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

Ruofei Du and Amitabh Varshney*
Augmentarium, Department of Computer Science and the Institute for Advanced Computer Studies (UMIACS)
University of Maryland, College Park

Figure 1: *Results from our system, Social Street View. (a) shows the rendering results in a regular display. Users can look through the museum in Paris for geo-tagged artworks inside as well as the dishes in nearby restaurants. (b) shows the stereo rendering results in a VR headset. Geo-tagged images are automatically aligned with building geometry and laid out aesthetically. (c) shows the deployment in an immersive curved screen environment with 15 projectors. Users can explore hundreds of social media messages near a New York city street at a resolution of 6k × 3k pixels. Please refer to the supplementary video at http://augmentarium.umd.edu and http://socialstreetview.com.*

Abstract

This paper presents an immersive geo-spatial social media system for virtual and augmented reality environments. With the rapid growth of photo-sharing social media sites such as Flickr, Pinterest, and Instagram, geo-tagged photographs are now ubiquitous. However, the current systems for their navigation are unsatisfyingly one- or two-dimensional. In this paper, we present our prototype system, Social Street View, which renders the geo-tagged social media in its natural geo-spatial context provided by immersive maps, such as Google Street View. This paper presents new algorithms for fusing and laying out the social media in an aesthetically pleasing manner with geospatial renderings, validates them with respect to visual saliency metrics, suggests spatio-temporal filters, and presents a system architecture that is able to stream geo-tagged social media and render it across a range of display platforms spanning tablets, desktops, head-mounted displays, and large-area room-sized curved tiled displays. The paper concludes by exploring several potential use cases including immersive social storytelling, learning about culture and crowd-sourced tourism.

Keywords: spatial-temporal virtual reality; social media; street view; geographical information systems; mixed reality; WebGL

Concepts: •**Computing methodologies** → **Virtual reality;** •**Information systems** → Geographic information systems;

*emails: {ruofei, varshney}@cs.umd.edu

Web3D '16,, July 22 - 24, 2016, Anaheim , CA, USA
ISBN: 978-1-4503-4428-9/16/07 $15.00
DOI: http://dx.doi.org/10.1145/2945292.2945294

1 Introduction

Social media plays a vital role in our lives because of its interactivity, versatility, popularity, and social relevance. Every day, billions of users create, share, and exchange information from their lives among their social circles [Kaplan and Haenlein 2010]. Social media spans several modalities that include text, photos, audio, videos, and even 3D models. In addition to the content visible on the social networks, social media also consists of metadata that is useful for understanding the relationship amongst users, sentiment mining, and propagation of influence. Specifically, digital photographs often embed metadata such as time of creation, location of creation (through GPS coordinates), and camera parameters, that are included in the EXIF data fields. In spite of the availability of such rich spatio-temporal metadata, the current-generation social media content is most often visualized as a linear narrative, rarely in a 2D layout, and almost never in a natural immersive space-time setting. For small screens of mobile devices, perhaps a linear narrative ordered by time or relevance, makes the most sense given the limitations of interaction modalities and display real-estate. However, in immersive virtual environments such as those afforded by virtual and augmented reality headsets, a spatio-temporal view of the social media in a mixed reality setting may be the most natural.

Immersive interfaces that interleave visual navigation of our surroundings with social media content have not yet been designed. *NewsStand* [Teitler et al. 2008], *Flickr* [Serdyukov et al. 2009], and *Panoramio*[1] have taken important first steps towards this goal by using a user's geo-location information on 2D maps to display content. Still, we have not come across any system that (a) enables

[1]Panoramio:http://www.panoramio.com

user exploration of social media in an immersive 3D spatial context in real time, and (b) allows temporal filtering of social messages in their spatial setting. Such a system will facilitate new genres of social interactions in spatial contexts mediated through virtual and augmented reality. These could be widely adopted in immersive social storytelling, learning about culture, and crowd-sourced tourism.

As a proof-of-concept, we have developed a prototype system called the *Social Street View* (SSV) (Fig. 1), the first immersive social media navigation system for exploring social media in large-scale urbanscapes and landscapes. Given a requested location, SSV builds a 3D world using tiles of panorama data from *Google Street View*[2] and *Bing Maps Streetside*[3], depths and normal maps, and road orientations. It then downloads geo-tagged data near the requested location from public-domain social media sites such as *Instagram* and *Twitter*. After building the 3D world, SSV renders the social media onto buildings or as virtual billboards along the road. The user can see photos of food uploaded by social media users next to the relevant restaurants, visual memories with friends in specific locations, identify accessibility issues on roads, as well as preview the coming attractions on scenic drives and nature hikes. The main contributions of our work are:

- conception, architecting, and implementation of Social Street View, a mixed reality system that can depict geo-tagged social media in an immersive 3D environment,
- blending multiple modalities of panoramic view metadata, including depth maps, normal maps, and road orientation, with social media meta data including GPS coordinates and time of creation,
- enhancing visual augmentation by using maximal Poisson-disk sampling and image saliency metrics,
- using WebGL and WebVR to achieve cross-platform compatibility across a range of clients including smartphones, tablets, desktop, high-resolution large-area wide-field-of-view tiled display walls, as well as head-mounted displays.

2 Background and Related Work

Our work builds upon a rich literature of prior art on creation of immersive maps as well as related work in visual management of geo-tagged information, analysis of geo-tagged social media and mixed reality in immersive maps.

2.1 Immersive Maps

In this paper we refer the term **immersive maps** to refer to online services that provide panoramic 360° bubbles at multiple waypoints. Since Google Street View (GSV) debuted in 2007, several immersive maps have covered over 43 countries throughout the world. Most street views are captured using a car equipped with a spherical array of cameras. For places inaccessible to ordinary cars, volunteers and trekkers on foot, tricycle, trolley, camel, snow mobile and even underwater apparatus capture immersive panoramas [Anguelov et al. 2010]. Therefore, the latest immersive maps include not only outdoor urbanscapes, but also indoor areas, rural areas, forests, deserts, and even under-water seascapes. Images from a spherical array of cameras can recover depth and reconstruct 3D point clouds using structure-from-motion algorithms [Torii et al. 2009]. Recently, laser scanners are being coupled with the cameras to directly acquire depth with omnidirectional panoramas.

[2]Google Street View: http://www.google.com/maps/streetview

[3]Bing Maps Streetside: http://www.bing.com/mapspreview

2.2 Visual Management of Geo-tagged Information

As a geographic information system, Social Street View is most closely related to Panoramio, Newsstand [Teitler et al. 2008], and PhotoStand [Samet et al. 2013]. These systems accomplish visual management of geo-tagged information on 2D maps. Panoramio is one of the first systems that collects user-submitted scenery photographs and overlays them on a 2D map. NewsStand [Teitler et al. 2008] is a pioneering system that allows users to interactively explore photos directly from news articles depending on the query location and the zoom-level on a 2D map). Its successor, TwitterStand [Sankaranarayanan et al. 2009] is able to visualize tweets on a 2D map by using geo-tagged information as well as inferring geospatial relevance from the content of the tweets. Recently, PhotoStand [Samet et al. 2013], has shown how to visualize geo-tagged images from real-time scraped news on a 2D map. A primary distinction between Social Street View and the above systems is the use of immersive maps instead of 2D maps. In virtual environments, immersive maps require us to address challenges such as visual clutter, design of content layout in 3D, and registration.

In 3D environments, the well-known view-management system by Bell *et al.* [2001] registers user-annotated text and images to a particular point in 3D space. Their algorithm reduces visual discontinuities in dynamically labeled and annotated environments. Our system does not involve manual steps to visualize social media in immersive maps. SSV also reduces the visual clutter by maximal Poisson-disk sampling or road orientations. Recent efforts in novel social media visualization interfaces include Social Snapshot [Patro et al. 2010], Photo Tourism [Snavely et al. 2006] and 3D Wikipedia [Russell and Martin-Brualla 2013]. Instead of using immersive maps, Photo Tourism uses image-based modeling and rendering for navigating thousands of photographs at a single location. However, since Photo Tourism performs 3D scene reconstruction from unstructured photos, it is slow, taking a few hours to process a few hundred photos. The system also discards noisy, dark, cluttered photos due to registration failure. Similarly, 3D Wikipedia automatically transforms text and photos to an immersive 3D visualization but suffers from a relatively slow bundle adjustment and multi-view reconstruction. In contrast, Social Street View takes advantage of the large-scale 2.5D data and can visualize multiple geo-relevant photos in immersive environments at interactive rendering rates.

2.3 Analysis of Geo-tagged Social Media

An important question to consider for us is the accuracy of the geo-tagged media corresponding to the real geographic location. Zielstra and Hochmair [2013] conducted an experiment to investigate the positional accuracy of 211 image footprints for 6 cities in Europe by comparing the geo-tagged position of photos to the most likely place that they were taken at, based on image content. In this study, they found Panoramio has a median error distance of 24.5 meters and Flickr images have a median error distance of 58.5 meters. With the growing popularity of global positioning system (GPS) and Wi-Fi positioning systems on mobile devices, the geospatial metadata information is increasingly reliable and accurate. We have observed that occasionally there may be some irrelevant social media based on location query, but this is not common. Nevertheless, we have implemented a mechanism for users to report the abuse of geo-tagged social media in our system.

2.4 Mixed Reality in Immersive Maps

Past work on mixed reality in immersive maps generally required users to manually augment content for immersive maps. Devaux and Paparoditis [2010] added laser-scanned depth data to some street views and enabled users to manually add images or videos at their desired 3D position. Their system also had additional interactive features including human labelling, crowd sourcing mode to blur faces, and localizing and measuring objects. In contrast, by automatically extracting proximal social media content, SSV dynamically enhances the user experience in immersive maps and allows user to focus on social media interactions. Korah and Tsai [2011] convert large collections of LIDAR scans and street-view panoramas into a presentation that extracts semantically meaningful components of the scene via point-cloud segmentation. In addition, they propose an innovative compression algorithm and also show how to augment the scene with physics simulation, environmental lighting and shadows. Past research on analysis of immersive maps also addresses the important problems of segmentation, human recognition, and accessibility identification. For instance, Xiao and Quan [2009] propose a multi-view semantic segmentation framework using pair-wise Markov Random Fields (MRF) to differentiate ground, buildings, and people in street views.

3 System Architecture

In this section we discuss the overview of the system architecture of Social Street View presented in Fig 2. Social Street View consists of a street view scraper, a social media scraper, distributed SQL databases, a web-server powered by Apache and PHP, and optional modules such as a temporal filter, a geo-location filter, and a computer-vision-based face filter.

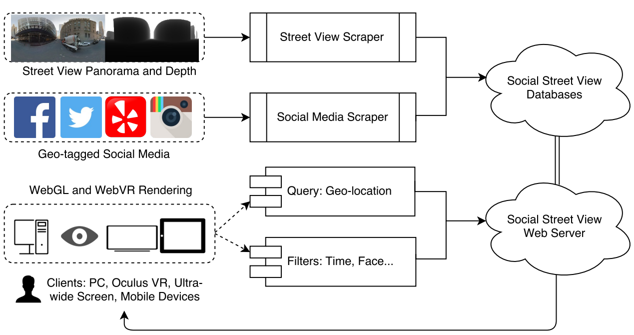

Figure 2: *The work flow of Social Street View. Our system streams data from two scrapers based on users' geo-location requests and renders social media in WebGL. Users can access the system via any WebGL-supported browser on a desktop, a tablet, a head-mounted display, or an immersive room-sized tiled display (see Fig. 1).*

3.1 Street View Scraper

Our street view scraper is a custom web-scraper tool written in Javascript and PHP that downloads GIS-related panoramic images and metadata at any geolocation where Google street view (GSV) data is available. Our tool is inspired by `GSVPano.js`[4], but we additionally scrape normal maps and road orientations. Currently, we request all street view data from Google Maps API. Each location query is analyzed by regular expressions, and thus it can be either a mailing address (e.g. 129 St., New York) or a pair of latitude and longitude coordinates (e.g. $40.2384, -70.2394$). For each location query, the scraper finds the closest panorama and downloads the following types of data:

1. **Tiles of panoramic images** with five types of resolution: from highest 13312×6656 pixels to the lowest 832×416 pixels. A stitched panorama is shown in Fig. 3(a).
2. **Depth map** from GSV meta data with a coarser resolution of 512×256 pixels. We normalize and up-sample the depth and normal maps in our GLSL shaders (Fig. 3(b)).
3. **Normal map** of 512×256 pixels (Fig. 3(c)).
4. **Road orientation and heading direction** indicates the travel direction of the GSV car or trekker.
5. **Geolocation and other information** including latitude, longitude, image age, and adjacent panoramas' hash IDs.

[4]GSVPano.js: https://github.com/heganoo/GSVPano

Figure 3: *The illustration of (a) stitched panoramic image tiles, (b) depth map and (c) normal map from Google Street View. The depth map is visualized in a yellow-red-black color scheme, where black indicates* 255 *meters or more, red indicates* 128 *meters and yellow indicates* 0 *meters. The normal map contains a 3D normal vector for each pixel. We visualize the normal data by converting the normal vector to HSV color space with blue-purple hues.*

3.2 Mining Social Media

Our geo-tagged social media scraper is a back-end program written in PHP. Tuchinda *et al.* [Tuchinda et al. 2008] have proposed to model the web services as information sources in a mediator-based architecture and have built exemplary application, Mashup. Using similar architecture, Social Street View is able to integrate information from several web services. For now, we use Instagram as the major source of social media in our proof-of-concept system. Since Instagram only allows users to upload images or videos from mobile devices, it largely avoids incorrect geo-tagged data from desktop clients. Our social media scraper collects the following data:

1. **Geospatial and textual location** including latitude and longitude coordinates, street names and user-tagged location name.
2. **Media type** indicating whether it is an image or a video.
3. **Caption and tags** containing text information.
4. **Published Time** containing the exact time-stamp when published.
5. **User comments and likes** reflecting the popularity level.
6. **URL** provides a link to the images or videos on the web.

Instagram API supports social-media query based on both geolocation and time. We use two distance thresholds, $\alpha = 10m$ and $\beta = 5km$ for dense urban areas and rural areas, respectively. Given a street view panorama, the scraper first requests social media within a radius of α on Instagram. If nothing is found, the scraper increases the threshold to β and queries again. If either threshold returns social media content, we send out R requests to acquire data over the past R months (we typically use $R = 12$). This allows Social Street View to answer spatial queries with a temporal filter, such as "*What do people wear in winter at this location?*.

3.3 Servers and Relational Databases

At present, we use distributed MySQL databases to store information of visited immersive maps and social media to reduce response time for duplicated or similar queries. One of the important components of our system is to build spatial data structures to efficiently answer proximity queries relating geo-tagged social media with spatially-located immersive panoramas. To accomplish this effectively, we build a bipartite graph that establishes edges between social media message nodes on one side with the immersive panorama nodes on the other side. This allows us to quickly answer what social media messages are relevant to be shown in any panorama. Since this can result in an all-pairs quadratic relationship, we needed to do this in an efficient manner. Once the two scrapers complete their tasks, the back-end servers build the bipartite graph in a separate thread: $G = \langle V, S, E \rangle$, where $V = \{v_i\}$ are the visited street views and $S = \{s_i\}$ are scraped social media, and $E = \{\langle v_i, s_j, d_{ij} \rangle \mid v_i \in V, s_j \in S\}$ are the edges between V and S. The weights of edges d_{ij} are defined by the distance between s_i and v_i according to the Haversine formula [Robusto 1957]:

$$\alpha_{ij} = \sin^2(\frac{\varphi_i - \varphi_j}{2}) + \cos(\varphi_i) \cdot \cos(\varphi_j) \cdot \sin^2(\frac{\lambda_i - \lambda_j}{2}) \quad (1)$$

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

$$d_{ij} = R \cdot \beta_{ij} \quad (3)$$

where φ_i, λ_i and φ_j, λ_j are the latitude and longitude of s_i and v_j respectively; R is the radius of the earth ($R = 6371km$); the result d_{ij} is the great-circle distance between s_i and v_j. Since both social media and street views are indexed by a $B+$ tree, the insertion and query without building the graphs takes $O(\log |V| + L)$ time, where $L = 100$ is the maximum number of queried social media. There could be an additional cost of sorting based on users' query and filters. However, the maintenance of the bipartite graph may take $O(k \cdot |V|)$ for k newly scraped social media S_k. To solve this issue, given a street view v_i, SSV calculates $S_i = \{s_j \mid \forall s_j \in S \wedge \langle v_i, s_j, d_{ij} \rangle \in E\}$ in $O(\log |V| + L)$. If $S_i = \emptyset$, Social Street View returns S_k and builds the bipartite graph at the back-end for the next query; otherwise, we return $S_k \cup S_i$ for more results. Thus, for each panorama, the streaming time is $O(\log |V| + L \log(L))$.

3.4 Social Street View Interface

Using WebGL and WebVR, we have designed and implemented an open-source and cross-platform system which is easy to access via most modern browsers and is shown in Fig. 4. Users can query any location in the input field on the top panel. The left optional panel has filters to specify query words, temporal (month and hour) ranges, distance ranges, number of faces and rendering controls. The bottom panel is an optional 2D visualization. We use `Three.js`[5], a cross-browser and GPU-accelerated Javascript library. We use `Bootstrap`[6] and `jQuery`[7] for 2D elements of the user interface.

4 Social Media Layout Algorithm

In this section, we present our approach to blend the visualization of social media together with street view panoramas in an immersive mixed reality setting.

[5]Three.js: http://www.threejs.org
[6]Bootstrap: http://www.getbootstrap.com
[7]jQuery: https://www.jquery.com

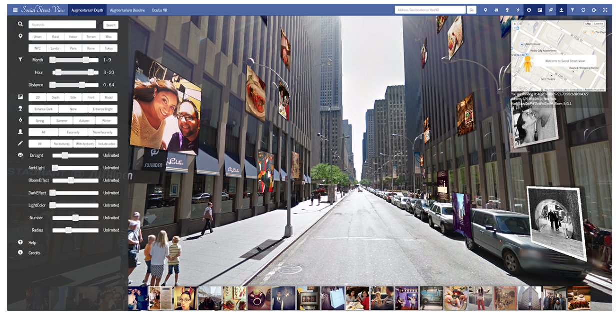

Figure 4: *The Social Street View interface powered by WebGL and WebVR.*

4.1 Baseline: 2D Visualization

Since there is limited previous work on visualizing multiple social media images in immersive maps, we first devised a basic 2D solution for users to have a quick glimpse of social media near their location. This is shown in the bottom panel of Fig 4. The users can click on any image to see a higher-resolution (640×640) image or video, text caption message, geo-location, and timestamp data. In addition, the user can use the image to link to Instagram to *comment*, *like*, or *forward* the social media. By clicking the geo-location of the social media, the user can navigate to the closest street view panorama to the image using Social Street View. Compared with Stweet that places a single Tweet message on a top layer above Google Street View with limited interactivity, this basic 2D visualization provides multiple queries near the center of panorama with richer information. In addition, the users can filter out the social media based on a desired time range and distance to the center of panorama.

4.2 Uniform Random Sampling

From Social Street View's server, the client acquires a subset of images or videos $\hat{S} = \{s_i \mid i = 1 \ldots N\}, \hat{S} \subseteq S$. Our goal is to place them naturally in low-saliency areas in an immersive map so that the social media rendering will minimize the visual clutter in a users view. To do this, we first stitch the tiles of panoramic images into a rectangle T. Next, we apply a projection $\mathscr{P} : T \rightarrow \Omega$ from T to the sphere Ω, thus building an immersive panoramic map with an inside camera looking outwards. For each point $\mathbf{p}_i = (u_i, v_i) \in T$, the corresponding point $\mathbf{q}_i = (x_i, y_i, z_i) \in \Omega$ is projected uniformly on a sphere. The easiest way to place social media is to randomly sample N points $\tilde{P} \subseteq T$ from the panorama. For each $\tilde{p} = (\tilde{u}, \tilde{v}) \in \tilde{P}$, calculate the corresponding 3D position $\tilde{q} = (\tilde{x}, \tilde{y}, \tilde{z}) \in \Omega$ as the center of the social media. As shown in Fig. 5(a), while we find that we can blend many interesting social media in this interactive mixed-reality world such as photography, food and people there are several drawbacks in the way they are laid out. To address this, we propose the following desiderata:

1. Ground-level context is important for way-finding for pedestrians and drivers. Therefore a system blending social media with immersive maps should minimize the rendering of the social media at or near the ground level.

2. Since billboards and other structures in real world are often aligned with physical landmarks, it is desirable to align social media with the proximal geometric structures.

3. Social media imagery should be reasonably spaced apart to avoid visual clutter and overlaps, if at all possible.

Figure 5: *Results before and after applying the depth map, the normal map and maximal Poisson-disk sampling. (a) shows the random layouts generated uniformly on the sphere, (b) shows results after using the depth map, (c) shows results after applying the normal map, (d) show results after using maximal Poisson-disk distribution for laying out photographs, (e) shows the original street view image, (f) and (g) visualize the depth map and the normal map respectively, for reference*

We next discuss how we accomplished the above goals in our work.

4.3 Depth and Normal-map-driven placement of social media

A depth map $D = \{d_i\}$ in which each depth value d_i corresponds to a pixel p_i in T, could be used to filter out points that are too close (e.g., the ground) or too far away (e.g., the sky). Additionally, we scale the size of images based on depth to give a perspective effect. The result is shown in Fig 5(b). This minimizes the images that are projected on to the ground or the sky.

We use the normal map $\mathbf{N} = \{\mathbf{n}_i\}$ to project the images onto the surfaces of buildings. Denoting the normal vector of the ground as $\mathbf{n}_g = (0, 1, 0)$, we define the ground level Ω_g as follows:

$$\Omega_g = \{q_i \mid \forall q_i \in \Omega \wedge \|\mathbf{n}_i - \mathbf{n}_g\| < \delta\} \quad (4)$$

where $\|\mathbf{n}_i - \mathbf{n}_g\|$ is the Euclidean distance between two vectors $\mathbf{n}_i$ and $\mathbf{n}_g$, $\delta = 0.5$ is a user-defined threshold. Next, for each sampled point $\tilde{\mathbf{p}}_i$, we use the corresponding normal vector $\mathbf{n}_i$ and rotate the social media to the correct orientation. The results are illustrated in Fig. 5(c). As one can see, the images are now well-aligned with the geometry of the buildings. However, the rendering still suffers from visual clutter and overlaps.

4.4 Maximal Poisson-disk Sampling

We use maximal Poisson-disk sampling to solve the problem of visual clutter. Poisson-disk distribution has been widely used in the field of computer graphics for global illumination [Shirley et al. 1991], object placement [McCool and Fiume 1992] and stochastic ray tracing [Cook 1986]. In our work, we follow the approach of the PixelPie algorithm devised by Ip *et al.* [Ip et al. 2013], which uses vertex and fragment shaders and GPU-based depth-testing features to efficiently implement the dart-throwing algorithm for maximal Poisson-disk sampling.[8]

After sampling points from the building surfaces, we sort them according to their depth. We preferentially place more *popular* social media closer, where popularity is defined in section 4.7. We outline this approach in Algorithm 1.

ALGORITHM 1: Social Media Layout using Poisson-disk Samples

Input: N sorted social media images $\hat{S} = \{s_i \mid i = 1 \dots N\}$, acquired from SSV servers.
Output: A set of image planes to display social media: $I = I_1 \dots I_M, M \le N$.
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 $i \leftarrow 1 \dots \min(N, |\tilde{P}|)$ **do**
  Place I_i with texture from $s_i \in \hat{S}$ at the projected position $\tilde{\mathbf{q}}_\mathbf{i} \leftarrow \mathscr{P}(\tilde{\mathbf{p}}_\mathbf{i})$;
  Rescale I_i according to the corresponding depth value: $\tau_i \leftarrow \tau / d_i$ for perspective visual effects;
  Rotate I_i so that it is perpendicular to the normal vector $\mathbf{n_i} \leftarrow \mathbf{N}(u_i, v_i)$;
  Add I_i to the result set: $I \leftarrow I \cup I_i$;
end

This provides us an aesthetic layout to display social media blended in with the immersive panorama. A screenshot of the resulting placement after this algorithm appears in Fig 5(d).

[8] Code of PixelPie: http://sourceforge.net/projects/pixelpie/

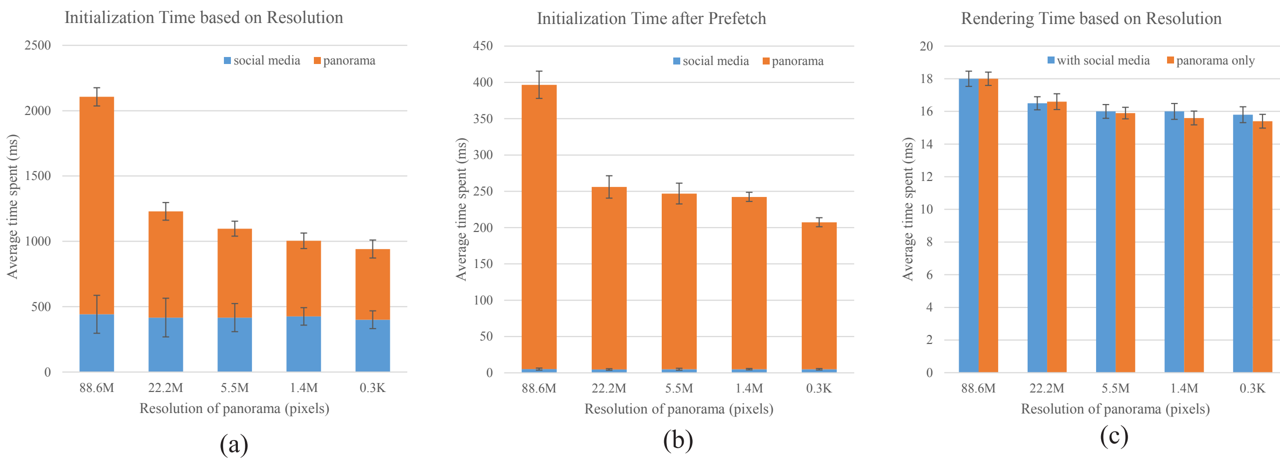

Figure 6: *Evaluation of processing time in different resolutions using* 100 *panoramas in Manhattan, (a) shows the initialization time decreases as the resolution goes down, (b) shows that by pre-fetching, initialization time is reduced by over* 3 *times; (c) shows after initialization, the rendering time costs about 16ms (about* 60 *FPS) in WebGL while the rendering of social media does not affect the rendering performance much.*

4.5 Placement of Social Media in Scenic Landscapes

In open and scenic areas, there are a limited number of surfaces on which to place social media. However, with the high-level knowledge of where the Google Street View camera is traveling from and traveling to, it is possible to place social media along the way without depth or normal maps. To generalize our system from urbanscapes with buildings to more general scenic landscapes, we propose Algorithm 2. Results are shown in Fig. 8.

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**
  Set the position $\mathbf{q_i} \leftarrow (KR\cos o_i,\ h,\ KR\sin o_i)$ at height h and radius R;
  (Optional based on user's preference) Add a frontal image plane to I at $\mathbf{q_i}$;
  Set the translation $\mathbf{t} \leftarrow (T\cos(o_i + \frac{\pi}{2}), 0, T\sin(o_i + \frac{\pi}{2}))$ with constant T;
  for $k \leftarrow 1 \dots K$ **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

4.6 Post-processing, Rendering and Interaction

To enhance the visual effects of the social media in an immersive setting, our system allows the users to add shadows, glowing shader effects, and alpha blending to the virtual billboards that depict the social media message. We can also model the difference between daytime and nights by using a blooming shader and an additive layer based on depth and normal. To experience the static street view in different seasons, we have implemented particle systems to render snow, falling leaves, or cherry petals in the scene. Additionally, we implemented a simple ray tracer that enables users to click on social media to read associated text.

4.7 Filtering of Social Media

In crowded areas such as the New York city, it is almost impossible to visualize every message in Social Street View. One solution is to give preference to the most popular social media. However, quantifying popularity itself is subjective since popularity spans features such as comments, replies, creation time, likes, and number of times forwarded. We have adopted the following criteria as a substitute for popularity:

$$\frac{\alpha C_i + L_i}{\Delta T_i} \tag{5}$$

where, for a given social media s_i, C_i is the number of comments, L_i is the number of likes, and ΔT_i is age of the social media message. Since comments generally have a higher impact than likes, we scale comments by a user-defined scaling factor α.

To protect potential privacy concerns or to find celebrity figures from news, we also incorporate the face filter using Face++ API [9]. Users can also filter social media based on time and distance.

5 Experiments and Evaluation

We have carried out a number of experiments to evaluate the Social Street View system. Here we report some of our results for a variety of Google Street View resolutions and social media.

5.1 Dataset Acquisition and Hardware Setup

We scraped 100 Google Street View panoramas from Manhattan in New York City as our main dataset. We found over 84,055 social media images on Instagram within a query distance of 20 meters in these panoramas. For each query, our system returns 100 closest social media images according to the distance to the panorama by searching in $B+$ trees. The experiments were conducted in Google Chrome (Version 40.0.2214.115 m) with Nvidia Quadro K6000 and Intel Xeon CPU E5-2667 2.90GHz. The rendering resolution we used is 2650×1440 pixels. To reduce the effects of different network latency, we store all the panoramas and social media on the

[9] Face++: http://www.faceplusplus.com

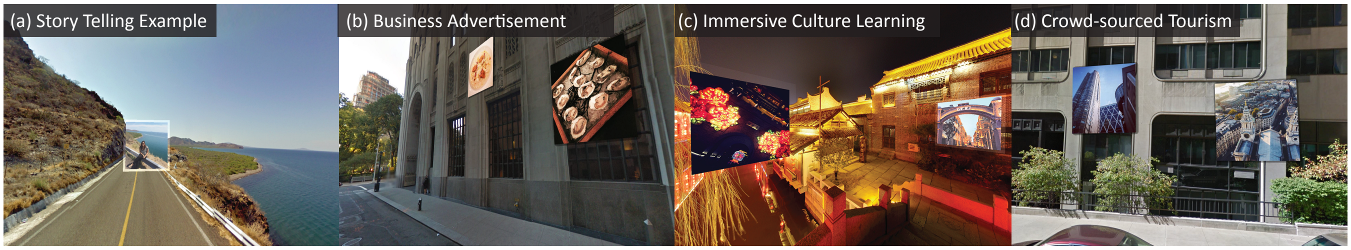

Figure 8: *Potential applications of Social Street View: (a) Users can link to Social Street View to tell immersive stories (b) Business owners can use Social Street View for impressive advertising (c) Children can learn culture from local social media (d) Tourists can preview the trip from crowd-sourced photographs embedded in the immersive maps.*

local disk of the workstation. We present the file sizes in Table 1 for a variety of panoramas.

Table 1: *Resolution, tile counts and file size of Google Street View (GSV) panoramic data*

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	$\sim 90K$

5.2 Evaluation of Initialization and Rendering Time

Interactivity and latency are of great importance for user navigation in the Social Street View. Fig. 6(a) shows the initialization time based on five different resolutions in Table 1. The initialization mainly includes the time spent in querying for panorama and social media, as well as loading texture to memory and WebGL initialization. We notice that when the size of the panoramic texture is reduced to 6656×3328, the overall time cost is reduced from approximate two seconds to one. Choosing an appropriate resolution based on a users display, downloading speed, and GPU power can therefore make a meaningful difference. We noticed that at least 900ms was being spent on initializing panoramas and social media. To reduce the time when switching between adjacent street views, we pre-fetch the data in memory for faster initialization. Google Street View uses progressive refinement to address latency by initially loading low-resolution images that are refined to higher-resolution ones over time. We can also rely on such an approach when the local disk is unable to support pre-fetching. In Fig. 6(b), we show how querying time is reduced by pre-fetching. Further, maximal Poisson-disk samples to place 100 social media can be generated at interactive rates. However, the texture loading can still take hundreds of milliseconds and we hope that it could be improved with further advances in the WebGL technology. In Fig. 6(c), we report the rendering time with and without social media. From the chart, it can be seen that the system runs at around 60 frames per second (fps) for all resolutions. Thus, the rendering of the social media does not affect the experience of navigating Social Street View.

5.3 Evaluation of Saliency Coverage

Saliency maps can represent regions where a user is likely to allocate visual attention in a fixed-time free-viewing scenario [Itti et al. 1998; Harel et al. 2006; Kim et al. 2010; Ip and Varshney 2011]. We compute image saliency using the Matlab tool by Hou *et al.* [Hou et al. 2011] to evaluate the social media coverage of saliency maps. An example of such a saliency map is shown in Fig 7 (a).

The average social media coverage of saliency maps over all the 100 immersive Google Street View panoramas is illustrated in Fig 7(b). Initially, the saliency map is covered by uniform random sampling algorithm at about 35%; after incorporating the depth map, most sky areas are filtered out but social media are highly likely to cover the vanishing point where saliency is high; after incorporating depth and normal maps, most social media are aligned to the building structures where saliency is low in most cases; with maximal Poisson-disk sampling, the social media distributes evenly and aesthetically, thus reducing the likelihood that several images could overlap in a high-saliency area. This is the reason why the maximal Poisson-disk sampling has a significantly lower standard deviation in saliency coverage in Figure 7 (b). In contrast, the uniform and random sampling, as well as approaches that rely only on depth-map-based placement, result in a larger coverage of higher saliency regions.

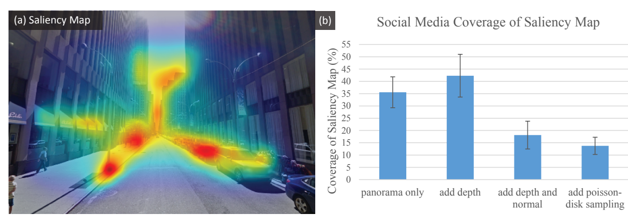

Figure 7: *(a) Street view with saliency map overlay. The visualization has a red-yellow-green-transparent scheme, where red indicates high saliency and transparent indicates low-saliency. (b) Evaluation results of saliency coverage from 100 immersive GSV panoramas.*

6 Use Cases and Discussion

While exploring Social Street View in a variety of scenarios using publicly available social media, we discovered a number of potential use cases that are promising in enhancing storytelling, business advertising, learning cultures and languages, and in visual analytics of social media in a spatio-temporal context.

6.1 Storytelling

Social Street View could greatly enhance the storytelling experience. For example, users could see photos from recent trips of their friends while allowing them to explore the 360° context in an immersive setting.

In Fig. 8(a), we present how social media stories can be more convincing using Social Street View. This panorama is along a road in Baja California Sur within Mexico[10]. Since this open road does not have vertical proximal structures, we use the scenic landscape layout mode here. Further, because it is along a long road we did not expect anyone to take photos and upload them to the social media. Nevertheless, our system found 3 images within a radius of 20 meters. In one of the images, Instagram user Daniela wrote on $July\ 12,\ 2014$:

> *Stuck in traffic on our way to Cabo with this awesome view #roadtrip #cabo #view #mexico*

When we pan and walk around this location, we are also impressed by *this awesome view*. We like to think of this as our system facilitating a democratized, crowd-sourced version of the *Kodak Picture Spot*.

6.2 Business Advertising

Social Street View can also be used for business advertising. For example, restaurant managers could showcase the social media photographs of their dishes shared by their customers in the context of the interior ambiance of their restaurants. Similarly, real-estate customers could view the neighborhood street view augmented by the dynamism of the social media of that community to get a better feel for their prospective neighbors.

Fig. 8(b) uses a panorama in *6 E 24th St, New York, United States*. In the rightmost image of dishes, an Instagram user frankiextah commented:

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

With mixed-reality rendering, Social Street View enhances future consumers' visual memories and makes it easier for them to seek *"amazing oysters"* around this place.

6.3 Learning Culture and Crowd-sourced Tourism

Immersive virtual environments have been used to protect world's cultural heritage and serve as a useful medium for cultural education [Addison 2001]. However, it is usually challenging to generate relevant captions and up-to-date photographs for each scene of a virtual environment. By blending crowd-sourced social media content with panoramic imagery, Social Street View can (with age-appropriate filters and curation) serve as an educational tool for children and researchers to learn cultures and languages in different cities and countries. As shown in Fig. 8(c), users can experience the holiday atmosphere of the Spring Festivals in Taierzhuang Ancient Town of China, where the oldest "living ancient canal" was built in the Ming Dynasty and the Qing Dynasty. Here again, because of a lack of sufficient vertical structures, one can enable the scenic-landscape mode and visualize recent photographs of the architecture taken by tourists in the daytime and at night.

Fig. 8 (d) presents an example of the crowd-sourced tourism in urban areas. Using face and popularity filters, users can get rid of most pictures with human faces and blend some high-quality photographs with a New York street. These photographs provide novel views for the user's exploring experience.

7 Conclusions and Future Work

In this paper we have presented Social Street View, a system to create immersive social maps that blend street view panoramas with geo-tagged social media. Our contributions include: (a) system architecture to scrape, query, and render geo-tagged street view and social media together on clients ranging from smartphones to tiled display walls to head-mounted displays using WebGL, (b) techniques to carefully layout and display social media on virtual billboards by a judicious combination of depth maps, normal maps, and maximal Poisson sampling, and (c) validating the efficiency of such mixed-reality visualizations for saliency coverage. We have also presented several potential use cases of exploring social media with temporal and spatial filters and storytelling with spatial context. Further supplementary material is available at http://augmentarium.umd.edu and http://socialstreetview.com.

There are several directions in which Social Street View could be extended. First, we are currently using low-level image saliency for evaluation but not for layout. With real-time or post-processed saliency maps, we can avoid placing social media on high-saliency regions to avoid occluding the important features in the panorama. Also, it is important to note here that saliency maps can incorporate low-level image features, as well as high-level semantics in guiding social media placement. Conversely, one can also extend previous work on visual persuasion of gaze direction in the context of social media by altering visual and geometric appearance attributes [Kim and Varshney 2006; Kim and Varshney 2008] Second, we plan to evaluate the relative merits of Social Street View use in high-resolution large-area screens compared to head-mounted displays. With the state-of-the-art interaction techniques such as gesture, speech, and head-tracking, our system could provide VR and AR users with real-time useful information from social media. Other potential future directions include 3D reconstruction, depth optimization and fusion, and spatial data mining of the social media with location-aware context. While our prototype system works well for mapping social media onto immersive panoramas of buildings and grounds in both urban and rural areas, it currently uses coarse depth maps. If accurate 3D reconstructions are available our system could experience a quantum leap in features and usability. The potential applications of visualizing geo-tagged social media are immense. With geo-tagged news datasets, journalists could generate impressive news stories using Social Street View. With geo-tagged restaurant datasets, customers could *see through* the restaurants in street view to explore the ambiance and the served dishes. With geo-tagged accessibility information from social media, such as missing curb ramps, we could enhance the street view to help the disabled.

Acknowledgements

We wish to thank *Shangfu Peng* from the Department of Computer Science at the University of Maryland, College Park for collaborating with visualizing social media in a 2D map (*i.e.*, Geollery at http://sametphp.umiacs.umd.edu/geollery) [Peng et al. 2014], *Tim Gray* and *Derek Yarnell* from UMIACS for helping set up Apache servers and SQL databases, *Sai Yuan* from the Department of Animal Science for helping with the vocal track of the demo, *Xuetong Sun* and *Sujal Bista* from UMIACS for helping with testing Social Street View in immersive virtual environments. We thank the anonymous reviewers from the ACM SIGGRAPH and Web3D community for their meticulous reading of our manuscript, as well as their insightful suggestions and comments.

This work has been supported in part by the NSF Grants 09-59979, 14-29404, 15-64212, the State of Maryland's MPower initiative, and the NVIDIA CUDA Center of Excellence. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the research sponsors.

[10]Geo-location: North 25.855319593, West 111.333931591

References

ADDISON, A. C. 2001. Virtual Heritage: Technology in the Service of Culture. In *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage*, ACM Press, 343–380.

ANGUELOV, D., DULONG, C., FILIP, D., FRUEH, C., LAFON, S., LYON, R., OGALE, A., VINCENT, L., AND WEAVER, J. 2010. Google Street View: Capturing the World at Street Level. *Computer 43*, 6 (June), 32–38.

BELL, B., FEINER, S., AND HÖLLERER, T. 2001. View Management for Virtual and Augmented Reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology - UIST '01*, 101–110.

COOK, R. L. 1986. Stochastic Sampling in Computer Graphics. *ACM Transactions on Graphics (TOG) 5*, 1, 51–72.

DEVAUX, A., AND PAPARODITIS, N. 2010. Increasing Interactivity in Street View Web Navigation Systems. In *Proceedings of the International Conference on Multimedia*, ACM Press, 903–906.

HAREL, J., KOCH, C., AND PERONA, P. 2006. Graph-Based Visual Saliency. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, MIT Press, Vancouver, British Columbia, Canada, 545–552.

HOU, X., HAREL, J., AND KOCH, C. 2011. Image Signature: Highlighting Sparse Salient Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 34*, 1 (July), 194–201.

IP, C. Y., AND VARSHNEY, A. 2011. Saliency-Assisted Navigation of Very Large Landscape Images. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 17*, 12, 1737–1746.

IP, C. Y., YALÇIN, M. A., LUEBKE, D., AND VARSHNEY, A. 2013. PixelPie: Maximal Poisson-disk Sampling with Rasterization. In *Proceedings of the 5th Conference on High-Performance Graphics - HPG '13*, ACM Press, vol. 2, 17.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI)*, 11, 1254–1259.

KAPLAN, A. M., AND HAENLEIN, M. 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons 53*, 1 (Jan.), 59–68.

KIM, Y., AND VARSHNEY, A. 2006. Saliency-Guided Enhancement for Volume Visualization. *IEEE Transactions on Visualization and Computer Graphics 12*, 5, 925–932.

KIM, Y., AND VARSHNEY, A. 2008. Persuading Visual Attention Through Geometry. *IEEE Transactions on Visualization and Computer Graphics 14*, 4, 772–782.

KIM, Y., VARSHNEY, A., JACOBS, D. W., AND GUIMBRETIÈRE, F. 2010. Mesh Saliency and Human Eye Fixations. *ACM Transactions on Applied Perception (TAP) 7*, 2, 12.

KORAH, T., AND TSAI, Y.-T. 2011. Urban Canvas: Unfreezing Street-View Imagery With Semantically Compressed LIDAR Pointclouds. In *Proceedings of 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, 271–272.

MCCOOL, M., AND FIUME, E. 1992. Hierarchical Poisson Disk Sampling Distributions. In *Proceedings of Graphics Interface '92*, vol. 92, Canadian Information Processing Society, 94–105.

PATRO, R., IP, C. Y., BISTA, S., AND VARSHNEY, A. 2010. Social Snapshot: A System for Temporally Coupled Social Photography. *IEEE Computer Graphics and Applications*, 1, 74–84.

PENG, S., SAMET, H., AND ADELFIO, M. D. 2014. Viewing Streaming Spatially-referenced Data at Interactive Rates. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 409–412.

ROBUSTO, C. C. 1957. The Cosine-Haversine Formula. *The American Mathematical Monthly 64*, 1, 38–40.

RUSSELL, B., AND MARTIN-BRUALLA, R. 2013. 3D Wikipedia: Using Online Text to Automatically Label and Navigate Reconstructed Geometry. *ACM Transactions on Graphics (TOG) 32*, 6, 1–10.

SAMET, H., ADELFIO, M. D., FRUIN, B. C., LIEBERMAN, M. D., AND SANKARANARAYANAN, J. 2013. PhotoStand: A Map Query Interface for a Database of News Photos. *Proceedings of the VLDB Endowment 6*, 12 (Aug.), 1350–1353.

SANKARANARAYANAN, J., SAMET, H., TEITLER, B. E., LIEBERMAN, M. D., AND SPERLING, J. 2009. TwitterStand: News in Tweets. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM Press, 42–52.

SERDYUKOV, P., MURDOCK, V., AND VAN ZWOL, R. 2009. Placing flickr photos on a map. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 484–491.

SHIRLEY, P., SUNG, K., AND BROWN, W. 1991. A Ray Tracing Framework for Global Illumination Systems. In *Proceedings of Graphics Interface '91*, Canadian Information Processing Society, 117–128.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3d. In *ACM Transactions on Graphics (TOG)*, vol. 25, ACM, 835–846.

TEITLER, B. E., LIEBERMAN, M. D., PANOZZO, D., SANKARANARAYANAN, J., SAMET, H., AND SPERLING, J. 2008. NewsStand: A New View on News. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '08*, ACM Press, no. 18, 1–10.

TORII, A., HAVLENA, M., AND PAJDLA, T. 2009. From Google Street View to 3D City Models. *Proceedings of 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, 2188–2195.

TUCHINDA, R., SZEKELY, P., KNOBLOCK, C. A., AND REY, M. 2008. Building Mashups By Example. In *The 13th International Conference on Intelligent User Interfaces*, 139–148.

XIAO, J., AND QUAN, L. 2009. Multiple View Semantic Segmentation for Street View Images. In *2009 IEEE 12th International Conference on Computer Vision*, IEEE, 686–693.

ZIELSTRA, D., AND HOCHMAIR, H. H. 2013. Positional Accuracy Analysis of Flickr and Panoramio Images for Selected World Regions. *Journal of Spatial Science 58*, 954, 251–273.