5-2022

VisualCommunity: a platform for archiving and studying communities

Suphanut Jamonnak

Deepshikha Bhati

Md Amiruzzaman

Ye Zhao

Xinyue Ye

See next page for additional authors

Authors

Suphanut Jamonnak, Deepshikha Bhati, Md Amiruzzaman, Ye Zhao, Xinyue Ye, and Andrew Curtis

VisualCommunity: a platform for archiving and studying communities

Suphanut Jamonnak[1] · Deepshikha Bhati[1] · Md Amiruzzaman[2] · Ye Zhao[1] · Xinyue Ye[3] · Andrew Curtis[4]

Abstract

VisualCommunity is a platform designed to support community or neighborhood scale research. The platform integrates mobile, AI, visualization techniques, along with tools to help domain researchers, practitioners, and students collecting and working with spatialized video and geo-narratives. These data, which provide granular spatialized imagery and associated context gained through expert commentary have previously provided value in understanding various community-scale challenges. This paper further enhances this work AI-based image processing and speech transcription tools available in VisualCommunity, allowing for the easy exploration of the acquired semantic and visual information about the area under investigation. In this paper we describe the specific advances through use case examples including COVID-19 related scenarios.

Keywords AI processing · Community study · Geo-narrative · Spatial video · Visualization system

✉ Suphanut Jamonnak
  sjamonna@kent.edu

1   Department of Computer Science, Kent State University, Kent, OH, USA

2   Department of Computer Science, West Chester University, West Chester, PA, USA

3   Department of Landscape Architecture and Urban Planning, Texas A&M University, College Station, TX, USA

4   Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA

🖄 Springer

Introduction

A "community"[1] is a complex construct of environment, geography, and behavior. It also contains multiple challenges, research topics, or success stories that can be investigated. While there is no single research frame to address this complexity, there is often a commonality in terms of visual data that ties these perspectives together. For example all might benefit from a video of a street. Similarly, that common source can be used as a data source, potentially enhanced with expert or local knowledge. For example, spatially coded data including sketch maps, videos, photos, and narratives can provide valuable insights into the granular processes at work in a community. These perspectives, often acquired from local residents, workers, or service professionals enrich or even replace more "official" data. One approach to collect such granular data is through using global positioning system (GPS) enriched video cameras, often mounted on a vehicle [1]. Narratives about the environment being traversed can also be recorded to create a spatial video geonarrative. The use of these methods can capture fine-scale changes and provides context to help explain process, discover processes and patterns, even across multiple time periods of collection, with insights gained being used to guide intervention (e.g., [1–5]). Topics could vary from crime or overdoses in high density residential or business districts, to environmental risks such as stagnant water and trash being predictive of disease outbreaks. Stakeholders can include various academic disciplines, planners and policymakers, to service providers and local community groups and residents.

Typically these data can include video, images, audio and associated transcriptions all of which can be tied together spatially with the GPS stream. While the potential use of these data in social studies and practices is large, widespread implementation is often hindered due to the lack of:

- A fully functional platform providing both data capture and data analysis functions, with a seamless linkage between them.
- Data processing capabilities for extracting semantic information from the geocoded videos and for transcribing the recorded audio narratives.
- A visualization platform supporting interactive exploration of the visual, semantic, and geospatial information.

Existing visual analytic (VA) tools for analyzing vehicle/human trajectories (e.g., [6]), geo-spatial events (e.g., [7]), geo-tagged social media (e.g., [8]), and multimedia data (e.g., [9]) cannot be directly applied to these types of geo-coded video datasets. Recently, a GeoVisuals system [10] was developed to visually explore similar spatially encoded datasets. However, this system is not easily applied by a community user (see "Motivation and tasks" scetion) due to (1) the lack of

---

[1] We understand the complex nature of the term "community"—in this paper we only use this to mean a small (within city/rural area/similar environmental) area.

processing functions and system integration, and (2) the complexity restricting non-technical users.

To help address this gap we develop a computational platform, named VisualCommunity, for easy community data capture, archiving, and analysis. This platform includes (1) a mobile APP (named as GeoVideo) helping users capture geo-coded video and narrative data with their mobile phones [11], and (2) a visualization system for easy and interactive study of the various acquired data. AI tools for semantic image segmentation and speech transcription are included, so that users can automatically process the collected data to extract visual and textual semantic information, which can also be linked to the associated location on the map.

A key aspect of VisualCommunity's creation is usability and transferability amongst users; domain experts were consulted throughout its development process; the mobile APP has been developed for both iOS and Android systems and is available in the corresponding APP stores [11]; and the visualization system is publicly available for download onto desktops or laptops environments. In this paper, we introduce the design and development of the platform, and present examples to illustrate functionality and effectiveness. We also discuss current limitations and the future work plan.

Related work

Community study with geospatial videos and narratives

Collecting and analyzing spatially encoded video data has proved to be an effective way of capturing, mapping and analyzing community processes. This has been aided by the widespread availability of GPS technologies, including smart phones and spatially supported cameras. In effect this means that researchers, professionals, and even community members can now record the type of visual, and spatially precise data that were previously unavailable through cost [2]. Adding in a simultaneously recorded commentary, often described as a geonarrative, further enriches these data with context. As a result, multiple topics including health, crime [12], and disasters have been studied using this approach both within the United States and in various challenging oversea environments [3, 13–15]. The challenge has always been that the associated software, including spatial data processing and analysis tools have lagged behind the technological and methodological advances. In this regards, only bespoke (and therefore niche limited) visualization systems allow users to fully explore spatial video and geo-narratives associated with their geo-trajectories.

Urban study with visual information

The utility of online geo-locatable imagery such as Google Streetview (GSV) [16] has further opened the potential to include neighborhood audits into research. Topics have included assessing disaster related damage [17], understanding the association

between the built environment and health [18], finding green areas [19], locating criminogenic environments [20, 21], and even identifying animal activity [22]. However, while useful, these data also have multiple limitations, including limited times for data collection, which can be problematic if the goal is to assess specific environments to an event that had happened, such as after a disaster, or to capture temporal change. One such challenge is that while there have been technological advances in the equipment used, the ability to manipulate and visualize the collected data lags behind. Our platform helps address this gap in a way not previously available.

In the proposed platform street level images were used as a data source to search for localized detail [12, 23–25]. The images were segmented with machine learning (ML) tools and the identified semantic categories (e.g., greenery, building, road, etc.) were utilized to characterize, cluster, and visualize urban forms. These methods utilize street images without narratives, while our platform explores spatial videos linked to geo-narratives.

Visual analytics systems for urban study

While as mentioned there have been various visual analytic methods and tools developed (e.g., [6, 26]) to fully leverage geo-spatial data, spatial videos combined with geonarratives provides both a new data direction and associated challenges. The team had previously built GeoVisuals [10] to interactively manage, visualize, and analyze spatial video and narratives using a set of visualization widgets and interaction functions. This offered a variety of data investigations based on keywords extracted from geo-narratives tied to images and locations on the videos. Users could quickly find important locations based on term frequency and sentiments. However, GeoVisuals required that the source data had already been acquired and processed (meaning having separate video and audio) while the extracted keywords, trajectories, and videos were integrated with a special data structure within a spatial database. This previous system also did not benefit from AI. More detailed comparisons are shown in "Motivation and tasks" section.

Overview

Target community data

Many workers, stakeholders, and social scientists often try to understand a community and tackle social problems at fine geographic scales. One data source to achieve this is a combination of community scale visuals enhanced by firsthand insights and opinions. To meet this research need, VisualCommunity incorporates the following heterogeneous data:

- *Geospatial video* that collects continuous video of an environment with each frame being geotagged.

- *Audio narrative* often recorded at the same time of the video by expert or informed individuals who provide a commentary on the environment being traversed, often identifying Points of Interest (POI).
- *Geo-trajectory* that associates the spatial video and narrative with GPS locations along the recording trip.
- *Geo-structures* of the environment such as streets and POIs.

These data items are often archived and organized together in a *data-capture trip*.

Motivation and tasks

Domain researchers in fields such as geography, public health, criminology, and other urban focused social sciences identified a set of research requirements if spatial video and geonarratives were to be used as an effective research method.

One identified challenge was that there was no single integrated system for data collection and analysis. A variety of off-the-shelf video cameras (e.g., mounted on cars) were used to collect data. Quite often the video players supplied by the camera's manufacturer were used to view the videos. The geotrajectories were displayed again either in the camera's own software, or through options available in a GIS (e.g., ArcGIS). To leverage the narratives, the audio had to be converted to text for browsing and searching. The overall process was fragmented and lacked the consistency often required for scientific inquiry.

Second, when the existing VA system, GeoVisuals [10], was used, domain experts could perform visual analytic over visual+semantic information. However, they identified several limitations: (1) GeoVisuals did not support data capture and processing. They had to load the videos into a database and the audio still had to be transcribed prior to upload. (2) GeoVisuals required the install of a spatial database, a Web server service, and there were system configuration issues which troubled less technologically savvy researchers. (3) GeoVisuals integrated a set of visual interfaces with video mode and trip mode. Combining complex visual metaphors of maps, texts, videos, and pictures in an interface quickly became overwhelming for users. The coordination of views and interactions demanded a relatively long learning curve. All of these issues meant that the platform, while offering an improved research experience, suffered from widespread use.

VisualCommunity platform is developed to address these limitations. In Table 1, we summarized the itemized comparison and also identified a set of tasks for VisualCommunity including:

- **T1:** Supporting convenient capture of geo-coded video and audio narrative with smartphones or tablets.
- **T2:** Automatically extracting landscape images and their semantic contents from the video.
- **T3:** Automatically transcribing audio narratives to text.

Table 1 Comparison to existing domain approaches

| Tasks	Popular tools of domain users	GeoVisuals system	VisualCommunity platform
Community data capture	Video camera toolkits	No support	**T1:** Mobile APPs on iOS/Android
Video data processing	No processing	No support	**T2:** AI based image semantic segmentation
Audio data processing	Manual transcription	No support	**T3:** AI based speech transcription
Narrative study	Basic term search	Study with term frequency and sentiment	**T4:** Edit narratives, add comments and study with terms
Visual content study	Video and image browsing	Video and image browsing	**T5:** Browse video/image and study with semantic categories
Data analysis tools	ArcGIS, Video player, word	Complex GeoVisuals interface	**T6:** Concise visualization interface
System feature	Fragmented and manual work	Hard to install, no data capture and processing	**G1:** Integrated platform with easy installation
Data analysis	Inefficient and time-consuming	Powerful but difficult to learn and use	**G2:** Easy to learn and use
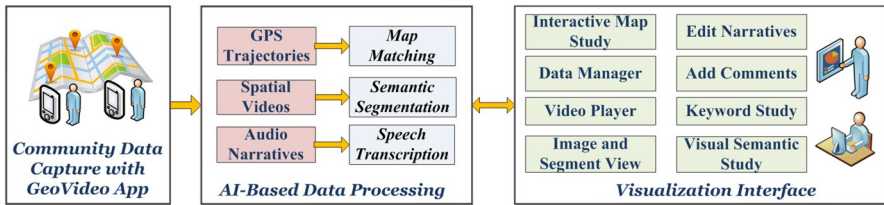
Fig. 1 VisualCommunity framework consists of (1) data capture with GeoVideo mobile APP; (2) AI-based data processing including map matching, semantic segmentation, and speech transcription; (3) visualization interface with coordinated visual components

- **T4:** Allowing users to edit narratives for errors and typos, and to add comments during analysis.
- **T5:** Facilitating users to perform analysis based on the extracted visual contents.
- **T6:** User friendly visualizations for non-technical users.

To implement these tasks, VisualCommunity was developed with two major goals:

- **G1:** Providing an integrated platform that enables a fluid workflow from data archiving, to visualizations, analysis, spatial inquiry and mining themes.
- **G2:** Making the platform easy to master and use for community data exploration. It should therefore be easy to deploy and install on different machines.

Workflow

VisualCommunity workflow and structure are illustrated in Fig. 1. The geo-coded video and narrative data can be conveniently captured by a GeoVideo mobile APP, which is available for iOS or Android platforms [11]. Section 4 shows its functions. The captured data can be easily transferred to the visual system running on desktops or laptops. Then, AI-based image semantic segmentation can automatically extract semantic categories (e.g., road, building, person, etc.) from consecutive video frames. Meanwhile, automatic speech transcription from a DNN (Deep Neural Network) model is applied which transforms audio data into a textual narration. This information is integrated into the system with the map matching corresponding trajectories to geographical context. We introduce the tools and functions in "AI-based data processing" section. Eventually, users can perform interactive data editing and exploration with intuitive visualization tools for the video, audio, image, and semantic information, which will be discussed in "VisualCommunity visualization system" section.
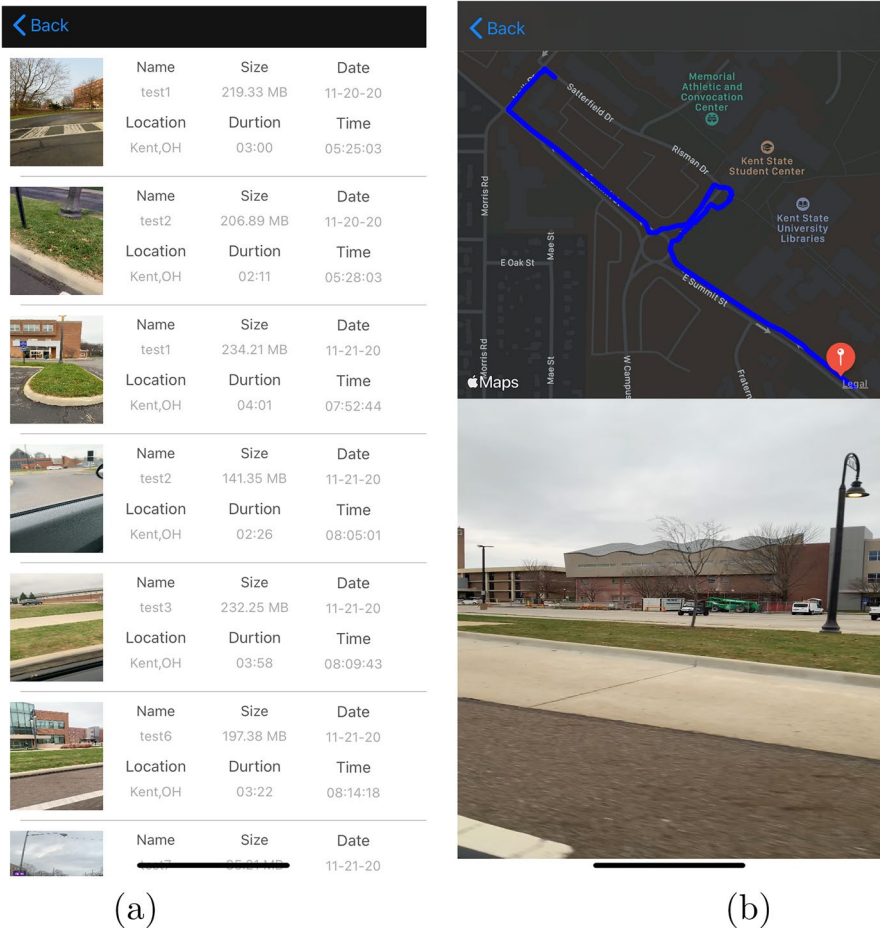
Fig. 2 GeoVideo interface snapshots (iOS version): **a** Listing recorded files. **b** Playing a recorded video along with its mapped geo-trajectory

Data archiving with GeoVideo app

Many domain users use cameras with internal GPS receivers mounted on vehicles (cars or bikes) or carried by hand to capture environmental data covering topics from post-disaster landscapes to infectious disease risk. While many use video cameras there is a need for more commonly available technologies such as a mobile phone or tablet integrating a GPS unit and a video camera. Some existing apps collect geo-trajectories such as [27, 28]. However, there is not a convenient mobile app available that can combine geo-videos with trajectories.
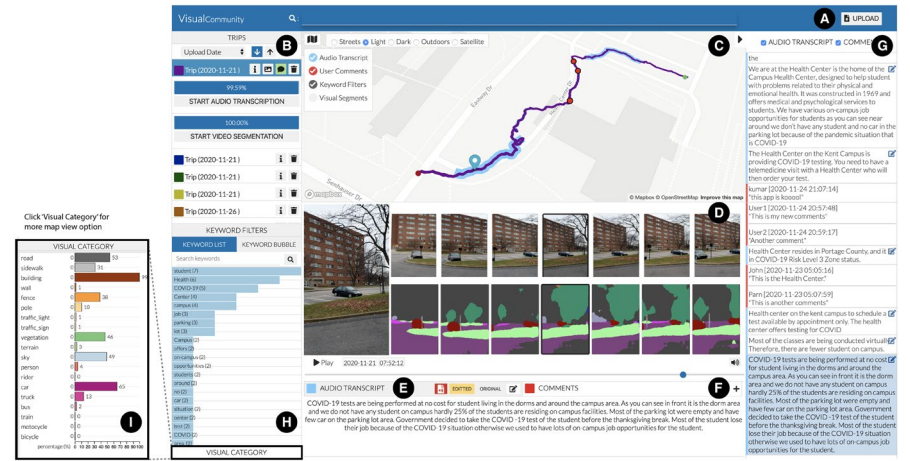
Fig. 3 VisualCommunity interface: **A** Upload data, **B** data management panel, **C** map view, **D** visual information view with video player, **E** narrative editor, **F** comment editor, **G** narrative view of speech transcript and user comments, **H** keyword filter, **I** visual category view

The GeoVideo App is designed for this purpose, to collect spatial video and audio together with a geographical trajectory. An active Internet connection is not required which is critical for oversea and generally resource challenged areas.

Once the app is opened it displays the current location of the user on a map view. Users can then start to record videos and trajectories. A list of recorded files is shown in Fig. 2a. The user can select one of them and play it. Figure 2b is the visualization interface where the selected video is played along with its corresponding trajectory displayed on the map. The map style can be easily changed. A marker moves along the trajectory to show the location of video content as the route progresses, with an option to be dragged anywhere on the route to show imagery at that location. The saved files which combine the video, audio, and GPS trajectory data are bundled for upload directly to the VisualCommunity visualization system for analysis.

GeoVideo stores the GPS trajectories in a .CSV (comma-separated values) format consisting of sequences of sampling points (longitude, latitude) for each second. Video clips are stored in .MP4 or .MOV format from mobile phones. The timestamps of the video are used to match the video to geographical locations.[2]

AI-based data processing

Once a user uploads a captured dataset to the VisualCommunity system, two AI algorithms are integrated based on deep learning neural networks. They are used for automatic information retrieval from the spatial video and audio narratives. The AI-based functions are used to automatically process the captured data, which are used

---

[2] GeoVideo has been developed and tested on different mobile platforms. It is available on Google Play App Store and Apple App Store.

later in extracting more meaningful information (see "Semantic content extraction from videos" and "Speech transcription from narratives" sections). In this study, we use PSPNet deep learning model to extract semantic object information from videos [29, 30], and DeepSpeech to extract transcripts from audio data [31]. Both of these AI-based tools are state-of-the-art models when this study was conducted. In "Scenario 3: data transfer and processing performance" section, we show an example of the effectiveness and latencies of these functions.

Semantic content extraction from videos

The video stream is broken down into image frames with a time interval of one second (e.g., using Ffmpeg libraries) though the time interval can be adjusted. These image frames are processed by the AI tool PSPNet for semantic content extraction. PSPNet is based on a Pyramid scene parsing deep neural network [29, 30]. This model is proven to provide better performance in benchmark datasets, such as PASCAL VOC2012 and cityscapes [29].

A pixel-level classification of each streetview image uses 19 categories including road, sidewalk, building, person, car, etc. In addition, the proportion of each in an image is recorded (for example, does "road" occupy 50% of the pixel). As shown in Fig. 3D, road (purple), car (red), meadow (green) are extracted from a sequence of image frames.

Speech transcription from narratives

The audio track is extracted and processed by the deep learning network of DeepSpeech [31]. DeepSpeech is an open-source embedded speech-to-text engine which can run in real-time on multiple types of devices. Integration occurs as a long audio stream split into smaller sections, which achieves a better transcription quality than if the whole stream is processed as one.

First, the AI algorithm transcribes the audio narrative to a text document consisting of a list of speech fragments. Each fragment is a natural language segment of the speaker's narration based on their talking speed, stopping points, and other attributes. It consists of multiple terms (i.e., keywords) while each term has unique term speaking length (audio length). As shown in the example in Fig. 3G, these fragments are visualized following the time sequence which is also the location on the trajectory. The text inevitably includes errors although DeepSpeech is one of the top engines in transcription accuracy. Therefore, it is important to provide a text edit function (Fig. 3E).

The textual narratives are processed to further extract semantic attributes for easy interaction. The system allows for keyword selection using TF (Term Frequency) or TF-IDF (Term Frequency-Inverse Document Frequency) weights as a guide. A text indexing structure supports the fast query response needed for interactive filtering.

The PSPNet and DeepSpeech tools are selected as being among the most reliable, and, importantly, they can run offline and on-device so that users do not have to rely on active remote servers during data exploration.

Visual community visualization system

Interface overview

The VisualCommunity interface is illustrated in Fig. 3 which includes a set of visualization functions:

- Upload data function (Fig. 3A) supports the upload of captured data from the GeoVideo app.
- Data management panel (Fig. 3B) allows users to manage these data and perform AI-based data processing. Users can load, select, remove, and process datasets organized as a list of trips.
- Map view (Fig. 3C) visualizes geo-trajectories where users can directly select an active trajectory. A location marker (i.e., blue color marker) on the active trajectory indicates the current location of a trajectory. Narrative-based insights linked with geographical structures are highlighted.
- Visual information view (Fig. 3D) shows the visual information at the current location in the form of video frames and their semantic objects. A video player allows for the user to play, pause, and drag the video to a specific location of a trip.
- Narrative editor (Fig. 3E) allows users to edit the transcribed speech narratives and correct the transcription errors.
- Comment editor (Fig. 3F) allows users to add extra comments and more description during data analysis.
- Narrative view (Fig. 3G) displays the description and insights, by location and video content, including both the transcribed narrative and the added comments.
- Keyword filter (Fig. 3H) presents the top keywords (in a list view or bubble view). Joint conditions are supported by Boolean operations on multiple keywords.
- Visual category view (Fig. 3I) visualizes the distribution of semantic categories (road, car, building, etc.) in the street scenes sourced from the spatial videos. A user can filter and extract critical images and their locations based on these categories.

All these functions are coordinated for interactive data exploration. Next, we introduce the visualization design and functions.
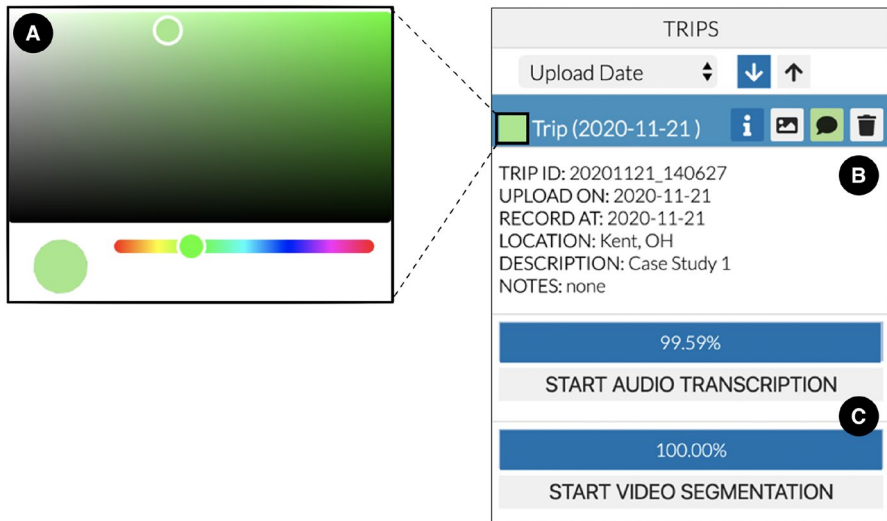
Fig. 4 Data list and processing: **A** Picking a specific trip color; **B** trip information such as ID, upload/record date, location and description; **C** AI-based processing with two modules: semantic segmentation and speech transcription

Data management and AI-based data processing

Users can load, select, and remove datasets which are organized as a list of trips (3B). For each trip, the users can start the semantic segmentation of images while also transcribing the audio narratives.

From the list, users can sort the data trips by their uploaded date, recorded date, or the size of narratives. Users can also select unique colors (Fig. 4A) for visualizing the trip trajectories on the map view.

Based on early prototype feedback, the system also allow users to browse the video and trajectory data before any AI-based data processing. This is important to perform an immediate quality control check and discard of problematic trips. As shown in Fig. 4C, users can start the AI-based processing of a selected trip, with progress visualized in percentages. It is also possible to pause and resume the segmentation and transcription processes at any time.

Map view and content visualization

Multiple trip trajectories are visualized on the map. A variety of map styles, such as streets, light, dark, outdoors, or satellite, can be chosen. One active trajectory is shown (in highlighted color) but it is possible to switch to any of the other inactive routes. On the active trajectory, a location marker indicates the current location of the images shown in the visual information view and the corresponding text in the narrative view. Users can drag the marker to change location.
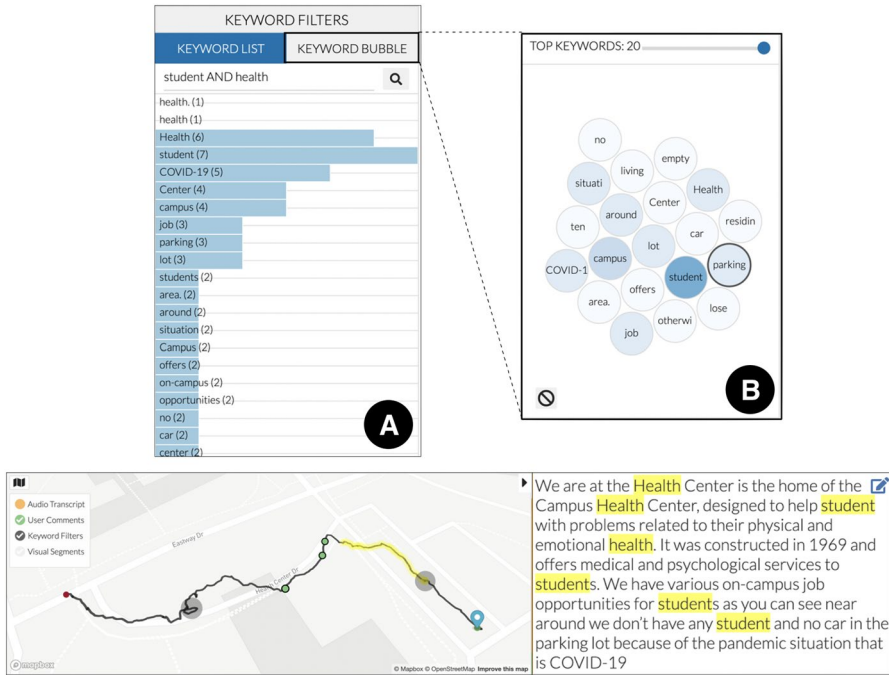
Fig. 5 Keyword filters: **A** Keyword lists are sorted by keyword frequency extracted from AI-based speech transcription and **B** keyword bubble contains a group of keywords colored by intensity of keyword frequency. All selected keywords are highlighted over trajectory and narrative

Four types of associated data contents can be highlighted over the trajectories displayed in different colors, including the narrative transcript, user comments, filtered keywords, and semantic category information. For example, in Fig. 3C, three locations (in red) indicate where users added comments, and two trajectory parts (in blue) show the geo-locations where narratives are available.

Visual information view and video player

The spatial video information of the active trip is shown in the visual information view (Fig. 3D). Two rows of images display the image frames corresponding to the current location. The frames of the exact location is highlighted with a black border, while a few frames prior and after are also shown to provide context. The top row presents the original images, while the bottom row shows the images with colorized visual categories (same as the colors in Fig. 3I). On the left of the two rows, the associated video plays with the ability to move through the frames by dragging.
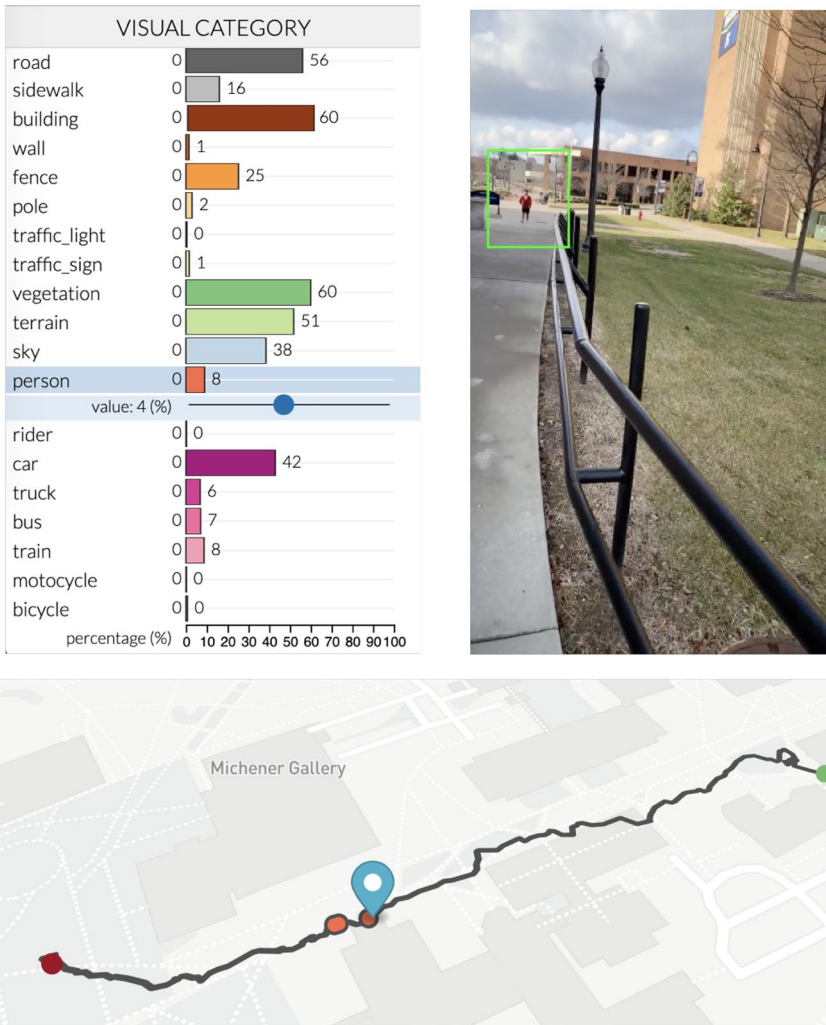
Fig. 6 Visual category based analysis based on semantic categories (e.g. road, building, traffic light, sky, etc.). The filtered values are displayed as a circle marker over the trajectory

Geo-narrative editing and comments

VisualCommunity allows a user to edit the text (Fig. 3E) at any time during analysis, with then option to listen to the original audio repeatedly if needed. It is also possible to add comments or third-party opinions at any location—for example it might be useful to emphasize that a place being described is clearly visible in the bottom of the image (Fig. 3F). The narratives and comments are shown in Fig. 3G. Users can also click on any item to change or remove it.

Keyword filter

The keyword filter (Fig. 3H) allows for the selection of keywords in the narratives or comments. At the same time the corresponding visual information and textural information are visualized, together with the locations on the map. This means that a researcher can use the three different media sources (image, text, map) to explore a location or scene in multiple non-linear ways. The filter uses two visualizations: a keyword list and a keyword bubble view, as shown in Fig. 5. In the keyword list (Fig. 5A), top keywords are shown and sorted by their term frequency in all texts. In the bubble view (Fig. 5B), top keywords are shown while the bubble color represents the frequency. Users can adjust the number of bubbles with a slider. Users can also directly input keywords in the text box. Hovering over each keyword will highlight those trajectory locations where the narrative text includes the keyword.

Users can also combine keywords which offer opportunities to explore locations and themes with semantics. In Fig. 5, using "student AND health" in the keyword filter creates highlighted parts on the map. Meanwhile, the corresponding keywords are highlighted in the narrative view as well.

Visual category view

At the current location, each semantic category of the video content, such as building, sidewalk, road, person, etc., is visualized as a bar with a specific color, as shown in Fig. 6. The length of the bar represents the maximum percentage of this category in any image from the active trip dataset. For example, Fig. 6 shows that "building" occupies a maximum of 54% in any street scene image and "sidewalk" has a maximum of 23%. Based on these hints, users can identify the visual content features. More importantly, it is possible to click on a category and use its percentage slider to define a specific threshold value. As shown in Fig. 6, users define a threshold of 30% for "building" meaning that only the locations with more than 30% buildings (based on the captured spatial video) are identified, as highlighted on the trajectory over the map (with red dots). Drilling down on these red dot locations reveals the corresponding images and narratives. It is also possible to drag the slider to adjust the threshold percentage to observe locations in real-time. In this way VisualCommunity supports data analysis based on the captured visual contents.

System implementation

For the visualization system, all the heterogeneous data are organized in local file systems with JSON formats. We avoid using a spatial database so as to reduce potential barriers to use based on the required installation and configuration expertise. The visualization system is implemented and distributed based on the Electron framework, in which D3.js and Mapbox libraries are employed. PSPNet and DeepSpeech are bundled inside, and GPU acceleration is automatically used when a GPU is available. The system can be installed and run on Windows, Linux, or Mac OS computers.
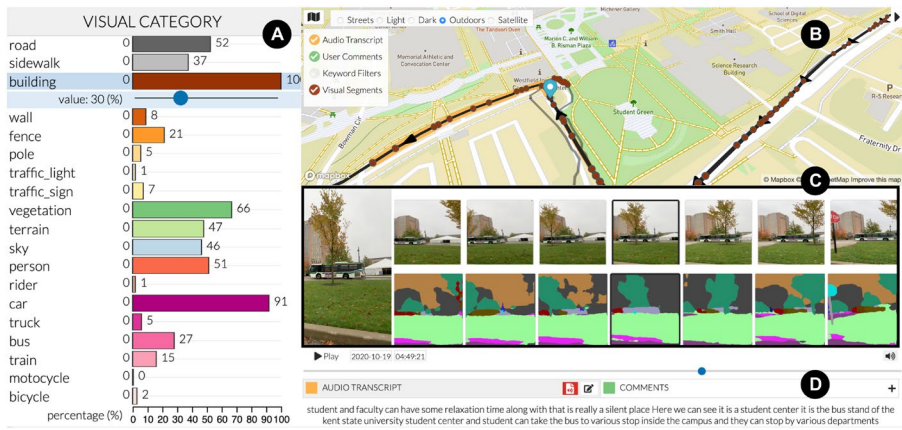
Fig. 7 Exploring virtual campus tours: **A** Finding the visual object "bus" with visual information view. **B** Studying related locations on the map. **C** Browsing landscape images and video. **D** Studying audio transcripts of campus information

Case study: using visualcommunity in COVID pandemic

In this section, we present a case study of using VisualCommunity system. This proof of concept test has been implemented during the COVID-19 pandemic. As a result of COVID-19 an university has imposed various health requirements and policies on staff, faculty and students, and on the physical campus environment. One outcome is the presence and movement of people on campus. VisualCommunity can be utilized to collect data and analyze the community dynamics and patterns, during different stages of the pandemic, which in turn can be used to further refine and improve campus policies. In an alternative application, the pandemic prevents campus visitors, such as prospective students and their families, from attending a traditional in-person campus tour. VisualCommunity can present a visual platform for them to tour the campus and discover information about the university.

The following two usage scenarios are presented: (1) campus tour guides record touring videos along with the typical narratives usually given during in-person tours. (2) university employees walk around the campus, capture spatial videos, and orally record their descriptions and findings related to the pandemic. These trips can be used to perform interactive campus data exploration based on the visual, semantic, and geospatial information. In this case study, the names and scenarios are fictitious to show the system utility.

Scenario 1: exploring virtual tours

Campus tour guides usually provide in-person tours for visitors such as high school students and their families. These visitors would like to "see" the campus, and acquire acquire information linked to campus places and life. Due to the pandemic, tours were halted but could be made virtual through the VisualCommunity framework. These virtual guided trips used GeoVideo to record videos and descriptions of

the campus. The virtual visitors can then use the visualization system to explore the campus. In this example, a (fictitious) visitor, Amy, conducts a virtual and interactive tour with the system.

As shown in Fig. 7A, Amy studied the trip datasets in the visual information view. Amy clicked on various categories to check the highlighted locations and their narratives recorded by the tour guides. In particular, Amy wanted to explore campus information with the visual object "building". By selecting this category, she set a threshold of 30% to extract those street scenes with more than 30% were buildings (identified by AI segmentation). A set of magenta points on the trajectories indicated where "building" was spotted. The same approach could be used to explore more topic categories such as "sidewalk" to assess campus walking paths, or as displayed in Fig. 7 "vegetation" to show how green the campus was.

Figure 7B displays one trip trajectory after Amy zoomed into the main street on campus where big buildings exist in the video frames. On the trajectory, one segment is highlighted in yellow showing where audio transcripts of the guides' narrative are available. Amy can read the transcript for more information (Fig. 7D) and check the street-view images as shown in Fig. 7C. She soon realizes the location is a bus station close to the student center. In Fig. 7C, Amy can drag these images, or directly drag the marker on the map, forward and backward to browse the environment surrounding this location. She can directly play the video as well. In such exploration, for example, Amy realizes that there are big tents around the student center. The transcript indicates that these tents are being used for COVID testing. For the tour guides who created this dataset, they could have used the same system to browse the data before sharing it with visitors, and add in comments about the tents that they might have missed during data capture, or provide updates if information has changed. Two advantages of this system as described are that it provides multiple options for the user to explore data on their own terms, even spatially, rather than simply viewing a more typical orientation video. Secondly, the example of the tents also shows how such a system can be used to archive temporal change, either over long durations (how a campus changes over the different phases of a multiple-year planning initiative), or short duration (how an environment changes in the days and weeks following a disaster - or in this case, a pandemic). A version of this second example will be explored next.

Scenario 2: studying campus situation

University staff used the GeoVideo app on walking trips across the university campus in order to archive the daily situation during the pandemic. The captured datasets included important campus locations, such as the health center, student recreation center, university library, student center, and more. These datasets were uploaded to the system and processed by the AI-tools. The data creators could also correct the errors in the narratives. After this, a (fictitious) university manager, Alice, explored the information in the visualization interface.

As shown in Fig. 8, Alice first selected one trip around the university's health center, whose trajectory was shown on the map (Fig. 8B). Alice explored the

Fig. 8 Studying campus situation during pandemic: **A** Searching interesting keywords. **B** Finding critical locations on the map. **C** Observing video snapshots. **D** Reading-related narratives. **E** Editing narratives and adding comments

related data in the geo-narratives, as shown in Fig. 8A. She first hovered over each top keyword to see where each was located. She then selected interesting keywords and combined them for searching. For instance, she used "student AND COVID AND campus" to find related locations, which were highlighted as the two yellow parts on the trajectory as shown in Fig. 8B. Alice could read the associated narrative text in which these words appeared. She clicked one paragraph on the narrative view (Fig. 8D). Meanwhile, the location marker on the map automatically moved to the corresponding location. This paragraph described that the health center continued to provide medical services but there had been relatively few students attending due to the pandemic. Figure 8C verifies this by displaying a video snapshot around the building in which there are no people or cars. The text provides further detail in that this might be partly due to the COVID test site

being located in the student center, instead of this health center. Few people came to the health center for other medical reasons. The high volume of cars around the student center might also provide justification for the university to move some testing capacity to the health center. Alice could add this idea / observation as a comment (Fig. 8E) with these comments being shared out as a report. What is key here is that the relative ease of data collection using a commonplace device like a smart phone could result in weekly, or even daily status updates.

Scenario 3: data transfer and processing performance

In this scenario, we use the example in (Fig. 3) to quantitatively show the effectiveness and latencies associated with the proposed solution.

First, GeoVideo APP can be used offline without any internet connection. This will eliminate the network latency and cost during the data capture process. For the captured video of about 30 min long, the file is about 2.1 GB which contains video, GPS and audio information. The file can be uploaded to VisualCommunity through the interface (Fig. 3). Here, the whole data transfer take approximately 5 min with a normal resident Internet speed.

The semantic segmentation and speech transcription process of this dataset are tested on a consumer desktop computer with an Intel i7-8700K CPU and 16 GB memory and an Nvidia GTX 1070 GPU with 8GB texture memory. With an implementation by TensorFlow, the semantic segmentation costs about 20 min and the speech transcription costs about 5 min. Here the processing time mostly depends on the duration of the recorded video (in this example, 30 min long). After all the data processing, the data exploration can be done over VisualCommunity interface. For example, when using Visual Category (Fig. 3I) to filter visual concepts related to keywords (e.g., "Car"), this task took about 2 min since the related visual and audio information need to be extracted from the dataset. The filtered visual elements are illustrated over the trajectory as circle markers. Dragging the location marker along the trajectory to each circle marker (Fig. 3C). Figure 3D shows a parking lot with cars inside. Here the dragging and visualization response cost less than 1 second. In the same way, users can perform interactive data exploration through the visualization software.

Deployment and feedback

VisualCommunity system has been publicized recently for test use. It is available at the website: http://vis.cs.kent.edu/VisualCommunity. It was employed by some domain scientists for different applications, such as investigating vegetation areas on sidewalks to analyze their walkability and the effects on different types of community residents (e.g., seniors, disabled); studying insights and suggestions from architects on community infrastructures, etc.

The preliminary feedback from the users was positive. First, the new Geo-Video app which is not available in existing tools (e.g., GeoVisuals) is convenient on mobile platforms. Second, the automatic tools can extract visual semantics from the video and transfer the speech to text. These were very practical challenges in their work. Third, the data exploration based on keywords is simpler in comparison to GeoVisuals. We also received a few limitations to be discussed in the next section. We will address the issues and further improve the software platform.

Discussion

There is a growing appreciation of multi-media mode, multi perspective, mixed data approach to investigating many of the challenges facing "communities" or other granular areas of interest. While spatially enhanced technologies, such as video cameras, have been used to good effect in both the United States and overseas, a common criticism for more widespread use is the lack of robust and user friendly software or platforms to fully leverage the insights contained within. In this paper we have addressed these needs through VisualCommunity, which has been developed while continually soliciting feedback from those same researchers and critics. More specifically, this new platform contributes to the following topics, while some limitations are to be addressed in future development:

Data exploration: VisualCommunity supports the visual exploration of data through being able to easily view video and the associated narrative. It also allows users to investigate these narratives in this space through both term (through key word) and image mining. We will further study visual storytelling techniques to report and promote the findings from the exploration.

Integration of AI functions in visualization system: The platform facilitates automatic processing of raw data through AI tools. The team also responded to a central directive of prototype evaluators that it is often not possible to rely on connecting to servers when in the field, or in certain locations. In response VisualCommunity utilizes free and off-line tools which are vitally important to the free visualization software. However, as AI tools (and the associated code) are continually developing, the platform has also been designed to be flexible and easily ingest such advances, and VisualCommunity can be adapted to customized versions that incorporate paid and online AI services if appropriate. For instance, a further advance can be achieved through the incorporation of new AI techniques to improve the quality of semantic segmentation (e.g., in bad weather) and voice transcription (e.g., for accent recognition and natural sentence detection). There is also a need to develop the theoretical frame that incorporates tools such as VisualCommunity with regards reliability, bias, and sustainability of AI, not only in terms software performance and utility, but also in how these affect society.

Software framework coupling data capture and exploration: While there was a need for better tools to investigate data being collected using existing technologies, a further frequently discussed limitation was when such equipment was not available. There was a need to be able to also capture the type of data that would feed into

VisualCommunity. This is now possible through the GeoVideo app function which in effect allows even a smart phone or tablet to become the only "extra" kit required.

A further need addressed by VisualCommunity is usability in any (including resource challenged) environments. While this point has been raised regarding the AI implementing, the same challenge faces software download/compilation, and data upload. In developing a practical visualization software, the software installation and functionalities, have equal importance to the visualization design itself. To further address this need future versions will incorporate the remote upload of GeoVideo data.

Conclusion

A computational and visualization platform, VisualCommunity, is developed to support the archiving and exploration of spatial multimedia data (spatial videos and geo-narratives) for small area (community/neighborhood/micro environmental) research and practice. It includes a mobile data capture tool and provides a cross-platform visual analytic tool on desktop and laptop computers. It allows social scientists, researchers, residents, and administrators to utilize the advances more fully in mobile technology currently available.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Rosenstein, B. (2002). Video use in social science research and program evaluation. *International Journal of Qualitative Methods, 1*(3), 1–38. https://doi.org/10.1177/160940690200100302
2. Mills, J. W., Curtis, A., Kennedy, B., Kennedy, S. W., & Edwards, J. D. (2010). Geospatial video for field data collection. *Applied Geography, 30*(4), 533–547.
3. Curtis, A., Blackburn, J. K., Smiley, S. L., Yen, M., Camilli, A., Alam, M. T., et al. (2016). Mapping to support fine scale epidemiological cholera investigations: A case study of spatial video in Haiti. *International Journal of Environmental Research and Public Health, 13*(186), 1–13.
4. Curtis, A., Blackburn, J. K., Widmer, J. M., Jr., & Morris, J. .G. . (2013). A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: Mapping health risks using spatial video in Haiti. *International Journal of Health Geographics, 12*(21), 1–15.
5. Curtis, A., Curtis, J. W., Kennedy, S. W., Kulkarni, A., & Auer, T. (2013). A methodology for assessing dynamic fine scale built environments and crime: A case study of the lower 9th ward after hurricane Katrina. *Crime modeling and mapping using geospatial technologies* (pp. 203–225). New York: Springer.
6. Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., & Zhao, Y. (2017). Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems, 18*(8), 2232–2249. https://doi.org/10.1109/TITS.2017.2683539

7. Miranda, F., Doraiswamy, H., Lage, M., Zhao, K., Gonçalves, B., Wilson, L., Hsieh, M., & Silva, C. T. (2017). Urban pulse: Capturing the rhythm of cities. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 791–800. https://doi.org/10.1109/TVCG.2016.2598585

8. Ertl, T., Chae, J., Maciejewski, R., Bosch, H., Thom, D., Jang, Y., & Ebert, D. S. (2012). Spatio-otemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Proceedings of the 2012 IEEE conference on visual analytics science and technology (VAST)*. VAST '12

9. Chen, M., Botchen, R. P., Hashim, R. R., Weiskopf, D., Ertl, T., & Thornton, I. M. (2006). Visual signatures in video visualization. *IEEE Transactions on Visualization and Computer Graphics, 12*(5), 1093–1100. https://doi.org/10.1109/TVCG.2006.194

10. Jamonnak, S., Zhao, Y., Curtis, A., Al-Dohuki, S., Ye, X., Kamw, F., & Yang, J. (2020). GeoVisuals: a visual analytics approach to leverage the potential of spatial videos and associated geonarratives. *International Journal of Geographical Information Science, 34*(11), 2115–2135. https://doi.org/10.1080/13658816.2020.1737700

11. Bhati, D., Amiruzzaman, M., Jamonnak, S., & Zhao, Y. (2022). Interactive visualization and capture of geo-coded multimedia data on mobile devices. In J.-H. Kim, M. Singh, J. Khan, U. S. Tiwary, M. Sur, & D. Singh (Eds.), *Intelligent human computer interaction* (pp. 260–271). Cham: Springer.

12. Amiruzzaman, M., Curtis, A., Zhao, Y., Jamonnak, S., & Ye, X. (2021). Classifying crime places by neighborhood visual appearance and police geonarratives: A machine learning approach. *Journal of Computational Social Science, 4*(2), 813–837.

13. Curtis, A., Curtis, J. W., Shook, E., Smith, S., Jefferis, E., Porter, L., Schuch, L., Felix, C., & Kerndt, P. R. (2015). Spatial video geonarratives and health: Case studies in post-disaster recovery, crime, mosquito control and tuberculosis in the homeless. *International Journal of Health Geographics, 14*(1), 22. https://doi.org/10.1186/s12942-015-0014-8

14. Krystosik, A. R., Curtis, A., Buritica, P., Ajayakumar, J., Squires, R., Dávalos, D., Pacheco, R., Bhatta, M. P., & James, M. A. (2017). Community context and sub-neighborhood scale detail to explain dengue, chikungunya and zika patterns in Cali, Colombia. *PLOS ONE, 12*(8), 1–25. https://doi.org/10.1371/journal.pone.0181208

15. Curtis, A., Felix, C., Mitchell, S., Ajayakumar, J., & Kerndt, P. R. (2018). Contextualizing overdoses in Los Angeles's skid row between 2014 and 2016 by leveraging the spatial knowledge of the marginalized as a resource. *Annals of the American Association of Geographers, 108*(6), 1521–1536. https://doi.org/10.1080/24694452.2018.1471386

16. Vincent, L. (2007). Taking online maps down to street level. *Computer, 40*(12), 118–120. https://doi.org/10.1109/MC.2007.442

17. Zhai, W., & Peng, Z.-R. (2020). Damage assessment using google street view: Evidence from hurricane Michael in Mexico beach, Florida. *Applied Geography, 123*, 102252.

18. Nguyen, Q. C., Khanna, S., Dwivedi, P., Huang, D., Huang, Y., Tasdizen, T., Brunisholz, K. D., Li, F., Gorman, W., Nguyen, T. T., et al. (2019). Using google street view to examine associations between built environment characteristics and us health outcomes. *Preventive Medicine Reports, 14*, 100859.

19. Berland, A., & Lange, D. A. (2017). Google street view shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening, 21*, 11–15.

20. Kronkvist, K. (2013). *Systematic social observation of physical in inner-city urban neighborhoods through Google Street View: The correlation between virtually observed physical disorder, self-reported disorder and victimization of property crimes*. Faculty of Health and Society.

21. Vandeviver, C. (2014). Applying google maps and google street view in criminological research. *Crime Science, 3*(1), 13. https://doi.org/10.1186/s40163-014-0013-2

22. Olea, P. P., & Mateo-Tomás, P. (2013). Assessing species habitat using google street view: A case study of cliff-nesting vultures. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0054582

23. Shen, Q., Zeng, W., Ye, Y., Arisona, S. M., Schubiger, S., Burkhard, R., & Qu, H. (2017). StreetVizor: Visual exploration of human-scale urban forms based on street views. *IEEE Transactions on Visualization and Computer Graphics, 24*(1), 1004–1013. https://doi.org/10.1109/tvcg.2017.2744159

24. Li, L., Tompkin, J., Michalatos, P., & Pfister, H. (2017). Hierarchical visual feature analysis for city street view datasets. Visual analytics for deep learning, VIS2017.

25. Zeng, W., & Ye, Y. (2018). VitalVizor: A visual analytics system for studying urban vitality. *IEEE Computer Graphics and Applications, 38*(5), 38–53. https://doi.org/10.1109/MCG.2018.053491730

26. Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization, 12*(1), 3–24.
27. Guo, B., Liang, Y., Yu, Z., Li, M., & Zhou, X. (2016). From mobile phone sensing to human geo-social behavior understanding. *Computational Intelligence, 32*(2), 240–258.
28. Stella, E. (2018). An educational experience in virtual and augmented reality to raise awareness about space debris. PhD thesis, Politecnico di Milano.
29. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881–2890).
30. Chen, Y., Li, W., & Van Gool, L. (2018). Road: Reality oriented adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7892–7901).
31. Mozilla, M. L. (2020). Deep Speech, an open source embedded speech-to-text engine.