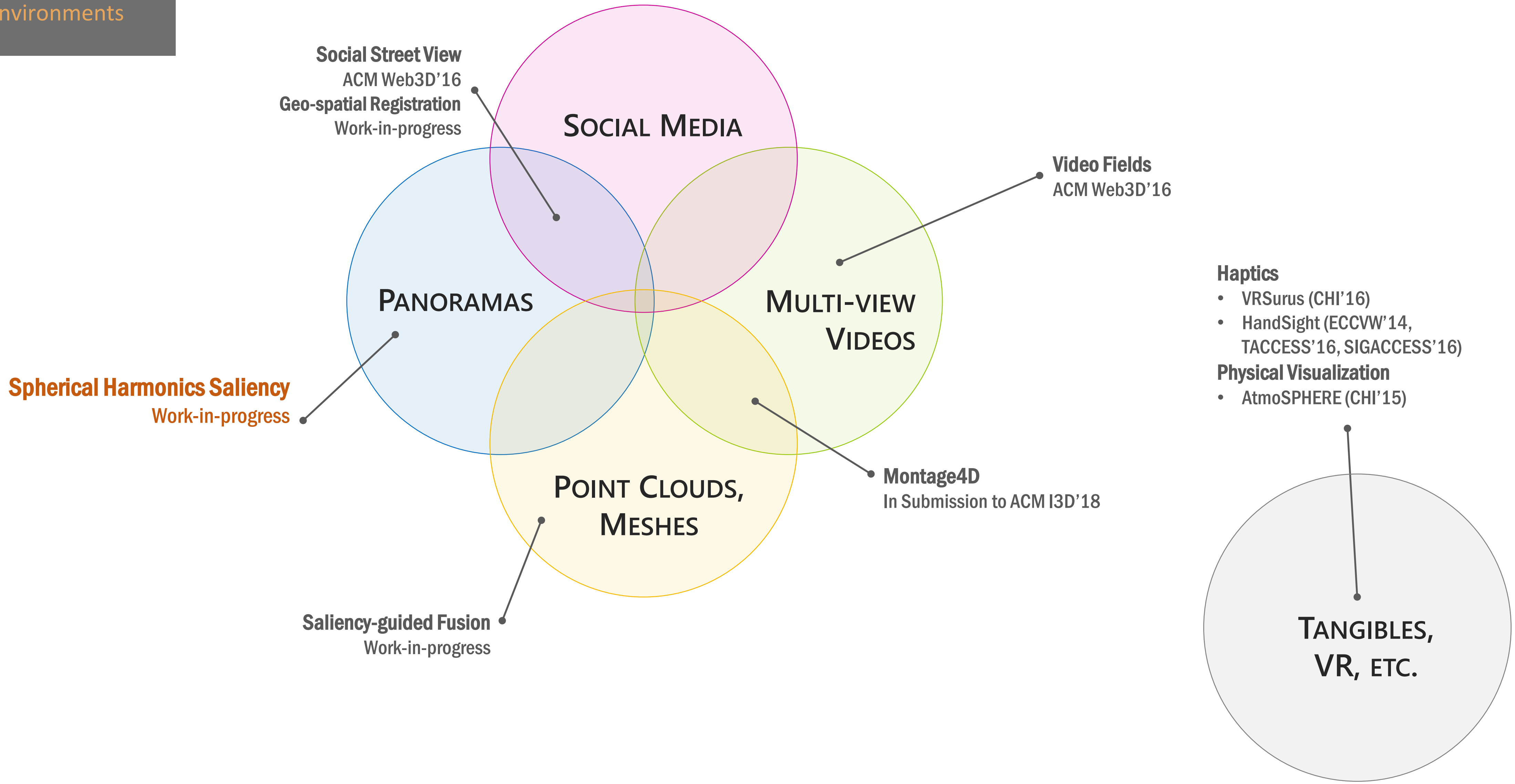
Storytelling

Baja California Sur, Mexico

Develop fast algorithms for
reconstructing simple geometries,
feature matching,
and **navigating the social street views without resampling.**

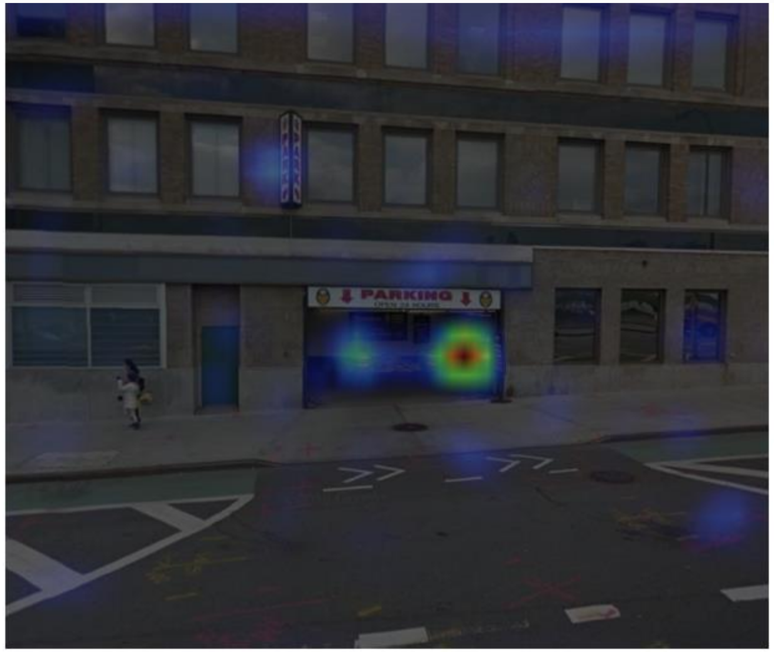
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



Motivation

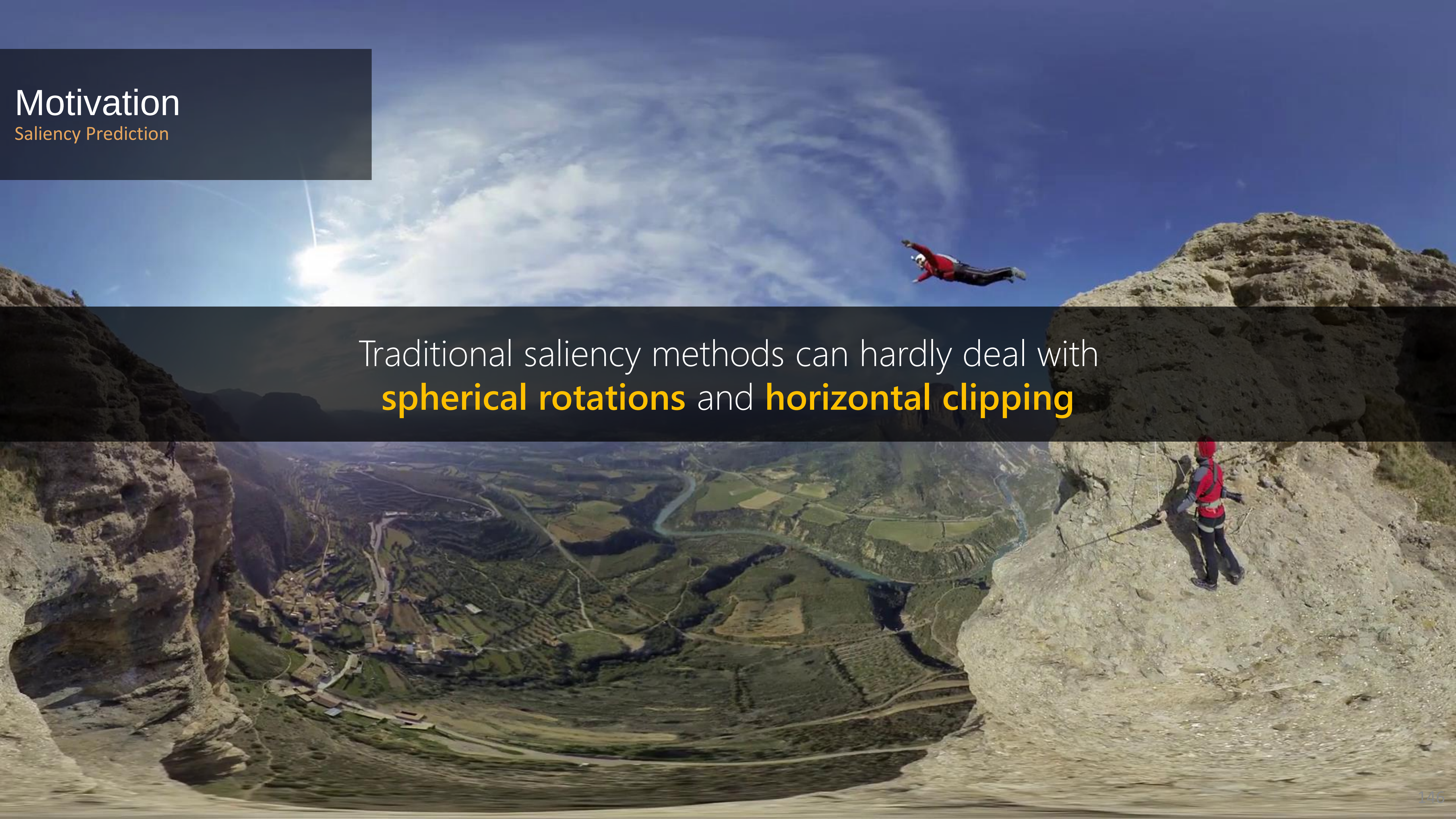
Social Media Overlay



How to efficiently generate the **saliency map for panoramas**?

Motivation

Saliency Prediction



Traditional saliency methods can hardly deal with **spherical rotations** and **horizontal clipping**

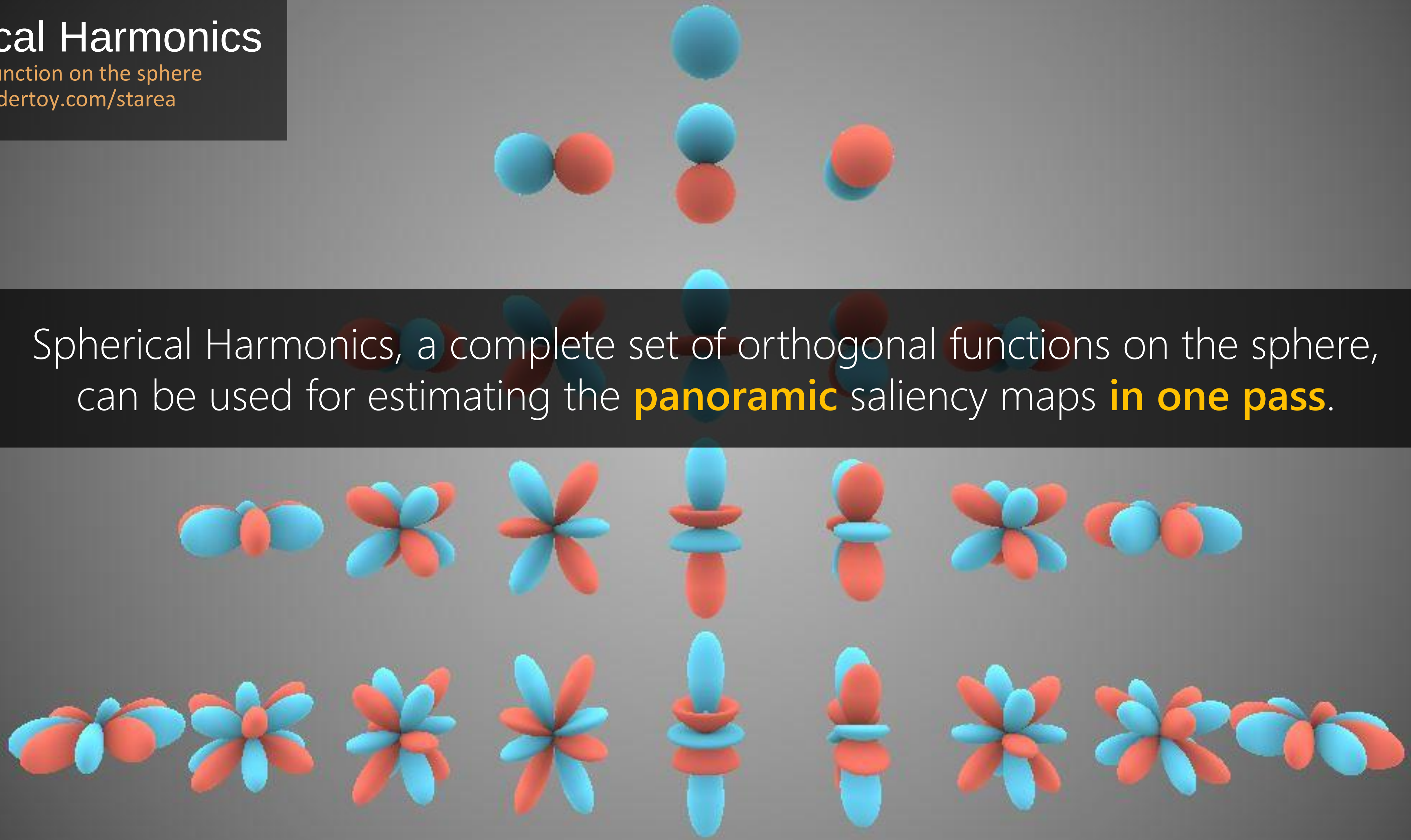
Motivation

Saliency Prediction

The sliding-window approach is too slow since it has to **cut the panorama into pieces**, and **stitch the saliency maps back**.

Spherical Harmonics

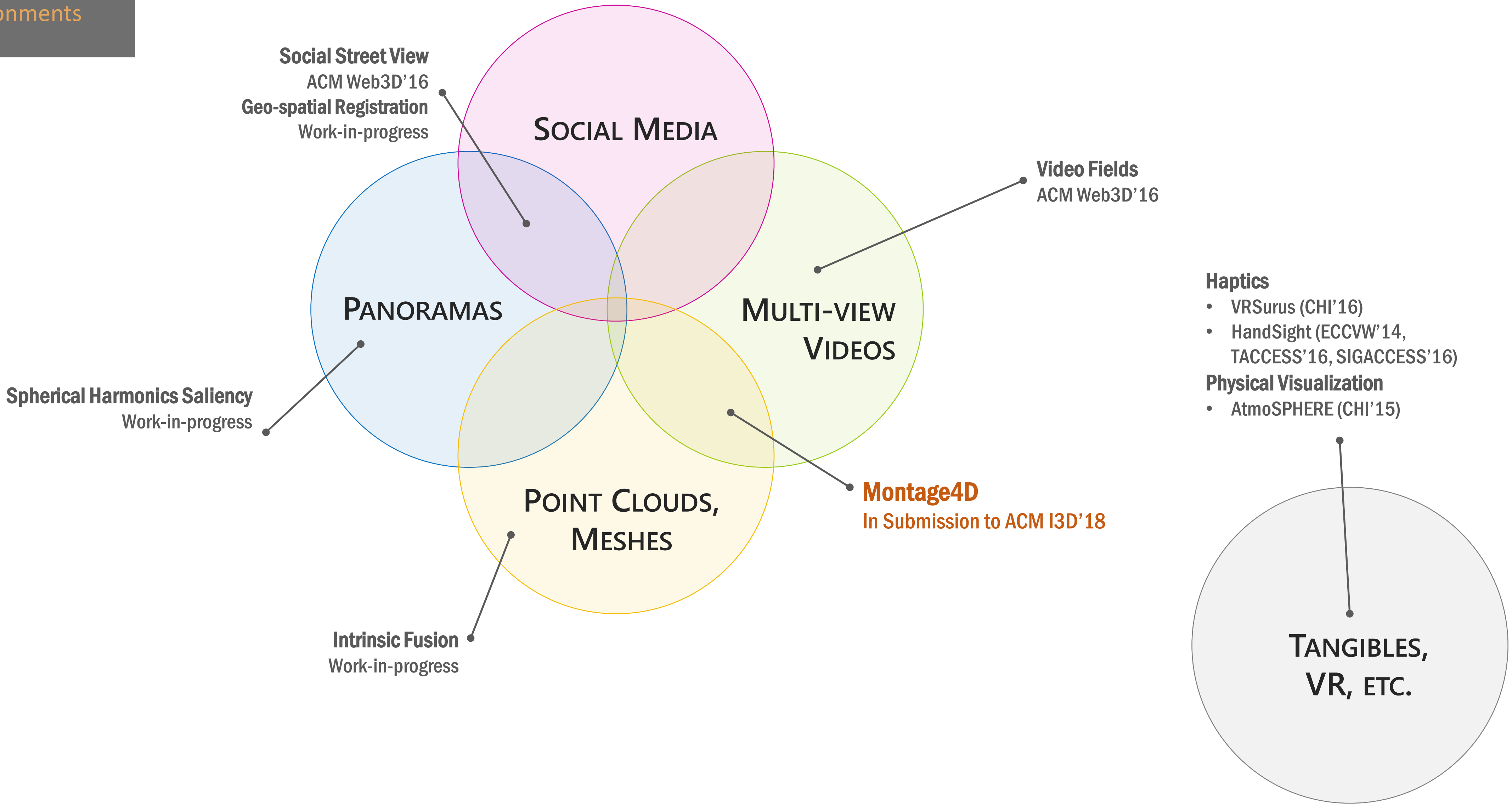
Orthogonal function on the sphere
Du. www.shadertoy.com/starea



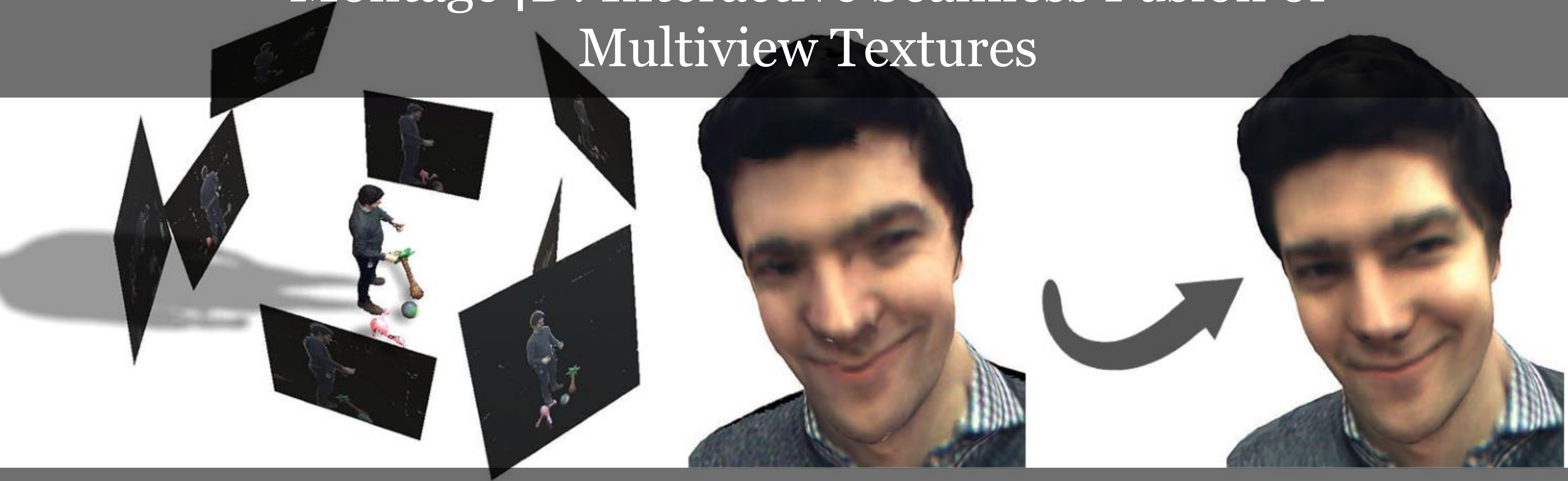
Spherical Harmonics, a complete set of orthogonal functions on the sphere, can be used for estimating the **panoramic** saliency maps **in one pass**.

Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



Montage4D: Interactive Seamless Fusion of Multiview Textures



Ruofei Du, Ming Chuang, Wayne Chang, Hugues Hoppe, and Amitabh Varshney

University of Maryland, College Park

Microsoft Research, Redmond

THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LABORATORY
AT THE UNIVERSITY OF MARYLAND

Microsoft Research

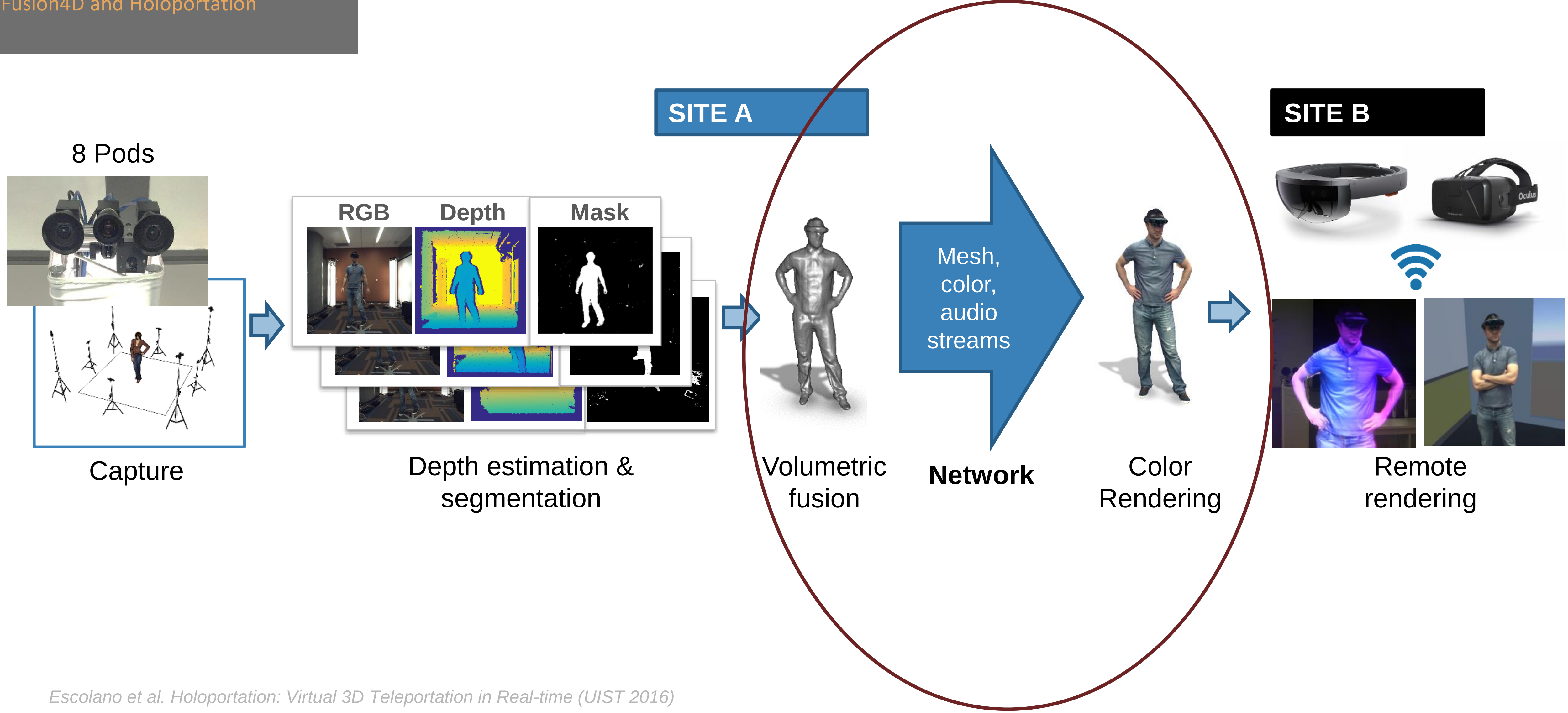
UMIACS

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

UNIVERSITY OF MARYLAND

Introduction

Fusion4D and Holoportation



Escolano et al. Holoportation: Virtual 3D Teleportation in Real-time (UIST 2016)

Fusing multiview video textures onto dynamic meshes with real-time constraint is **a challenging task**

30%

of the participants does not believe the 3D reconstructed person looks real

Motivation

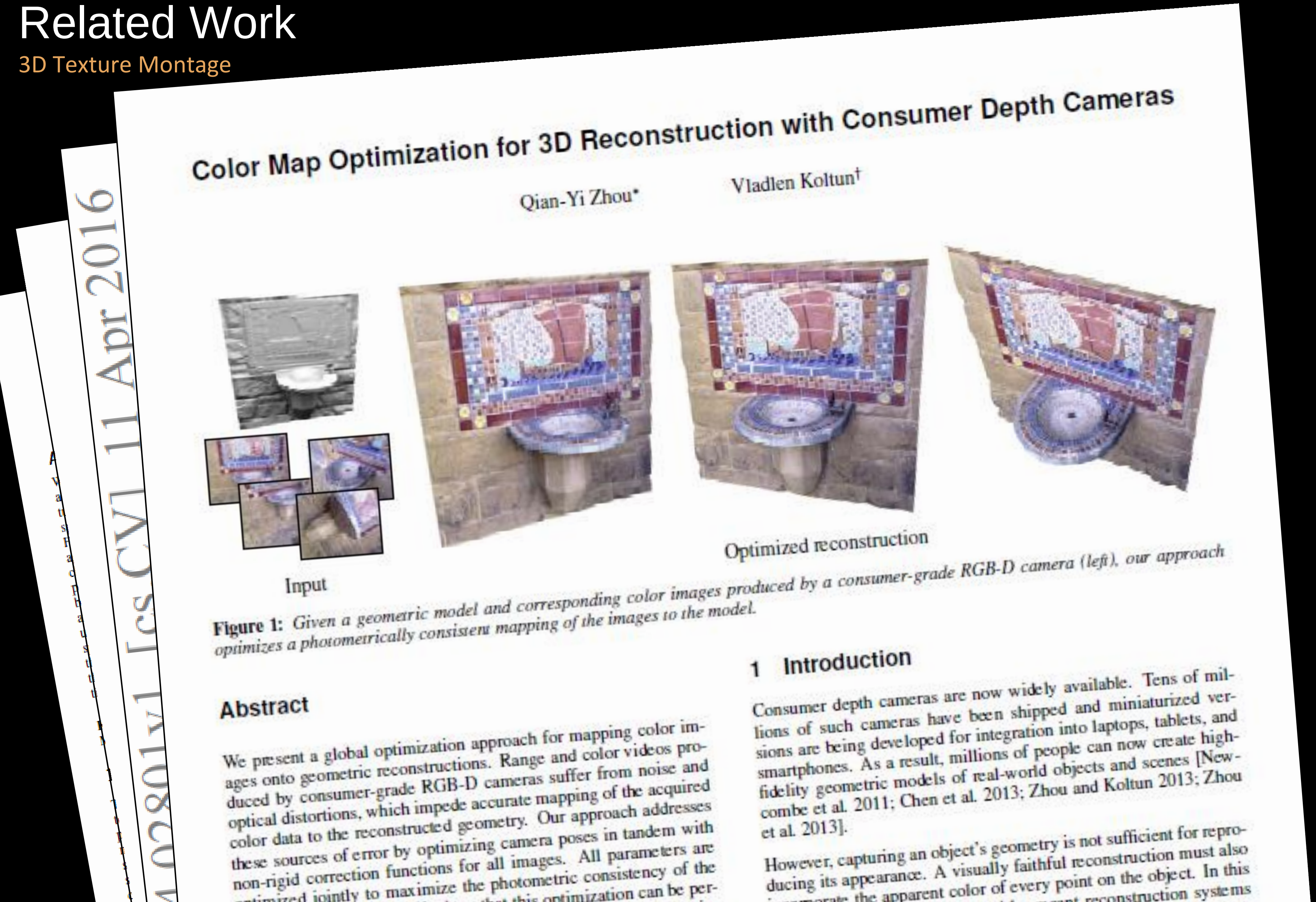
Visual Quality Matters

Motivation

Visual Quality Matters

Related Work

3D Texture Montage



Color Map Optimization for 3D Reconstruction with Consumer Depth Cameras

Qian-Yi Zhou* Vladlen Koltun†

Input

Optimized reconstruction

Figure 1: *Given a geometric model and corresponding color images produced by a consumer-grade RGB-D camera (left), our approach optimizes a photometrically consistent mapping of the images to the model.*

Abstract

We present a global optimization approach for mapping color images onto geometric reconstructions. Range and color videos produced by consumer-grade RGB-D cameras suffer from noise and optical distortions, which impede accurate mapping of the acquired color data to the reconstructed geometry. Our approach addresses these sources of error by optimizing camera poses in tandem with non-rigid correction functions for all images. All parameters are optimized jointly to maximize the photometric consistency of the

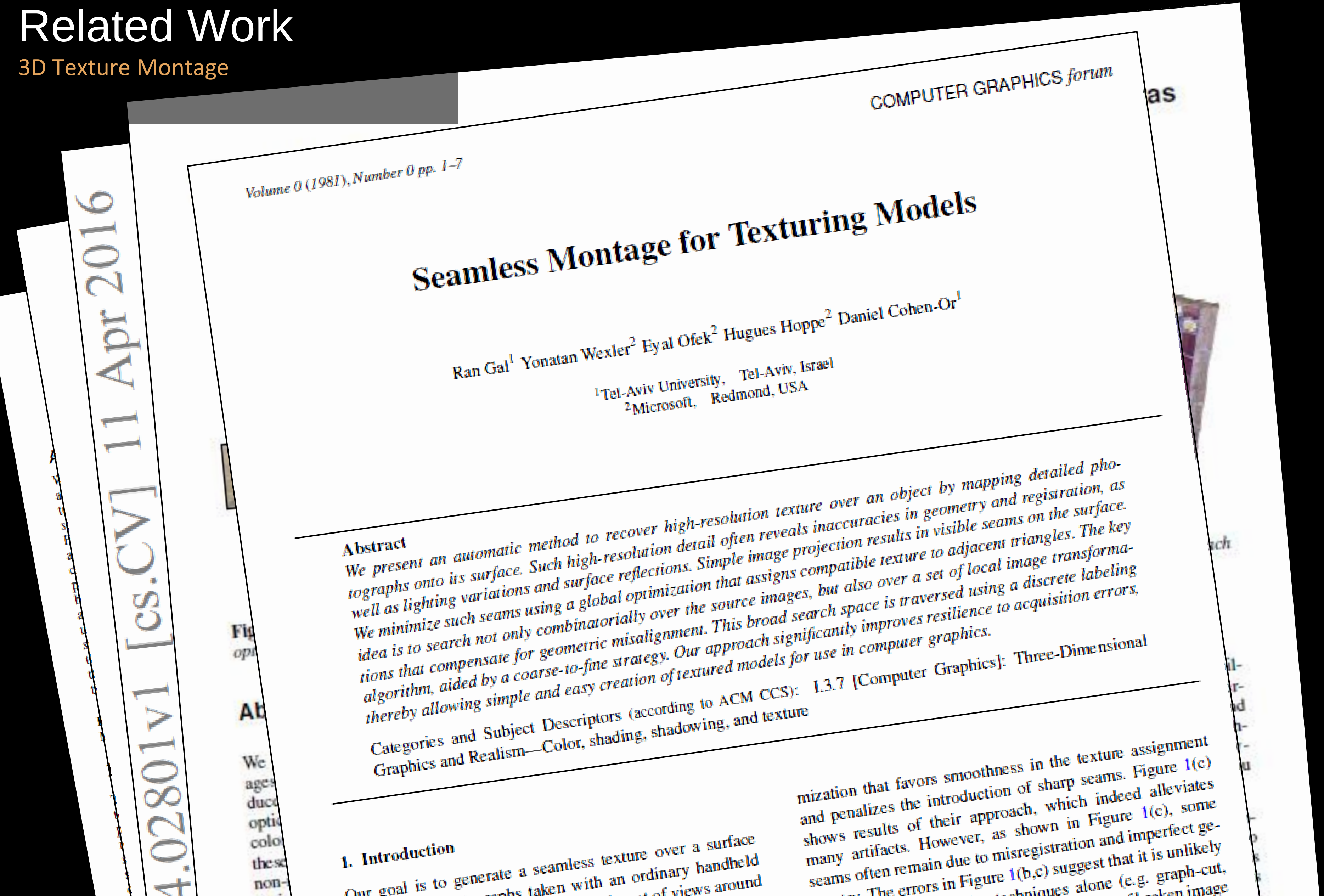
1 Introduction

Consumer depth cameras are now widely available. Tens of millions of such cameras have been shipped and miniaturized versions are being developed for integration into laptops, tablets, and smartphones. As a result, millions of people can now create high-fidelity geometric models of real-world objects and scenes [Newcombe et al. 2011; Chen et al. 2013; Zhou and Koltun 2013; Zhou et al. 2013].

However, capturing an object's geometry is not sufficient for reproducing its appearance. A visually faithful reconstruction must also ... the apparent color of every point on the object. In this

Related Work

3D Texture Montage



COMPUTER GRAPHICS *forum*

Volume 0 (1981), Number 0 pp. 1–7

Seamless Montage for Texturing Models

Ran Gal[1] Yonatan Wexler[2] Eyal Ofek[2] Hugues Hoppe[2] Daniel Cohen-Or[1]

[1]Tel-Aviv University, Tel-Aviv, Israel
[2]Microsoft, Redmond, USA

Abstract

We present an automatic method to recover high-resolution texture over an object by mapping detailed photographs onto its surface. Such high-resolution detail often reveals inaccuracies in geometry and registration, as well as lighting variations and surface reflections. Simple image projection results in visible seams on the surface. We minimize such seams using a global optimization that assigns compatible texture to adjacent triangles. The key idea is to search not only combinatorially over the source images, but also over a set of local image transformations that compensate for geometric misalignment. This broad search space is traversed using a discrete labeling algorithm, aided by a coarse-to-fine strategy. Our approach significantly improves resilience to acquisition errors, thereby allowing simple and easy creation of textured models for use in computer graphics.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

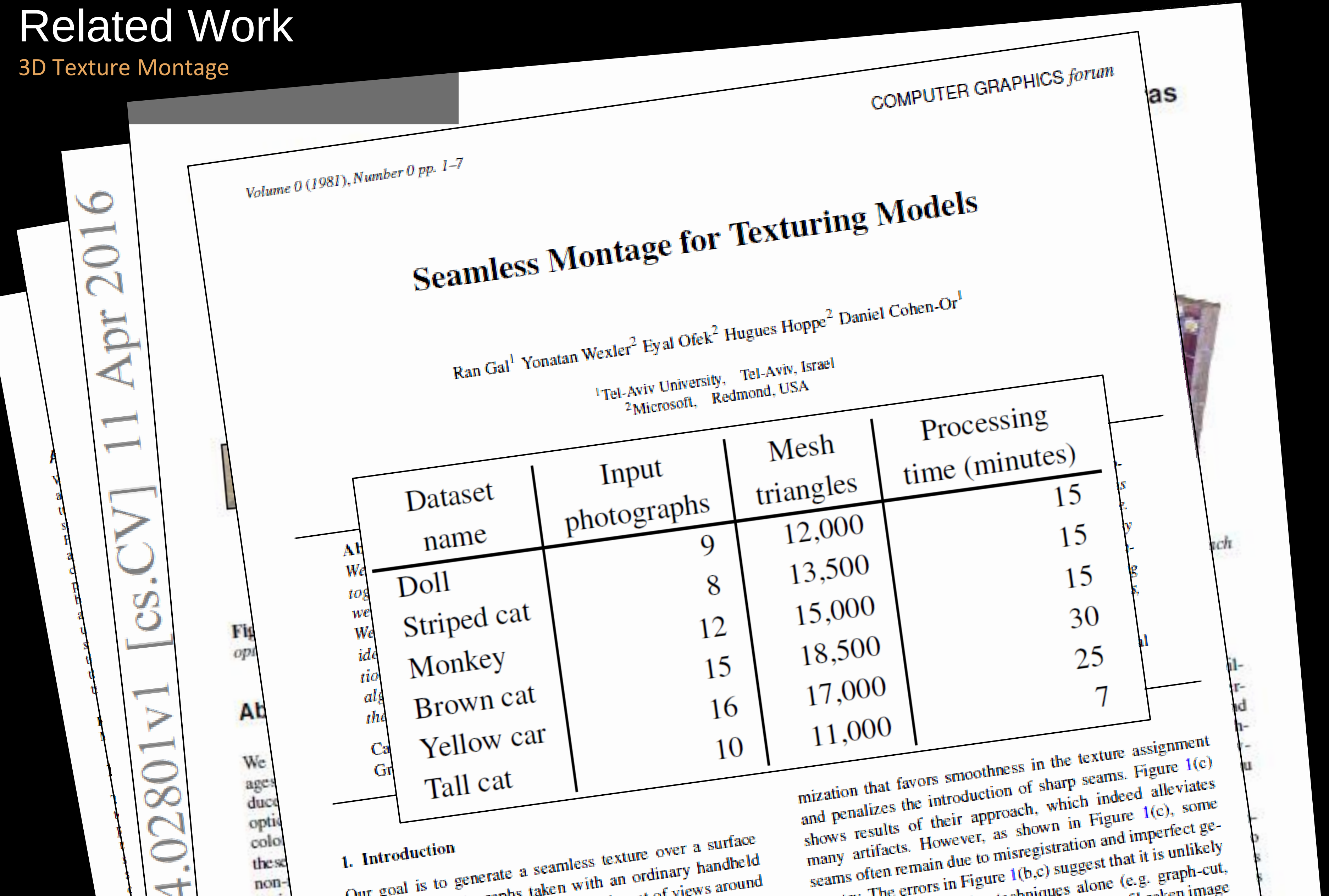
1. Introduction

Our goal is to generate a seamless texture over a surface ...

... mization that favors smoothness in the texture assignment and penalizes the introduction of sharp seams. Figure 1(c) shows results of their approach, which indeed alleviates many artifacts. However, as shown in Figure 1(c), some seams often remain due to misregistration and imperfect geometry. The errors in Figure 1(b,c) suggest that it is unlikely ...

Related Work

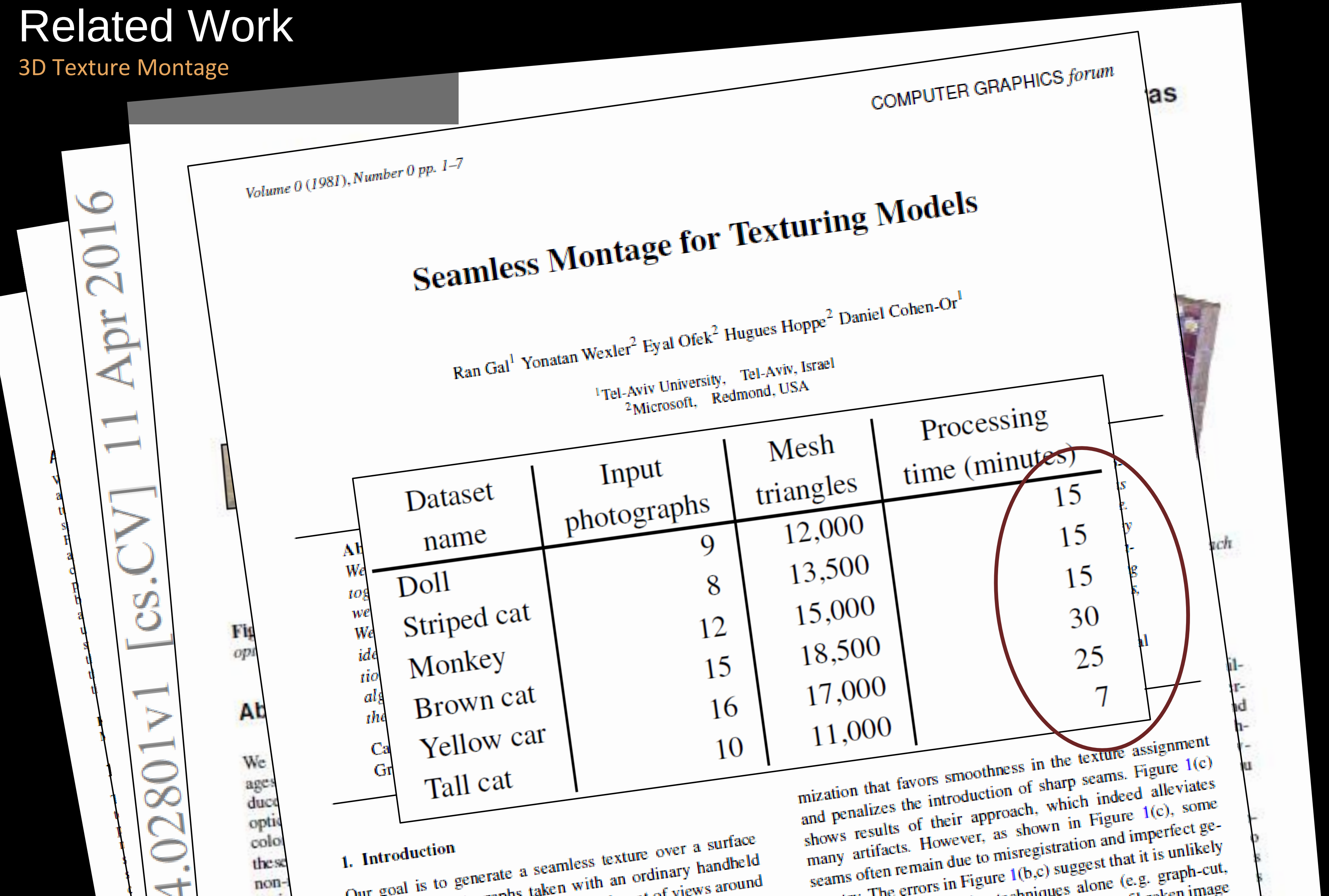
3D Texture Montage



Dataset name	Input photographs	Mesh triangles	Processing time (minutes)
Doll	9	12,000	15
Striped cat	8	13,500	15
Monkey	12	15,000	15
Brown cat	15	18,500	30
Yellow car	16	17,000	25
Tall cat	10	11,000	7

Related Work

3D Texture Montage



4.02801v1 [cs.CV] 11 Apr 2016

COMPUTER GRAPHICS *forum*

Volume 0 (1981), Number 0 pp. 1–7

Seamless Montage for Texturing Models

Ran Gal[1] Yonatan Wexler[2] Eyal Ofek[2] Hugues Hoppe[2] Daniel Cohen-Or[1]

[1]Tel-Aviv University, Tel-Aviv, Israel
[2]Microsoft, Redmond, USA

Dataset name	Input photographs	Mesh triangles	Processing time (minutes)
Doll	9	12,000	15
Striped cat	8	13,500	15
Monkey	12	15,000	15
Brown cat	15	18,500	30
Yellow car	16	17,000	25
Tall cat	10	11,000	7

1. Introduction

Our goal is to generate a seamless texture over a surface ...

mization that favors smoothness in the texture assignment and penalizes the introduction of sharp seams. Figure 1(c) shows results of their approach, which indeed alleviates many artifacts. However, as shown in Figure 1(c), some seams often remain due to misregistration and imperfect geometry. The errors in Figure 1(b,c) suggest that it is unlikely ...

Related Work

3D Texture Montage

High-Quality Streamable Free-Viewpoint Video

Alvaro Collet Ming Chuang Pat Sweeney Don Gillett Dennis Evseev David Calabrese
Hugues Hoppe Adam Kirk Steve Sullivan
Microsoft Corporation

Figure 1: *Examples of reconstructed free-viewpoint video acquired by our system.*

Abstract

We present the first end-to-end solution to create high-quality free-viewpoint video encoded as a compact data stream. Our system records performances using a dense set of RGB and IR video cameras, generates dynamic textured surfaces, and compresses these to a streamable 3D video format. Four technical advances contribute to high fidelity and robustness: multimodal multi-view stereo fusing RGB, IR, and silhouette information; adaptive meshing guided by automatic detection of perceptually salient areas; mesh tracking to create temporally coherent subsequences; and encoding of tracked textured meshes as an MPEG video stream. Quantitative

Our goal is to transform free-viewpoint video from research prototype into a rich and accessible form of media. Several system components must work together to achieve this goal: capture rigs must be easy to reconfigure and support professional production workflows; reconstruction must be automatic and scalable to high processing throughput; and results must be compressible to a data rate close to common media formats. Visual quality from any angle must be on par with traditional video, and the format must be renderable in real-time on a wide range of consumer devices.

In this paper, we discuss how we address these challenges to create an end-to-end system for realistic, streamable free-viewpoint video at significantly higher quality than the state of the art. Our approach does not require prior knowledge of the scene content. It handles

Related Work

3D Texture Montage

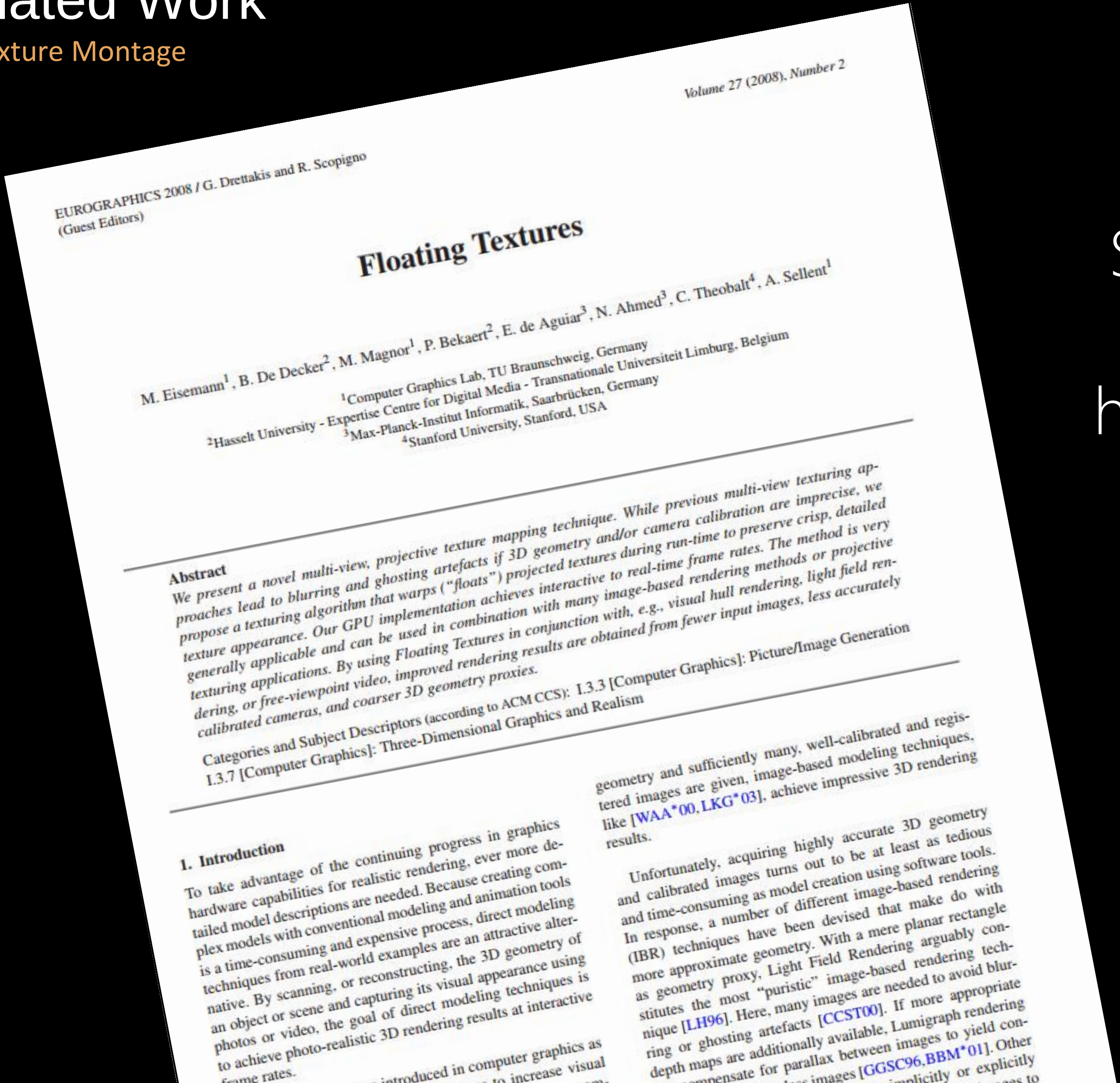
High-Quality Streamable Free-Viewpoint Video

Alvaro Collet Ming Chuang Pat Sweeney Don Gillett Dennis Evseev David Calabrese
Hugues Hoppe Adam Kirk Steve Sullivan
Microsoft Corporation

Scene	Frames	Avg points	Avg time/ frame (s)	Avg frames/ keyframe	Output Mbps
D.Duo	886	2.25M	29.1	24.6	12.1
Dress	1157	1.44M	27.7	27.5	8.6
Kendo	740	2.01M	26.0	35.2	8.3
Haka	173	2.70M	28.2	57.7	12.0
Lincoln	508	1.60M	27.5	72.6	7.9

Related Work

3D Texture Montage


EUROGRAPHICS 2008 / G. Drettakis and R. Scopigno
(Guest Editors)

Volume 27 (2008), Number 2

Floating Textures

M. Eisemann[1], B. De Decker[2], M. Magnor[1], P. Bekaert[2], E. de Aguiar[3], N. Ahmed[3], C. Theobalt[4], A. Sellent[1]

[1]Computer Graphics Lab, TU Braunschweig, Germany
[2]Hasselt University - Expertise Centre for Digital Media - Transnationale Universiteit Limburg, Belgium
[3]Max-Planck-Institut Informatik, Saarbrücken, Germany
[4]Stanford University, Stanford, USA

Abstract

We present a novel multi-view, projective texture mapping technique. While previous multi-view texturing approaches lead to blurring and ghosting artefacts if 3D geometry and/or camera calibration are imprecise, we propose a texturing algorithm that warps ("floats") projected textures during run-time to preserve crisp, detailed texture appearance. Our GPU implementation achieves interactive to real-time frame rates. The method is very generally applicable and can be used in combination with many image-based rendering methods or projective texturing applications. By using Floating Textures in conjunction with, e.g., visual hull rendering, light field rendering, or free-viewpoint video, improved rendering results are obtained from fewer input images, less accurately calibrated cameras, and coarser 3D geometry proxies.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

To take advantage of the continuing progress in graphics hardware capabilities for realistic rendering, ever more detailed model descriptions are needed. Because creating complex models with conventional modeling and animation tools is a time-consuming and expensive process, direct modeling techniques from real-world examples are an attractive alternative. By scanning, or reconstructing, the 3D geometry of an object or scene and capturing its visual appearance using photos or video, the goal of direct modeling techniques is to achieve photo-realistic 3D rendering results at interactive frame rates.

geometry and sufficiently many, well-calibrated and registered images are given, image-based modeling techniques, like [WAA*00, LKG*03], achieve impressive 3D rendering results.

Unfortunately, acquiring highly accurate 3D geometry and calibrated images turns out to be at least as tedious and time-consuming as model creation using software tools. In response, a number of different image-based rendering (IBR) techniques have been devised that make do with more approximate geometry. With a mere planar rectangle as geometry proxy, Light Field Rendering arguably constitutes the most "puristic" image-based rendering technique [LH96]. Here, many images are needed to avoid blurring or ghosting artefacts [CCST00]. If more appropriate depth maps are additionally available, Lumigraph rendering [GGSC96, BBM*01]. Other

Screen-space optical flow could fix some misregistration issues, but heavily relies on **RGB features**, and fails when **changing viewpoints**.

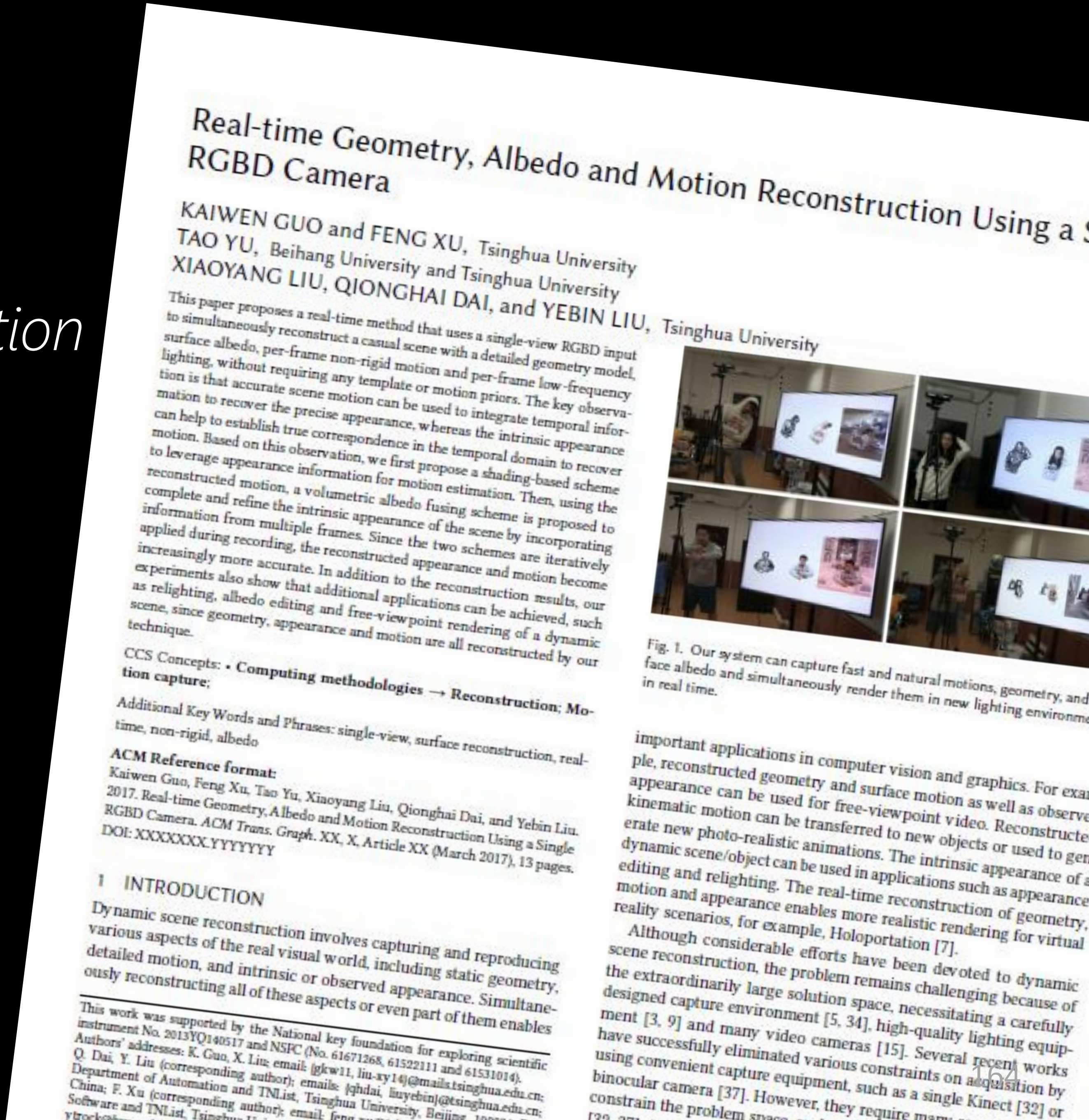
Related Work

3D Texture Montage

Up to now, few systems but *Holoportation* could fuse dynamic meshes **with multiple cameras** in real time.

This recent SIGGRAPH 2017 paper produces excellent dynamic reconstruction results,

but uses **a single RGBD camera**.



Real-time Geometry, Albedo and Motion Reconstruction Using a S
RGBD Camera

KAIWEN GUO and FENG XU, Tsinghua University
TAO YU, Beihang University and Tsinghua University
XIAOYANG LIU, QIONGHAI DAI, and YEBIN LIU, Tsinghua University

This paper proposes a real-time method that uses a single-view RGBD input to simultaneously reconstruct a casual scene with a detailed geometry model, surface albedo, per-frame non-rigid motion and per-frame low-frequency lighting, without requiring any template or motion priors. The key observation is that accurate scene motion can be used to integrate temporal information to recover the precise appearance, whereas the intrinsic appearance can help to establish true correspondence in the temporal domain to recover motion. Based on this observation, we first propose a shading-based scheme to leverage appearance information for motion estimation. Then, using the reconstructed motion, a volumetric albedo fusing scheme is proposed to complete and refine the intrinsic appearance of the scene by incorporating information from multiple frames. Since the two schemes are iteratively applied during recording, the reconstructed appearance and motion become increasingly more accurate. In addition to the reconstruction results, our experiments also show that additional applications can be achieved, such as relighting, albedo editing and free-viewpoint rendering of a dynamic scene, since geometry, appearance and motion are all reconstructed by our technique.

CCS Concepts: • Computing methodologies → Reconstruction; Motion capture;

Additional Key Words and Phrases: single-view, surface reconstruction, real-time, non-rigid, albedo

ACM Reference format:
Kaiwen Guo, Feng Xu, Tao Yu, Xiaoyang Liu, Qionghai Dai, and Yebin Liu. 2017. Real-time Geometry, Albedo and Motion Reconstruction Using a Single RGBD Camera. ACM Trans. Graph. XX, X, Article XX (March 2017), 13 pages.
DOI: XXXXXXX.YYYYYYY

Fig. 1. Our system can capture fast and natural motions, geometry, and surface albedo and simultaneously render them in new lighting environments in real time.

1 INTRODUCTION

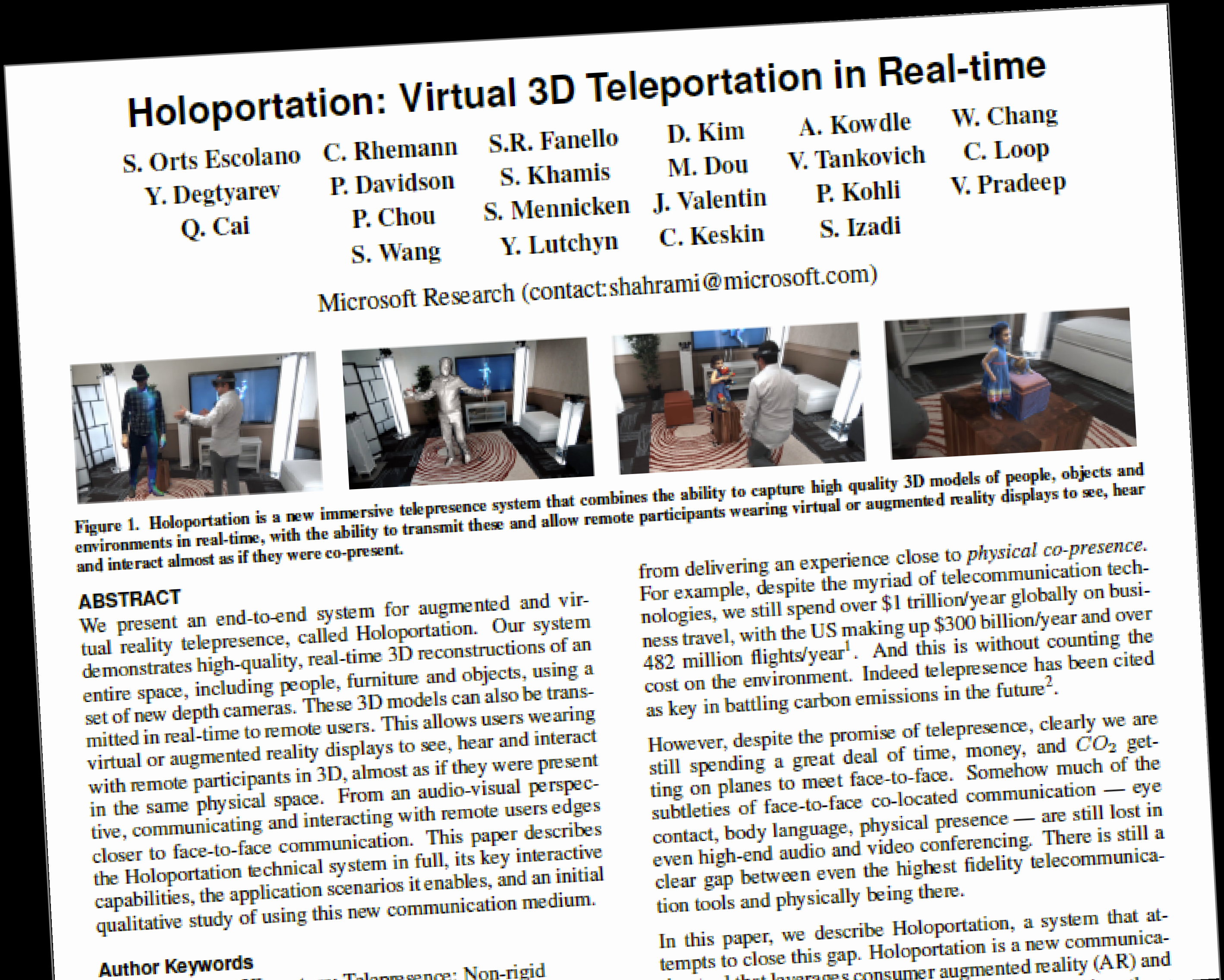
Dynamic scene reconstruction involves capturing and reproducing various aspects of the real visual world, including static geometry, detailed motion, and intrinsic or observed appearance. Simultaneously reconstructing all of these aspects or even part of them enables important applications in computer vision and graphics. For example, reconstructed geometry and surface motion as well as observed appearance can be used for free-viewpoint video. Reconstructed kinematic motion can be transferred to new objects or used to generate new photo-realistic animations. The intrinsic appearance of a dynamic scene/object can be used in applications such as appearance editing and relighting. The real-time reconstruction of geometry, motion and appearance enables more realistic rendering for virtual reality scenarios, for example, Holoportation [7].

Although considerable efforts have been devoted to dynamic scene reconstruction, the problem remains challenging because of the extraordinarily large solution space, necessitating a carefully designed capture environment [5, 34], high-quality lighting equipment [3, 9] and many video cameras [15]. Several recent works have successfully eliminated various constraints on acquisition by using convenient capture equipment, such as a single Kinect [32] or binocular camera [37]. However, they require many constrain the problem space

This work was supported by the National key foundation for exploring scientific instrument No. 2013YQ140517 and NSFC (No. 61671268, 61522111 and 61531014).
Authors' addresses: K. Guo, X. Liu; email: {gkw11, liu-xy14}@mails.tsinghua.edu.cn; Q. Dai, Y. Liu (corresponding author); emails: {qhdai, liuyebin}@tsinghua.edu.cn; Department of Automation and TNList, Tsinghua University, Beijing, China; F. Xu (corresponding author); email: feng... Software and TNList, Tsinghua U...

Related Work

3D Texture Montage



Holoportation: Virtual 3D Teleportation in Real-time

S. Orts Escolano, C. Rhemann, S.R. Fanello, D. Kim, A. Kowdle, W. Chang, Y. Degtyarev, P. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. Chou, S. Mennicken, J. Valentin, P. Kohli, V. Pradeep, S. Wang, Y. Lutchyn, C. Keskin, S. Izadi

Microsoft Research (contact:shahrami@microsoft.com)

Figure 1. Holoportation is a new immersive telepresence system that combines the ability to capture high quality 3D models of people, objects and environments in real-time, with the ability to transmit these and allow remote participants wearing virtual or augmented reality displays to see, hear and interact almost as if they were co-present.

ABSTRACT

We present an end-to-end system for augmented and virtual reality telepresence, called Holoportation. Our system demonstrates high-quality, real-time 3D reconstructions of an entire space, including people, furniture and objects, using a set of new depth cameras. These 3D models can also be transmitted in real-time to remote users. This allows users wearing virtual or augmented reality displays to see, hear and interact with remote participants in 3D, almost as if they were present in the same physical space. From an audio-visual perspective, communicating and interacting with remote users edges closer to face-to-face communication. This paper describes the Holoportation technical system in full, its key interactive capabilities, the application scenarios it enables, and an initial qualitative study of using this new communication medium.

Author Keywords

from delivering an experience close to *physical co-presence*. For example, despite the myriad of telecommunication technologies, we still spend over $1 trillion/year globally on business travel, with the US making up $300 billion/year and over 482 million flights/year[1]. And this is without counting the cost on the environment. Indeed telepresence has been cited as key in battling carbon emissions in the future[2].

However, despite the promise of telepresence, clearly we are still spending a great deal of time, money, and CO_2 getting on planes to meet face-to-face. Somehow much of the subtleties of face-to-face co-located communication — eye contact, body language, physical presence — are still lost in even high-end audio and video conferencing. There is still a clear gap between even the highest fidelity telecommunication tools and physically being there.

In this paper, we describe Holoportation, a system that attempts to close this gap. Holoportation is a new communica-

Normal Weighted Blending

$$w_i = V \cdot \max(0, \hat{n} \cdot \hat{v}_i)^{\alpha}$$

Visibility test

Normal vector

Texture camera view direction

Majority Voting for color correction

For each vertex, and for each texture, test **if the projected color agrees with more than half of the other textures**, if not, set the texture weight field to 0.

What is our approach for real-time **seamless** texture fusion?

Workflow

Identify and diffuse the seams

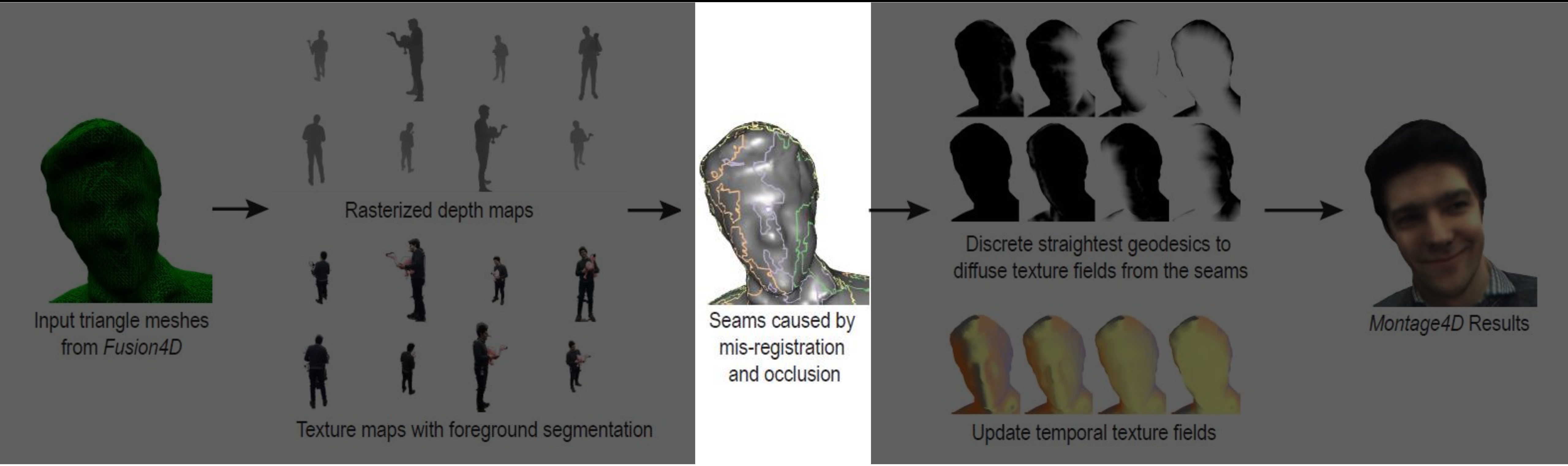
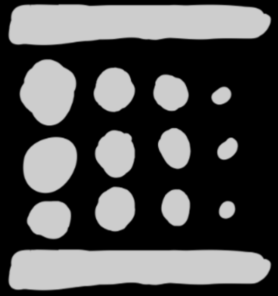


Figure 2: *The workflow of the Montage4D rendering pipeline.*

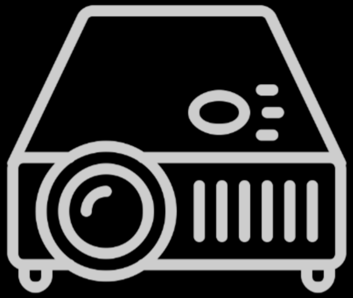
What are the **causes** for the *blurring* and *seams*?

Motivation

Causes for blurring



Blurring



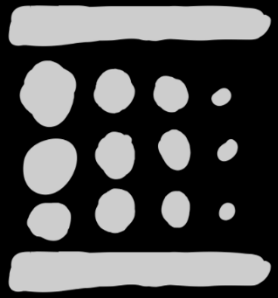
Texture projection errors

Inaccurate camera calibration

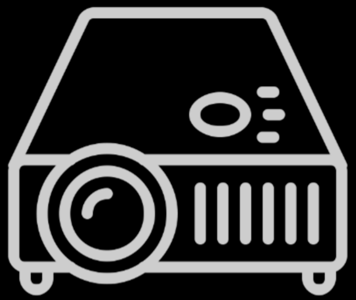
Normal-weighted blending

Motivation

Causes for blurring



Blurring



Texture projection errors

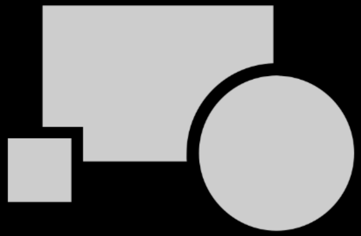
~~Normal-weighted blending~~

View-dependent rendering

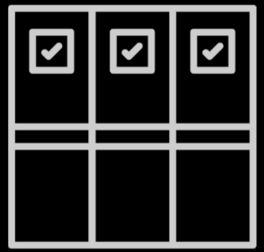
Motivation

Causes for Seams

Seams

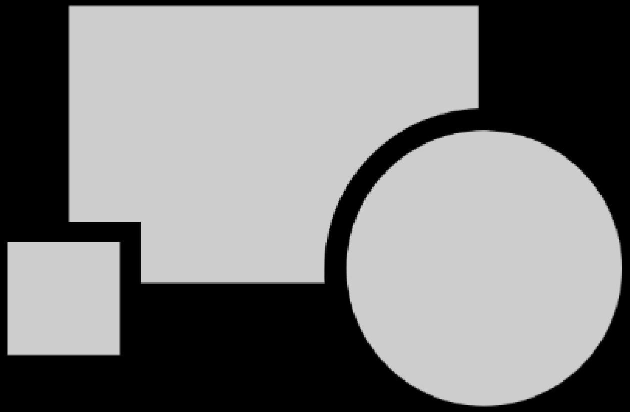


Self-occlusion



Majority-voting

Field-of-View



Self-occlusion

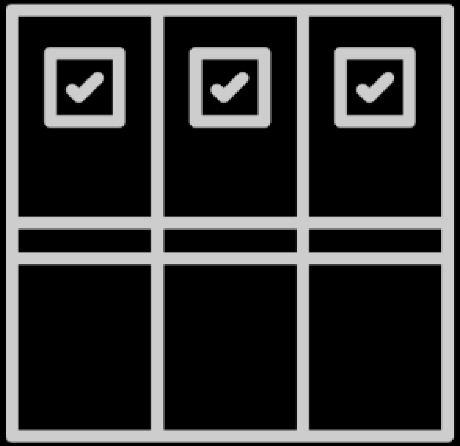
One or two vertices of the triangle are occluded in the depth map while the others are not.

Seams

Causes

Raw projection mapping results

Seams after occlusion test



Majority Voting

The triangle vertices have different results in the majority voting process, which may be caused by either mis-registration or self-occlusion.

Seams

Causes

Raw projection mapping results

Seams after occlusion test

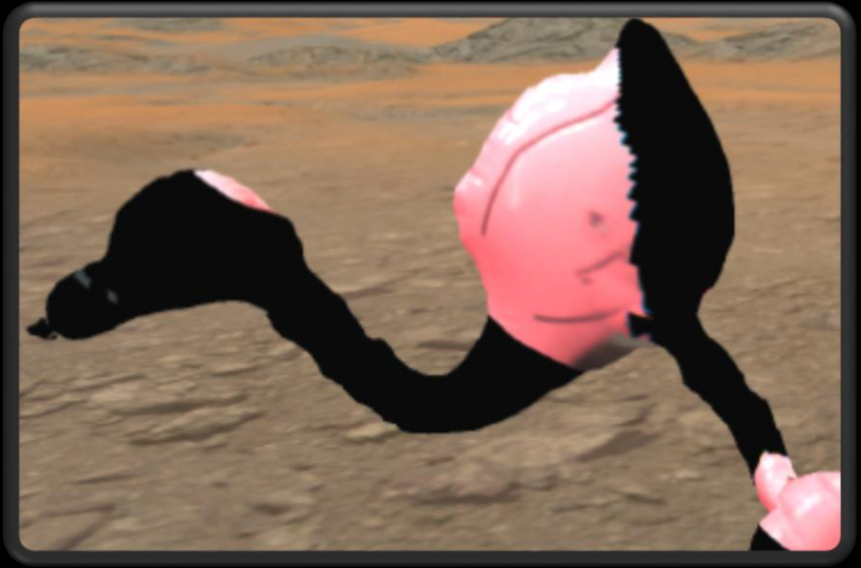
Seams after majority voting test

Field of View

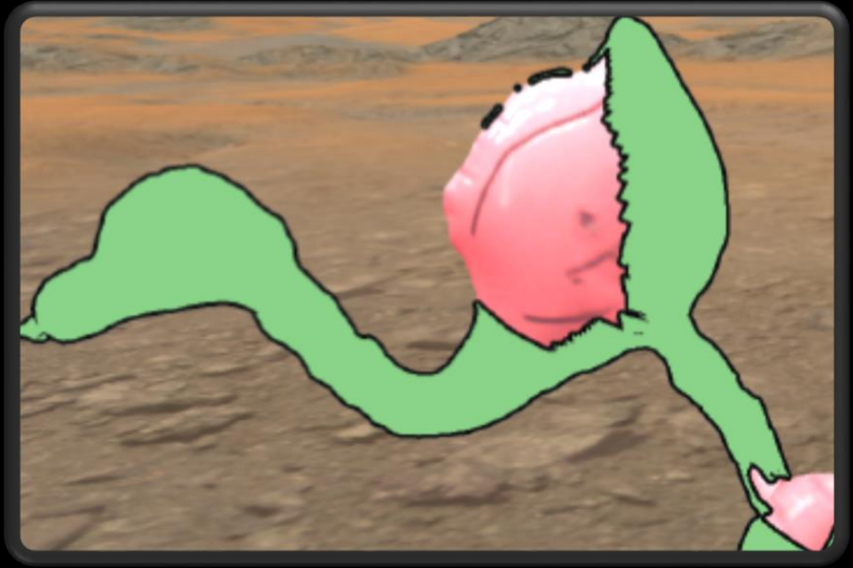
One or two triangle vertices lie outside the camera's field of view or in the subtracted background region while the rest are not.

Seams

Causes



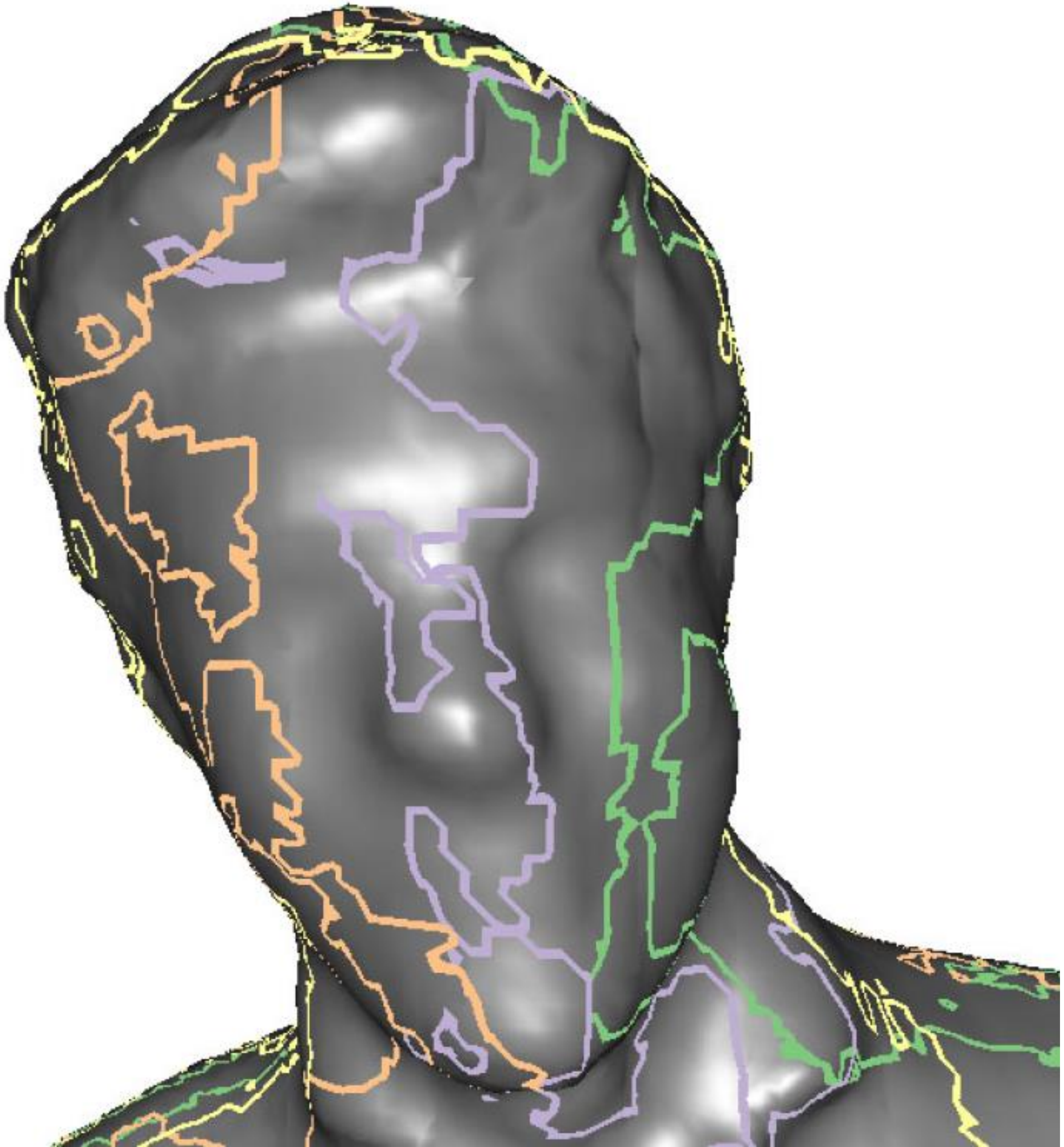
Raw projection mapping results



Seams after field-of-view test

Seams

Causes



For a static frame, how can we **get rid of** the annoying seams at interactive frame rate?

How can we spatially smooth the **texture (weight) field near the seams** so that we cannot see visible seams in the results?

Workflow

Identify and diffuse the seams

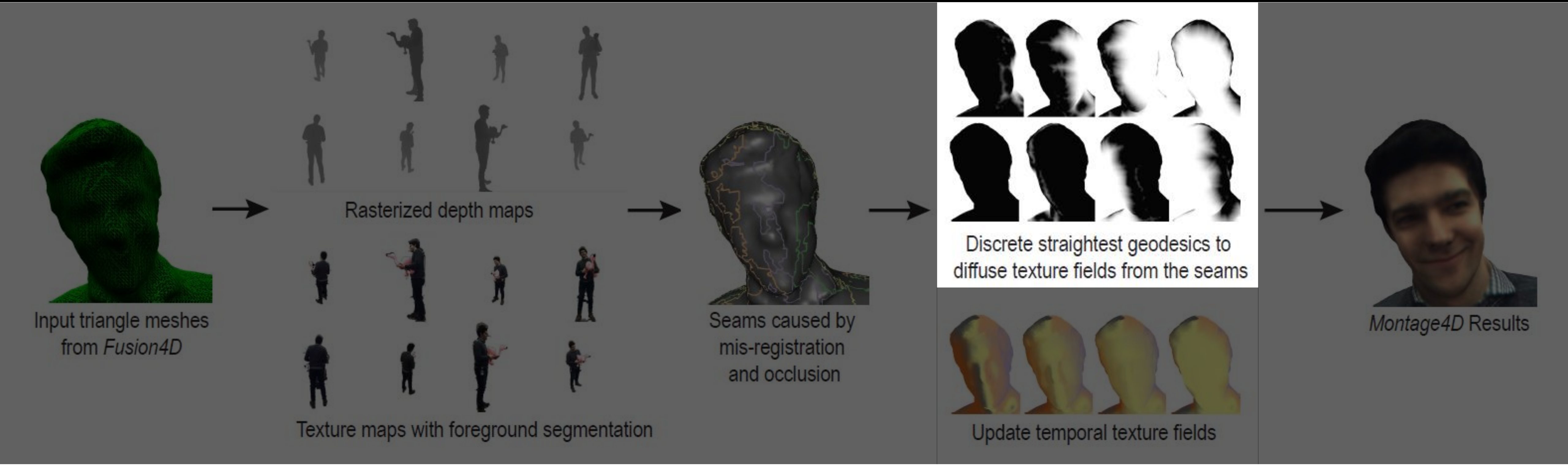


Figure 2: *The workflow of the Montage4D rendering pipeline.*

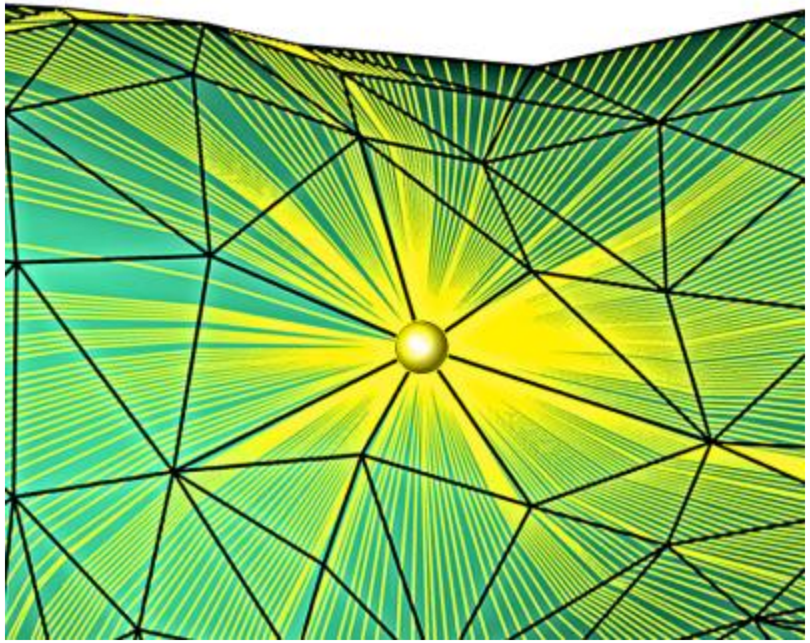
Geodesics

For diffusing the seams

Geodesic is the **shortest** route between two points on the surface.

Geodesics

For diffusing the seams



On triangle meshes, this is challenging because of the computation of **tangent directions**.
And shortest paths are defined on **edges** instead of the vertices.

Geodesics

For diffusing the seams

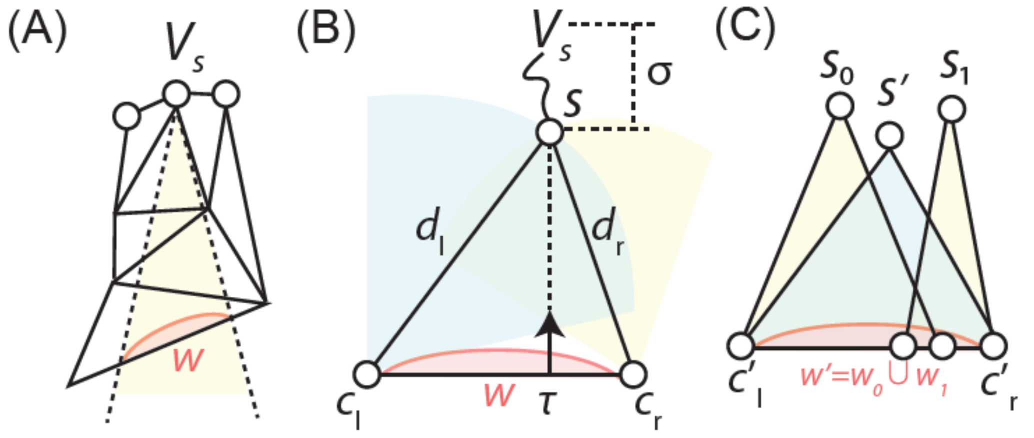


Figure 5: *Illustration of computing the approximate geodesics. (A) shows the concept of the geodesic window from a single source vertex. (B) shows the components within a window. (C) shows the merging process of two overlapping windows for approximiation.*

We use the algorithm by *Surazhsky* and *Hoppe* for computing **the approximate geodesics**.
The idea is to maintain **only 2~3 shortest paths** along each edge
to reduce the computational cost.

Approximate Geodesics

For diffusing the seams

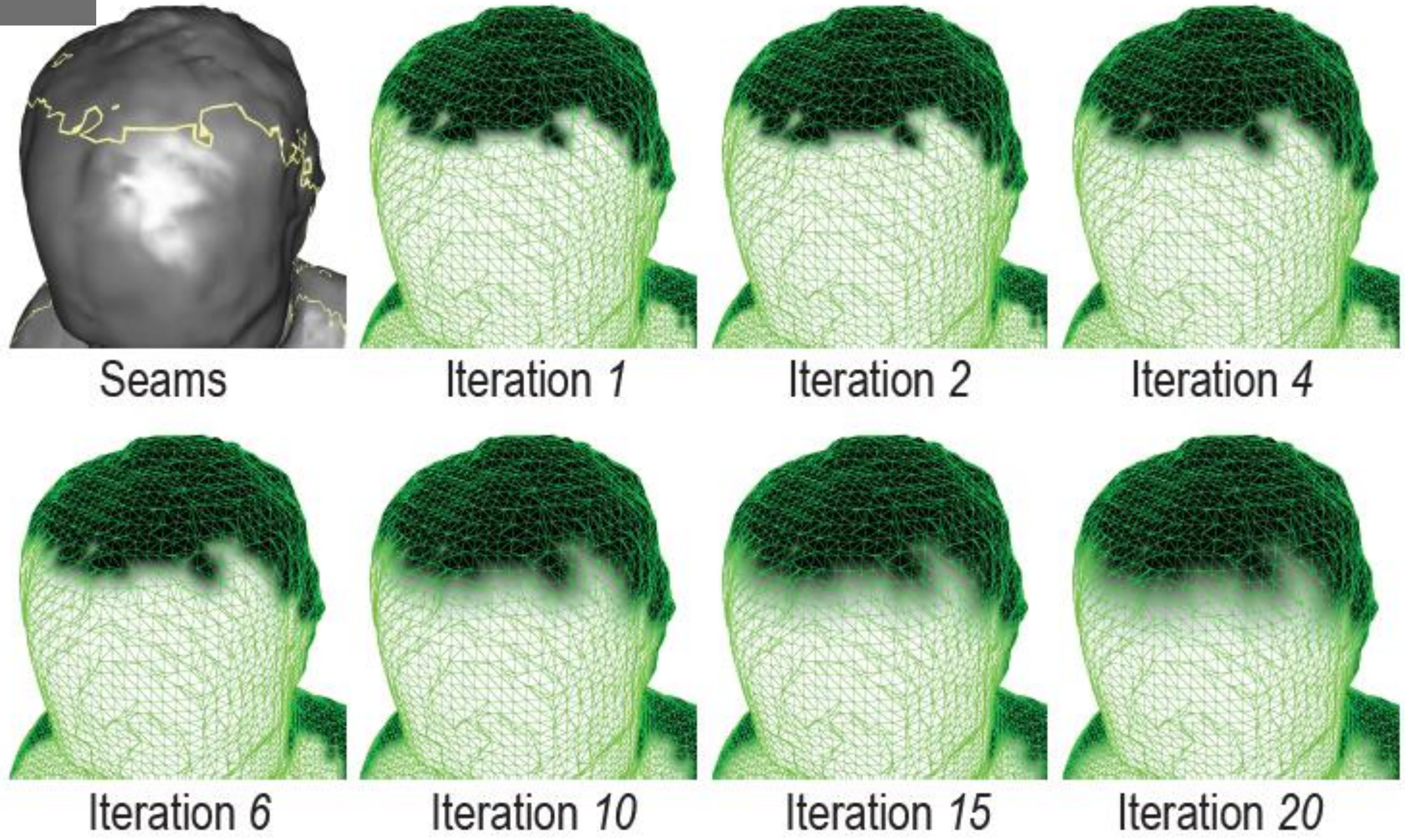


Figure 6: *Examples of the initial seam triangles and the propagation process for updating the geodesics.*

View-dependent Rendering

Spatially highlight the close views

Texture camera view direction

Global weight
Visibility% of view i

$$\mathscr{T}_{\mathbf{v}}^{i} = \mathscr{V}_{\mathbf{v}} \cdot g^{i} \cdot \gamma_{\mathbf{v}}^{i} \cdot \max(0, \hat{\mathbf{v}} \cdot \hat{\mathbf{v}}_{i})^{\alpha}$$

Visibility test

Geodesics

User camera's view direction

Workflow

Identify and diffuse the seams

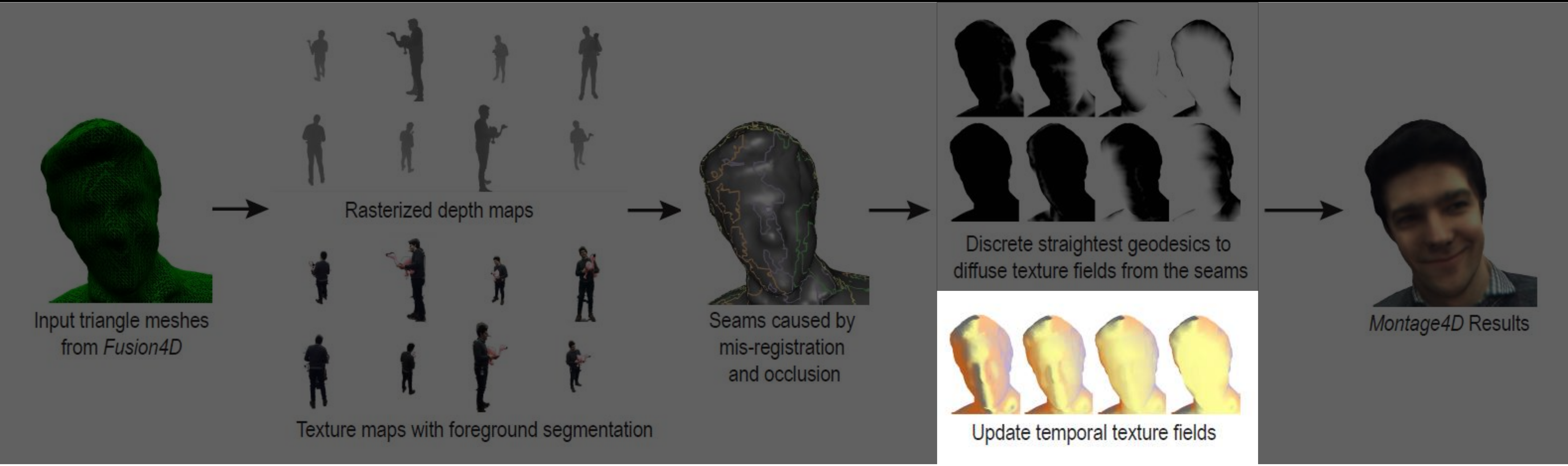


Figure 2: *The workflow of the Montage4D rendering pipeline.*

Temporal Texture Field

Temporally smooth the texture fields

Temporal smoothing factor of 0.02

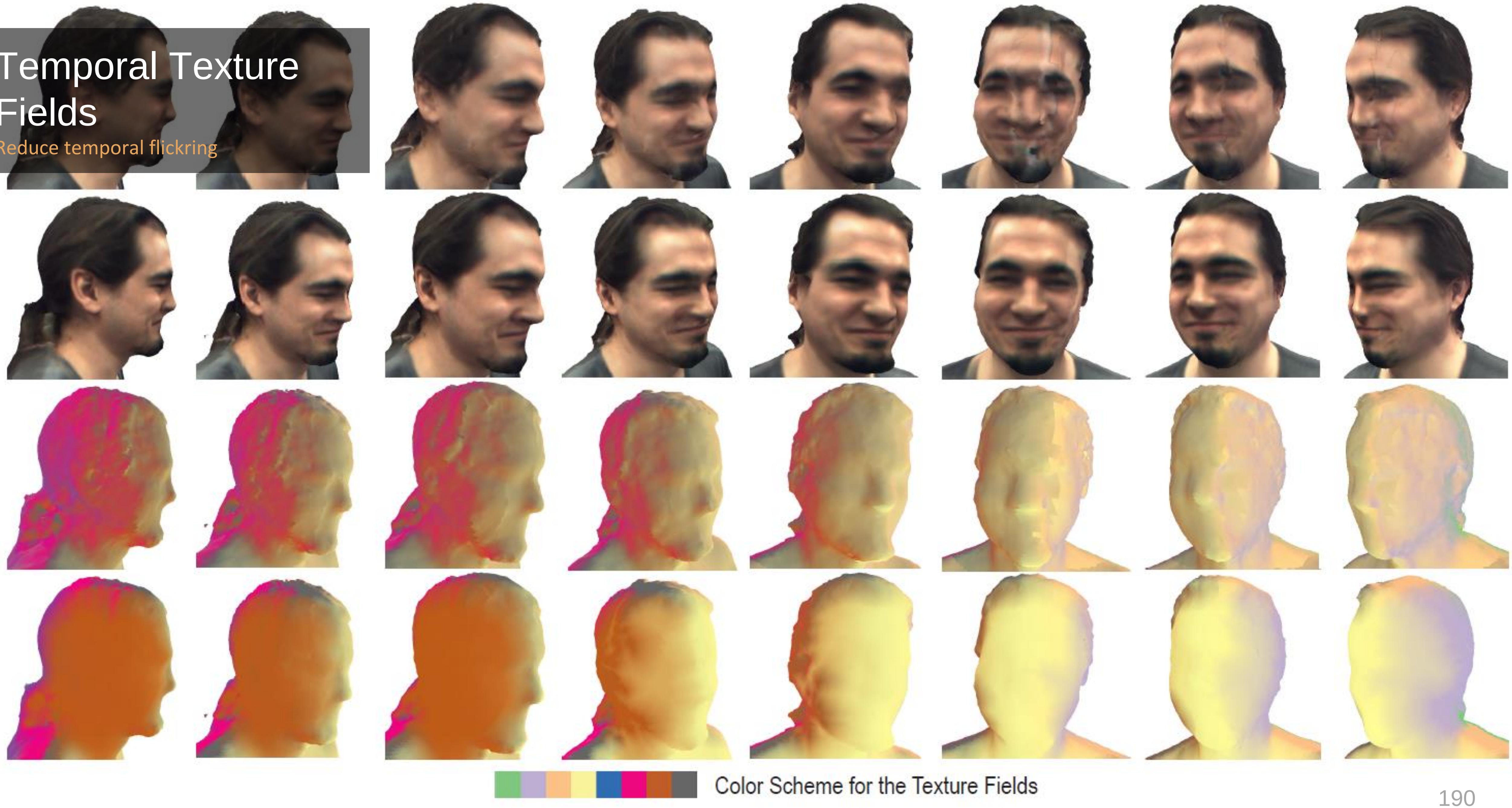
$$\mathscr{T}_{\mathbf{v}}^{i}(t) = \mathscr{T}_{\mathbf{v}}^{i}(t-1) + \lambda \nabla \mathscr{T}_{\mathbf{v}}^{i}(t)$$

Texture field of the previous frame

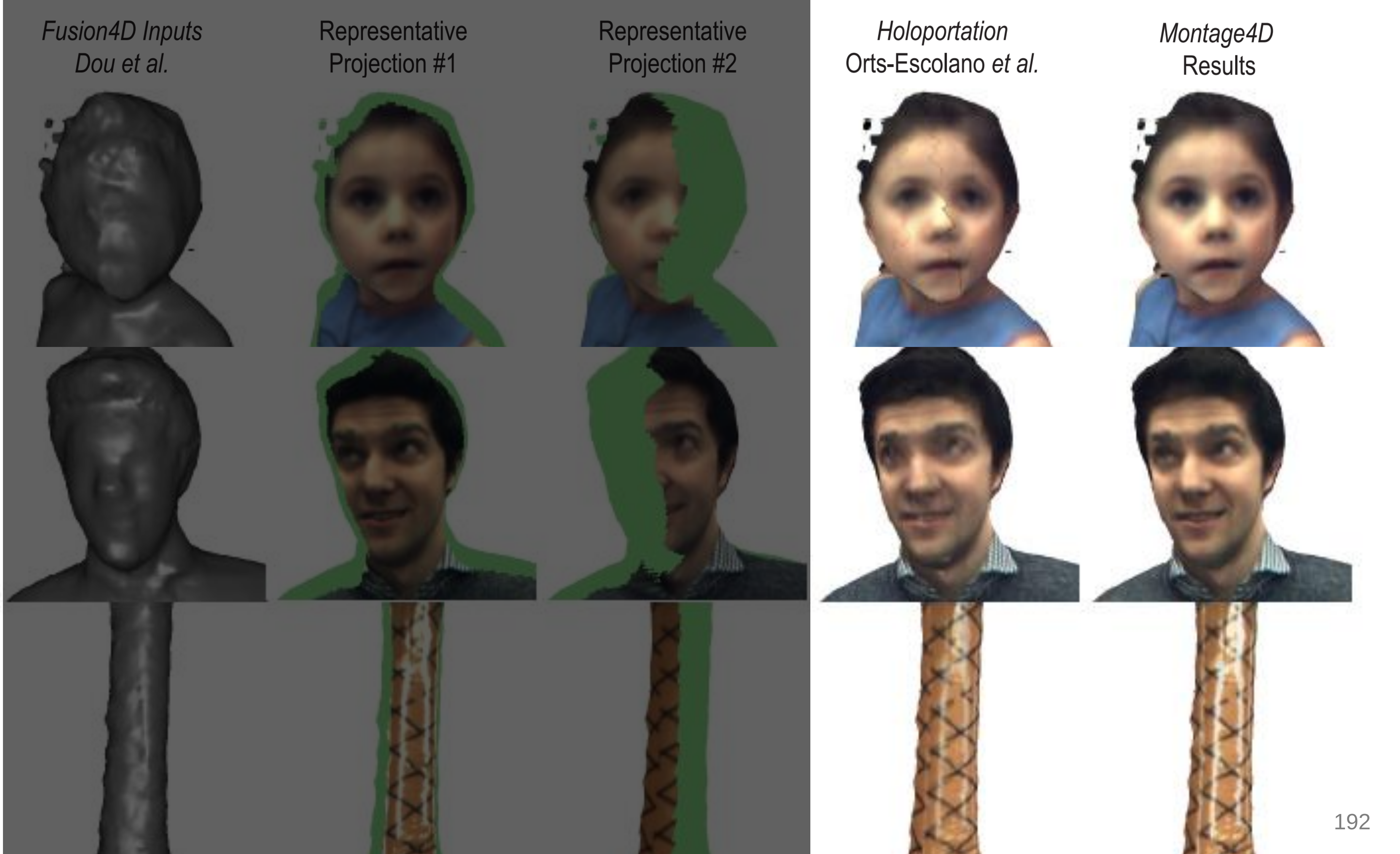
The gradient between the ideal texture field of
the *current* frame, and the value of the *previous* frame

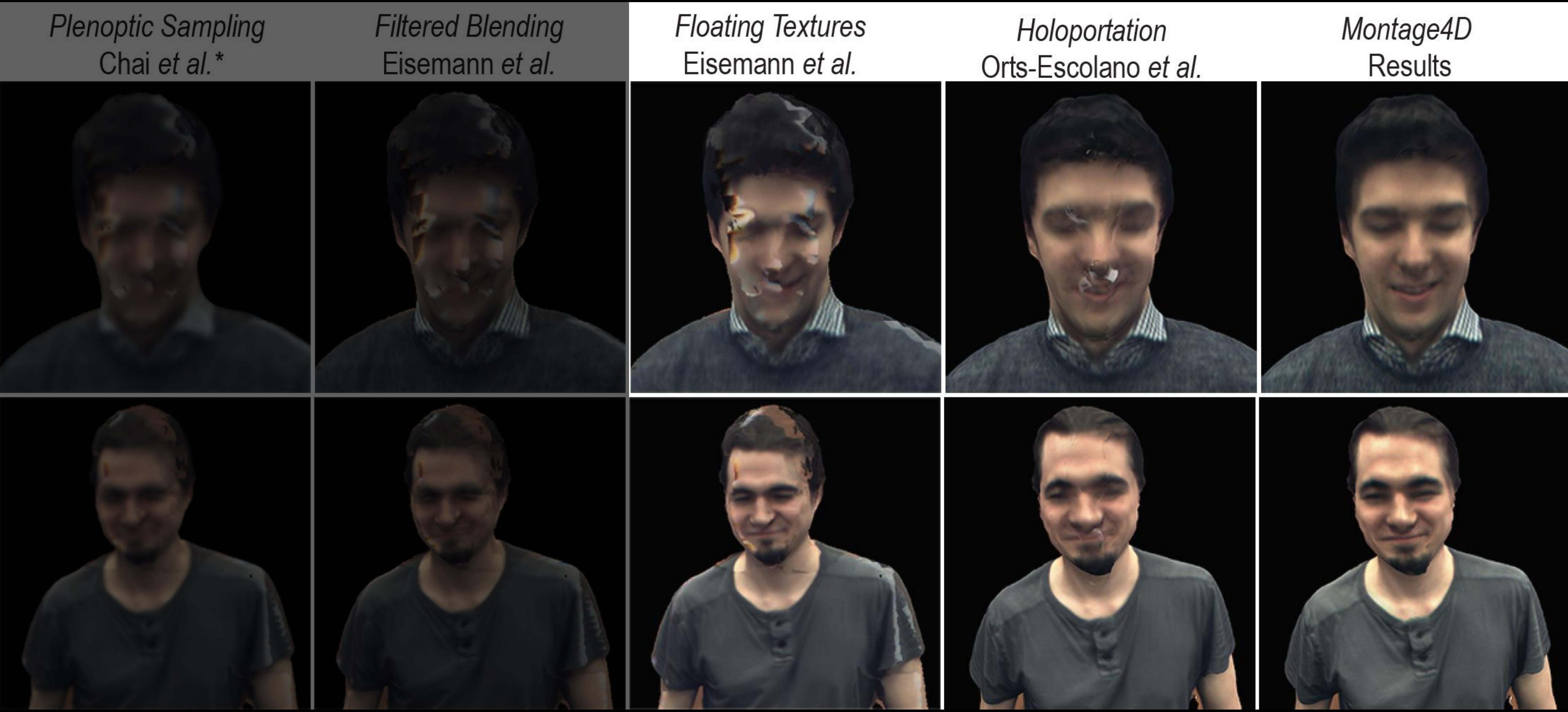
Temporal Texture Fields

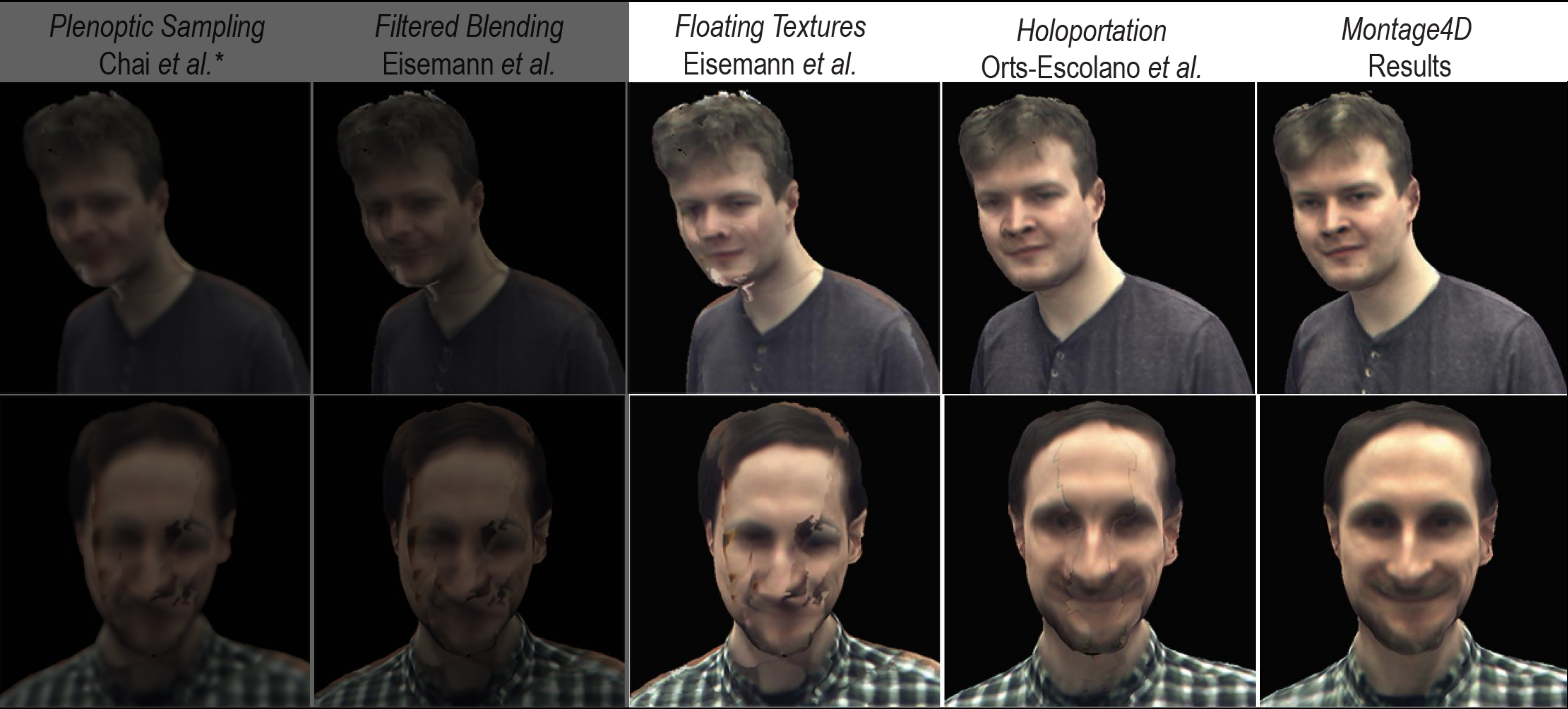
Reduce temporal flickring

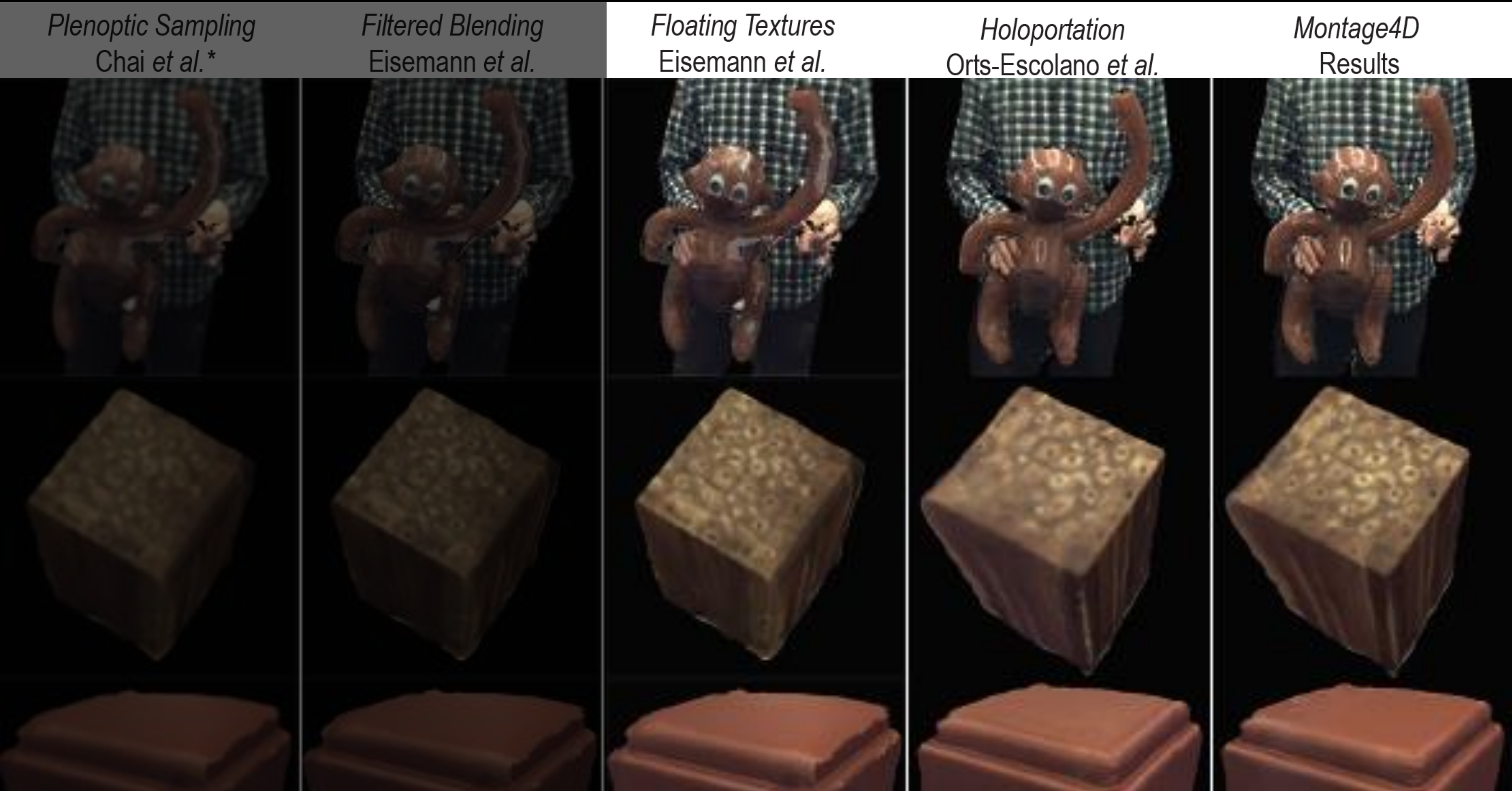



Fusion4D Inputs
Dou et al.
Representative
Projection #1
Representative
Projection #2
Holoportation
Orts-Escolano *et al.*
Montage4D
Results


Fusion4D Inputs *Dou et al.*
Representative Projection #1
Representative Projection #2
Holoportation Orts-Escolano *et al.*
Montage4D Results


Plenoptic Sampling
Chai *et al.**
Filtered Blending
Eisemann *et al.*
Floating Textures
Eisemann *et al.*
Holoportation
Orts-Escolano *et al.*
Montage4D
Results


Plenoptic Sampling
Chai et al.*
Filtered Blending
Eisemann et al.
Floating Textures
Eisemann et al.
Holoportation
Orts-Escolano et al.
Montage4D
Results


Plenoptic Sampling
Chai *et al.**
Filtered Blending
Eisemann *et al.*
Floating Textures
Eisemann *et al.*
Holoportation
Orts-Escolano *et al.*
Montage4D
Results

With additional computation for seams, geodesics, and temporal texture fields, is **our approach** still in real time?

Exmperiment

RMSE = Root mean squared error
FPS = frames per second

Table 1: *Comparison of root-mean-square error (RMSE) and frame rates between Holoportation and Montage4D methods*

Dataset	Frames	#vertices / frame	#triangles / frame	Holoporation RMSE	Holoporation FPS	Montage4D RMSE	Montage4D FPS
Timo	837	131K	251K	5.63%	227.2	3.27%	135.0
Yury	803	132K	312K	5.44%	222.8	3.01%	130.5
Sergio	837	215K	404K	7.74%	186.8	4.21%	114.3
Girl	1192	173K	367K	7.16%	212.56	3.73%	119.4
Julien	526	157K	339K	12.63%	215.18	6.71%	120.6

Exmperiment

RMSE = Root mean squared error
FPS = frames per second

Table 1: *Comparison of root-mean-square error (RMSE) and frame rates between Holoportation and Montage4D methods*

Dataset	Frames	#vertices / frame	#triangles / frame	Holoporation RMSE	[redacted]	Montage4D RMSE	[redacted]
Timo	837	131K	251K	5.63%	[redacted]	3.27%	[redacted]
Yury	803	132K	312K	5.44%	[redacted]	3.01%	[redacted]
Sergio	837	215K	404K	7.74%	[redacted]	4.21%	[redacted]
Girl	1192	173K	367K	7.16%	[redacted]	3.73%	[redacted]
Julien	526	157K	339K	12.63%	[redacted]	6.71%	[redacted]

The root-mean-squared-error of the rendering results have been great **reduced**.

Exmperiment

RMSE = Root mean squared error
FPS = frames per second

Table 1: *Comparison of root-mean-square error (RMSE) and frame rates between Holoportation and Montage4D methods*

Dataset	Frames	#vertices / frame	#triangles / frame	Holoporation		Montage4D	
				[redacted]	FPS	[redacted]	FPS
Timo	837	131K	251K	[redacted]	227.2	[redacted]	135.0
Yury	803	132K	312K	[redacted]	222.8	[redacted]	130.5
Sergio	837	215K	404K	[redacted]	186.8	[redacted]	114.3
Girl	1192	173K	367K	[redacted]	212.56	[redacted]	119.4
Julien	526	157K	339K	[redacted]	215.18	[redacted]	120.6

Our frame rate is slower, but still capable at **over 90 FPS**, for dynamic VR rendering.

Experiment

Break-down of a typical frame

Table 2: *Timing comparison between Holoportation and Montage4D for a typical frame*

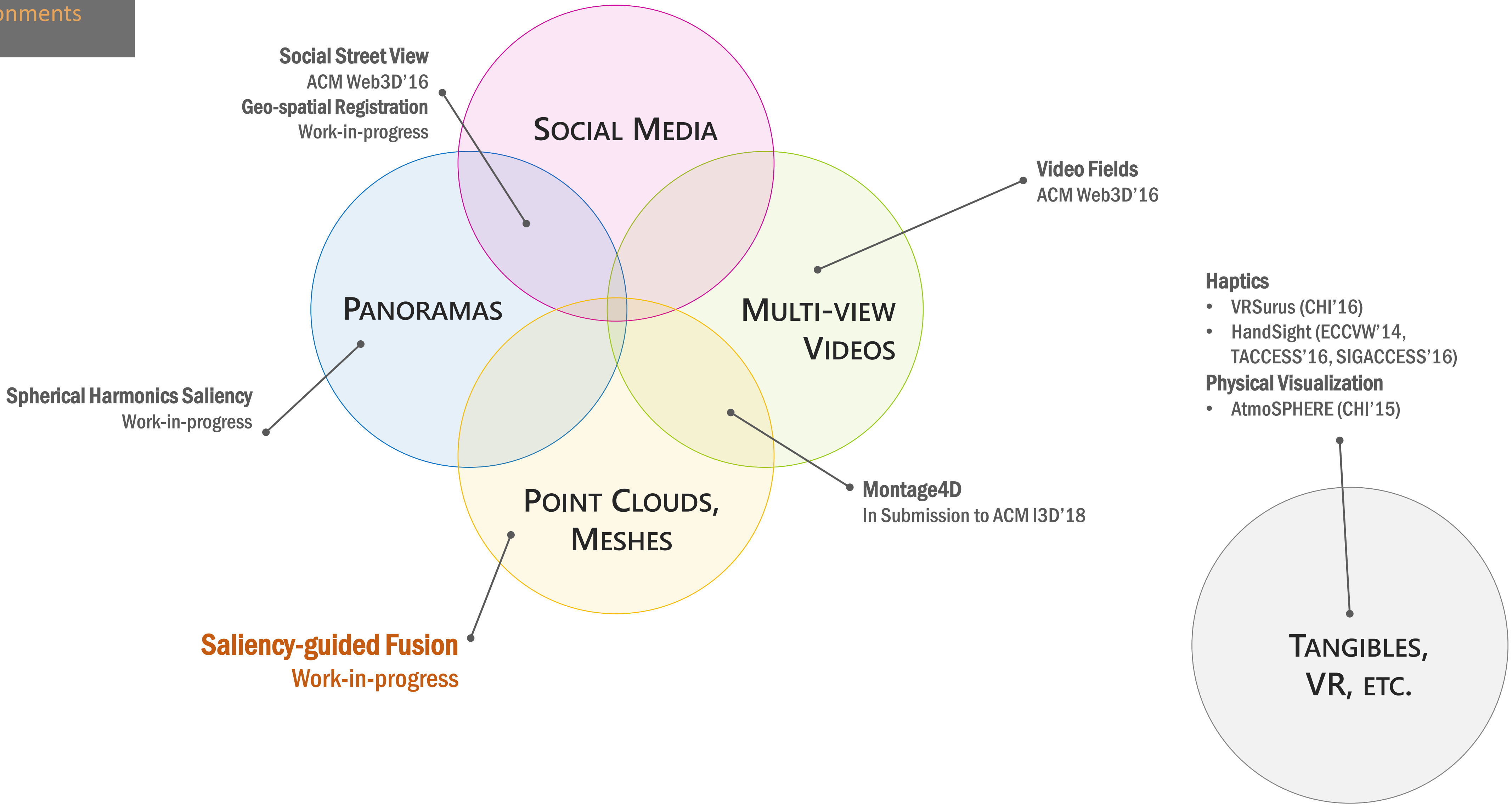
Procedure	Timing (ms)	
	Holoportation	Montage4D
Communication between CPU and GPU	4.83	9.49
Rendering and Texture Sampling	0.11	0.30
Rasterized Depth Maps calculation	0.14	0.13
Seams Identification	N/A	0.01
Approximate Geodesics estimation	N/A	0.31
Other events	0.12	0.18
Total	5.11	10.40

Most of the time is used in communication between CPU and GPU

In conclusion, **Montage4D** provides a practical texturing solution for real-time 3D reconstruction systems.
In the future, we envision that **Montage4D** is useful for **fusing the massive multi-view video data** into VR applications like *remote business meeting*, *immersive education*, and *family gathering*.

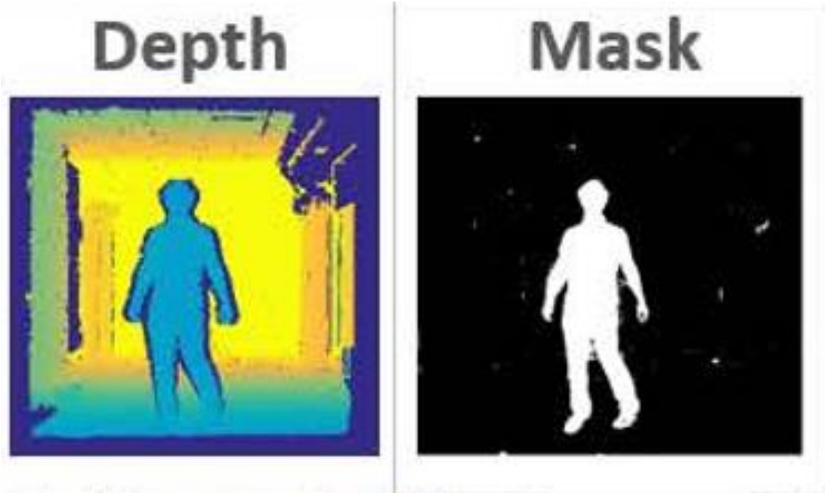
Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



Saliency-guided Fusion

Image Feature

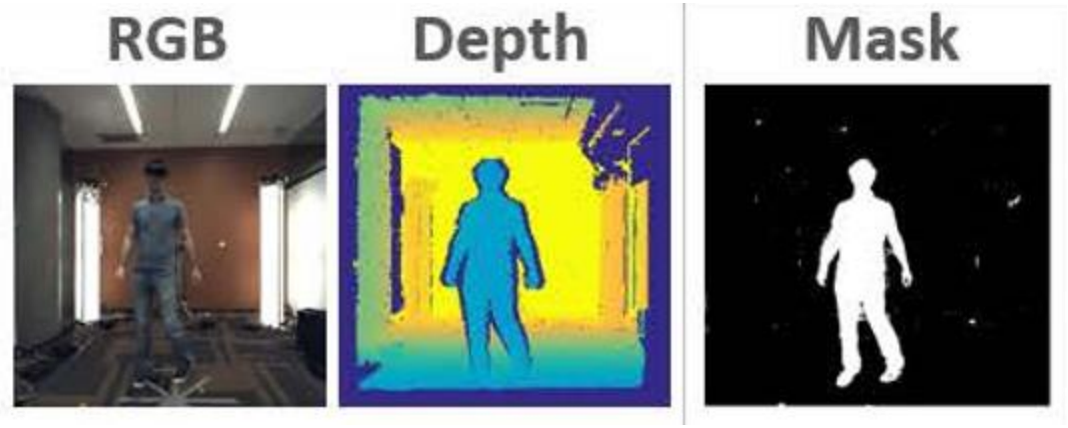


Depth Estimation and Segmentation

Volumetric Fusion

Saliency-guided Fusion

Image Feature



Full RGBD information and segmentation

Colored Volumetric Fusion

Colors Terms

Optimizing Energy Functions

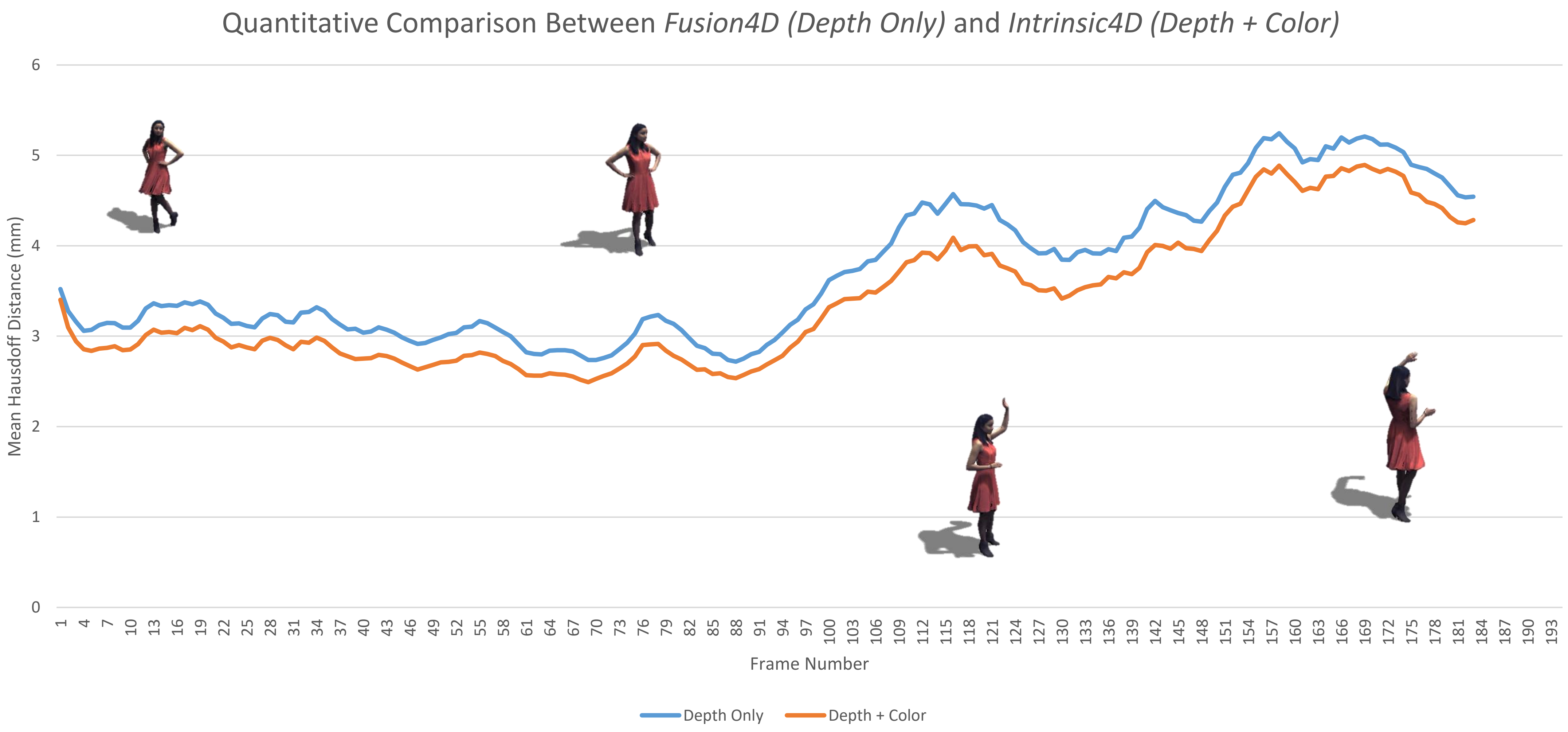
$$E(G) = \lambda_{depth} E_{depth}(G) + \boxed{\lambda_{color} E_{color}(G)} + \lambda_{corr} E_{corr}(G) + \lambda_{rot} E_{rot}(G) + \lambda_{smooth} E_{smooth}(G)$$

$$E_{depth}(G) = \sum_{n=1}^{N} \sum_{m \in \mathrm{V}_n(G)} \min_{x \in P(Depth_n)} \left|\left| Depth(v_m; G) - x \right|\right|^2$$

$$E_{color}(G) = \sum_{n=1}^{N} \sum_{m \in \mathrm{V}_n(G)} \min_{c \in P(Color_n)} \left|\left| Color(v_m; G)_{LAB} - c_{LAB} \right|\right|^2$$

Saliency-guided Fusion

Image Feature

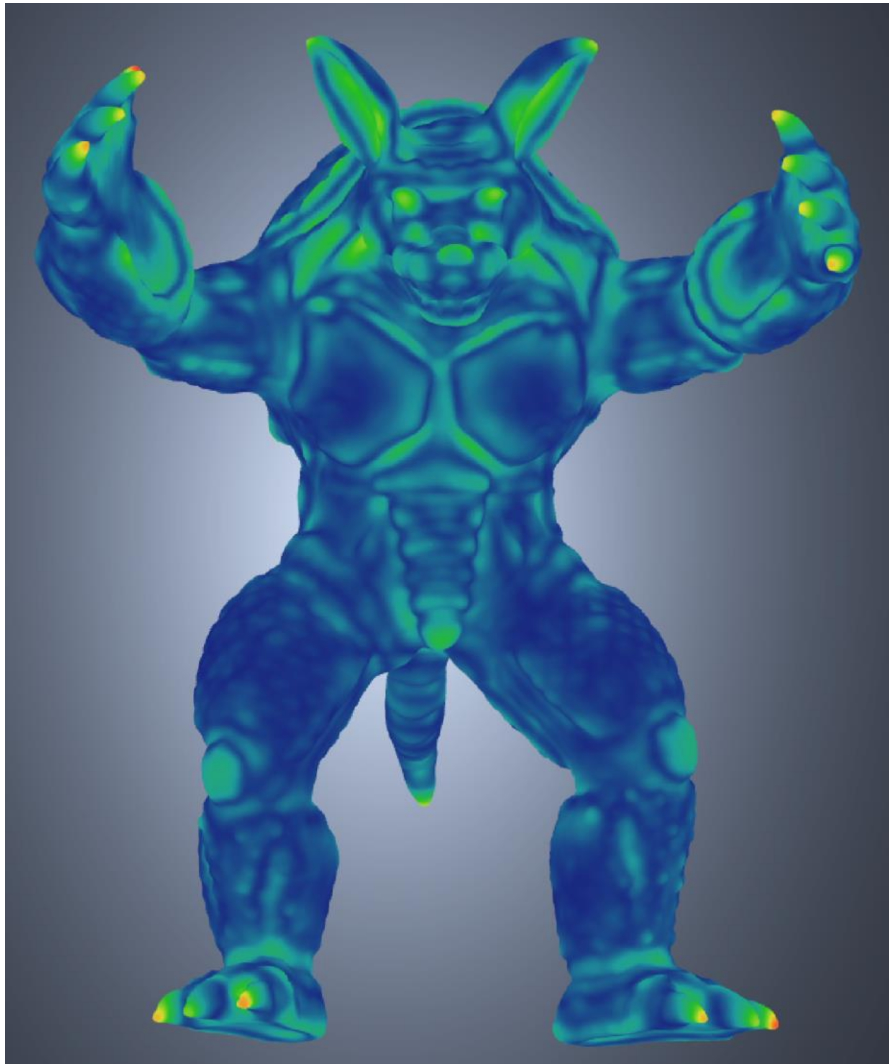


The frame rate **drops** from 30 FPS to 10 FPS with the color term…

How to **speed up** the dynamic reconstruction procedure,
with the performance improvement?

Mesh Saliency

Lee, Varshney, Jacobs. SIGGRAPH 2005

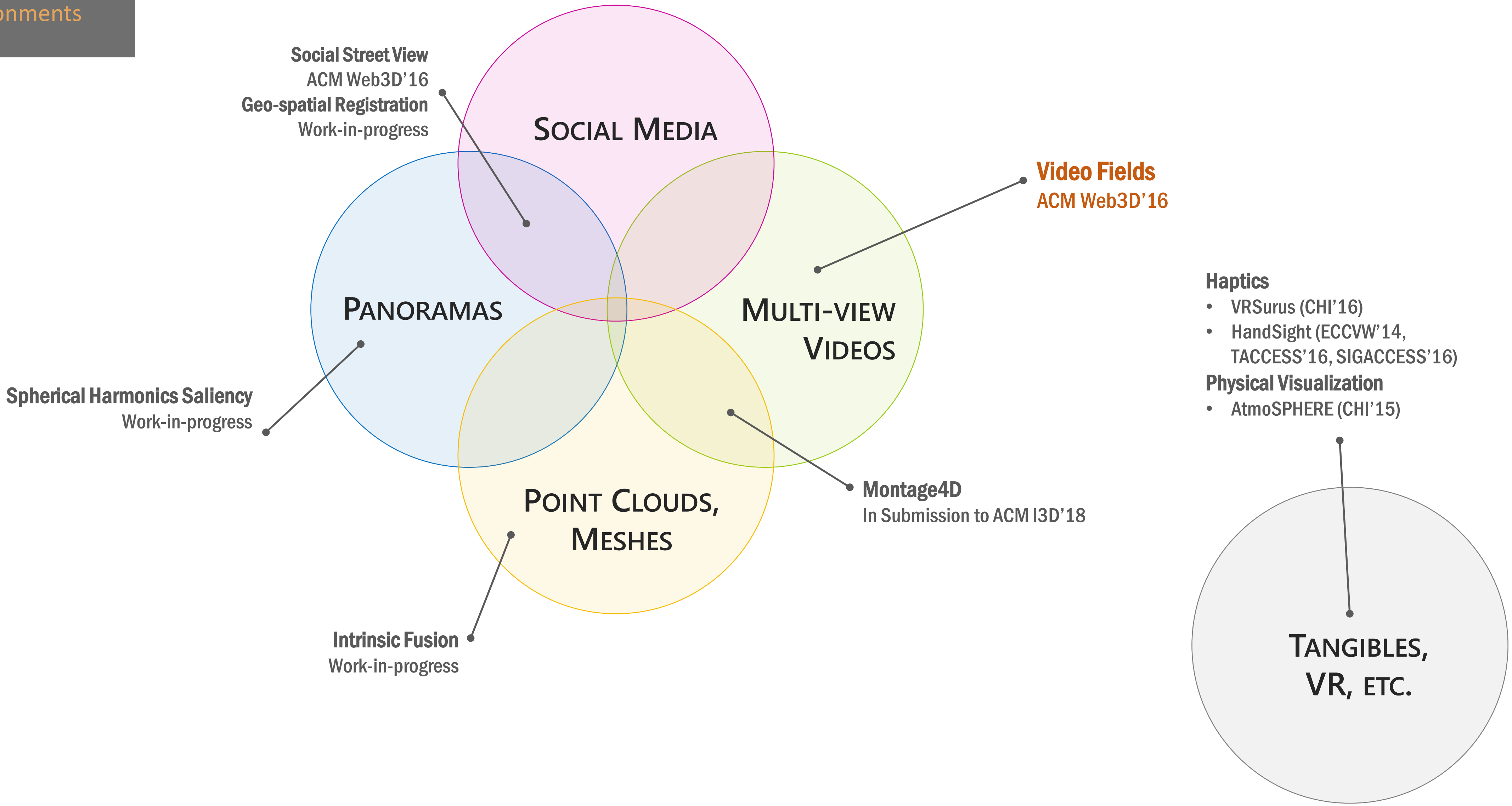


Use color-term optimization and more voxels for reconstructing **high-saliency** regions

Saliency is the vital key to balance the tradeoff between quality and speed.

Proposal

Fusing Multimedia Data Into
Dynamic Virtual Environments



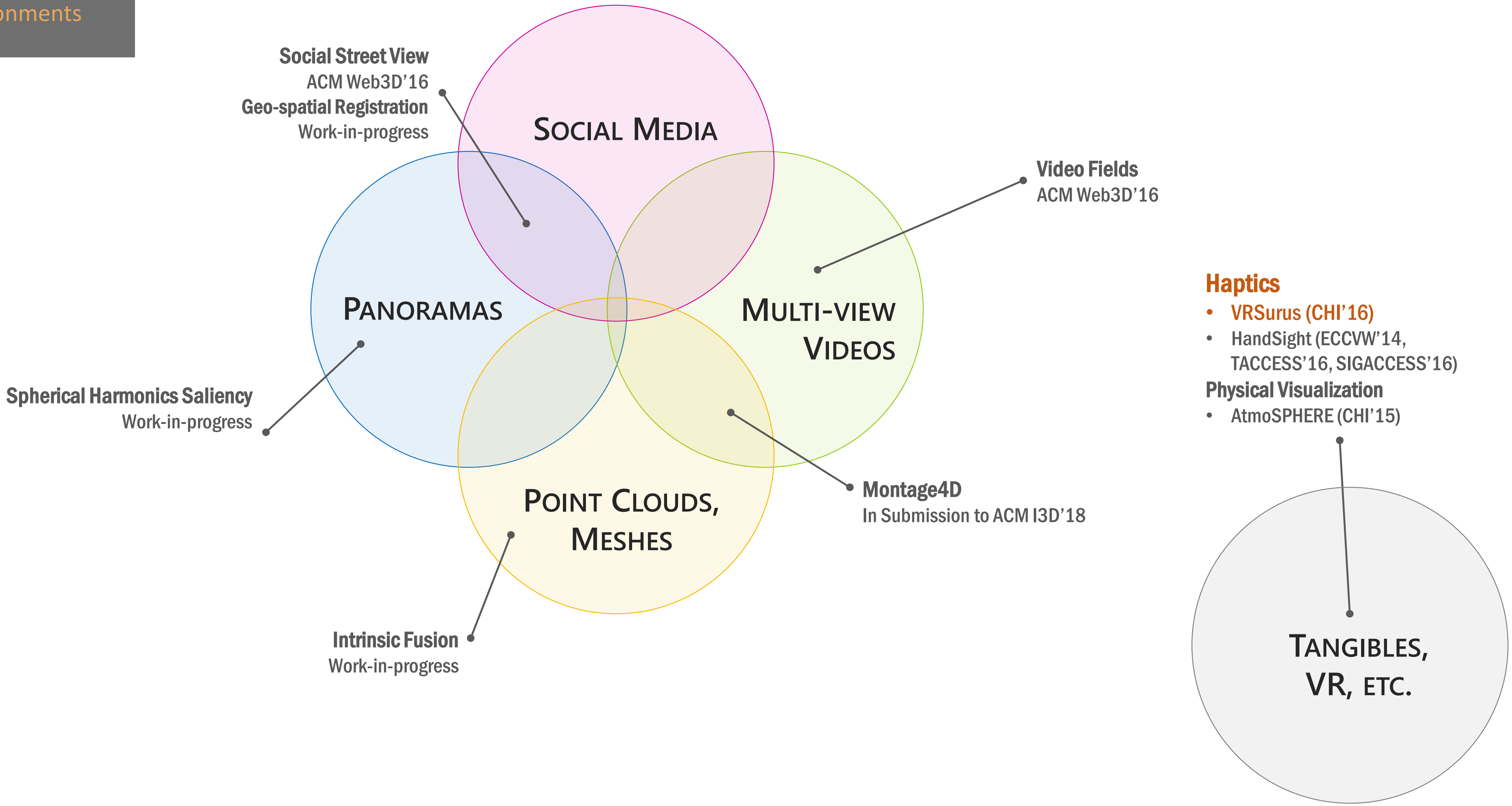
Video Fields: Fusing Multiple Surveillance Videos Into a Dynamic Virtual Environment

Ruofei Du, Sujal Bista, and Amitabh Varshney
www.Video-Fields.com
www.Augmentarium.com

Augmentarium | Department of Computer Science | UMIACS
University of Maryland, College Park
In Proceedings of the 21st Annual ACM SIGGRAPH Web3D Conference, 2016

Proposal

Fusing Multimedia Data Into Dynamic Virtual Environments



VRSurus: Enhancing Interactivity and Tangibility of Puppets in Virtual Reality (CHI 2016)

Ruofei Du and Liang He
University of Maryland, College Park

THE AUGMENTARIUM
VIRTUAL AND AUGMENTED REALITY LABORATORY
AT THE UNIVERSITY OF MARYLAND
GVIL
UMIACS
COMPUTER SCIENCE
UNIVERSITY OF MARYLAND
UNIVERSITY OF MARYLAND

Demo

VRSurus

Timeline

Paper deadline goals

Date	Project
Oct. – Nov. 2017	Spherical Harmonics Videos ACM I3D 2018
Nov. – Feb. 2018	Geo-spatial Registration with Social Street View ACM ToG / IEEE TVCG
Mar. – Aug. 2018	Saliency-guided Real-time 3D Reconstruction ACM ToG / IEEE TVCG

Paper List
Preliminary Work

1. **Du, R.**, Varshney, A. Social Street View: Blending Immersive Street Views with Geo-tagged Social Media.
   In proceedings of the 21st International Conference on Web3D, 2016. pp. 77-85. ACM. **(Best Paper Award)**
2. **Du, R**., Bista, S., Varshney, A. Video Fields: Fusing Multiple Surveillance Videos into a Dynamic Virtual Environment.
   In proceedings of the 21st International Conference on Web3D, 2016. pp. 165-172. ACM.
3. **Du, R**., Ming, C., Chang, W., Hoppe, H., Varshney, A. Montage4D: Interactive Seamless Fusing Multiview Video Textures.
   In submission to ACM SIGGRAPH Symposium on Interactive Graphics (ACM I3D) 2018.
4. **Du, R**., He, L. VRSurus: Enhancing Interactivity and Tangibility of Puppets in Virtual Reality.
   In Proceeding of the of Human Factors in Computing Systems. (CHI '16 EA) pp. 2454-2461. ACM
5. **Du, R**., Wills, K., Potasznik, M, Froehlich, J.E. AtmoSPHERE: Representing Space and Movement Using Sand Traces in an Interactive Zen Garden. In Proceeding of the of Human Factors in Computing Systems (CHI '16 EA). pp. 1627-1632. ACM.
6. Stearns, L., **Du, R.**, Oh, U., Catherine, Z., Findlater, L., David, R., Froehlich, J.E. Evaluating Haptic and Auditory Directional Guidance to Assist Blind Persons in Reading Printed Text Using Finger-Mounted Cameras.
   In ACM Transactions on Accessible Computing, 8(5), pp. 1-39. 2016.
7. Stearns, L., **Du, R.**, Oh, U., Wang, Y., Findlater, L., Chellappa, R., Froehlich, J.E. The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind.
   In Proceeding of the European Conference on Computer Vision (ECCV '14 Workshop). pp. 615–631. 2014.
8. Findlater, L., Stearns, L., **Du, R.**, Oh, U., Wang, Y., Chellappa, R., Froehlich, J.E.
   Supporting Everyday Activities for Persons With Visual Impairments Through Computer Vision-Augmented Touch
   In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, pp. 383-384, 2015.

Acknowledgement

Collaborators, advisors, labmates, and committee members

Committee | Augmentarium | MSR | Collaborators

Amitabh Varshney
varshney@cs.umd.edu

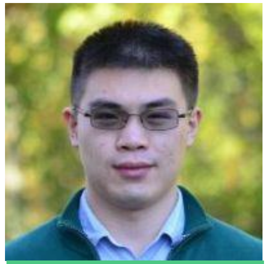
Matthias Zwicker
zwicker@cs.umd.edu

Furong Huang
furongh@cs.umd.edu

Sujal Bista
Sujal@cs.umd.edu

Hsueh-Chien Cheng
hccheng@cs.umd.edu

Eric Krokos
Sujal@cs.umd.edu



Xuetong Sun
xtsunl@cs.umd.edu

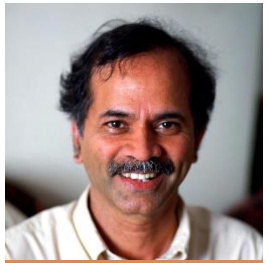
Xiaoxu Meng
xmeng525@umiacs.umd.edu

Sida Li
sidali@umiacs.umd.edu

Eric Lee
ericlee@umiacs.umd.edu

Jon Froehlich
jonf@cs.washington.edu

Leah Findlater
leahkf@uw.edu



Rama Chellappa
rama@umiacs.umd.edu

Lee Stearns
lee@leestearns.com

Liang He
edigahe@gmail.com

Sai Yuan
syuan@umd.edu

Wayne Chang
wechang@microsoft.com

Marek Kowalski
marek.kol4@gmail.com

Zoey Chen
qiuyuchen14@gmail.com

Spencer Fowers
sfowers@microsoft.com

Jeff Kramer
jekramer@Microsoft.com

Ben Cutler
bcutler@microsoft.com

Ming Chuang
mingchuang82@gmail.com

Hugues Hoppe
hhoppe@gmail.com

Acknowledgement

NSF | Nvidia | MPower | UMIACS

NSF

NVIDIA®

UMIACS

University of Maryland Institute for Advanced Computer Studies

Microsoft® Research

UNIVERSITY OF MARYLAND

UNIVERSITY OF MARYLAND

MPOWERING THE STATE

Thank you

Ruofei Du

ruofei@cs.umd.edu

Committee: Dr. Varshney, Dr. Zwicker, and Dr. Huang

Fusing Multimedia Data Into Dynamic Virtual Environments

Ruofei Du
ruofei@cs.umd.edu

Committee: Dr. Varshney, Dr. Zwicker, and Dr. Huang