UFO_Tracker: Visualizing UFO sightings

Vinh T. Nguyen
Computer Science Department
Texas Tech University
TX 79415, USA
vinh.nguyen@ttu.edu

Vung Pham
Computer Science Department
Texas Tech University
TX 79415, USA
vung.pham@ttu.edu

Tommy Dang
Computer Science Department
Texas Tech University
TX 79415, USA
tommy.dang@ttu.edu

***Abstract*—Visualizing and analyzing geospatial and temporal observations are common tasks for many application domains. In this paper, we introduce *UFO_Tracker*, a visual analytic tool for analyzing unidentified flying object sightings from the National UFO Reporting Center. The goal here is to give the user a higher level view of where different types of sightings occur, to investigate whether sightings are increasing or decreasing over time, to discover the connections between different events which might happen at different geographic areas, and to quickly identify typical incidents at a given period of time without reading the whole sightings through topic modelling. Multiple visualization and data mining techniques are combined to make sense the increasingly large UFO reports which get updated hourly. The usefulness of the application is evaluated through a case study where anon-expert in ufology can find some typical interesting sightings. Our application can also be able to detect some misleading events such as missile launch or fireworks on a specific day through keywords and topic extraction. One limitation of our application is the data which is not up-to-date when new sightings are posted since the application pulled and processed data locally. Our initial application targets UFO sighting reports. However, we believe our approach has wider applications in other research domains, such as analyzing text corpus obtained from social media.**

***Index Terms*—Unidentified Flying Objects, geospatial temporal visualizations, dot plots, topic modelings, word clouds.**

I. INTRODUCTION

Unidentified Flying Objects (UFOs) are commonly described as anything that can be seen in the sky but can not be identified as a known object. The story of UFOs began in 1947 in the US [1] when a business man Kenneth Arnold [2] saw a fast moving, glistening objects from the sky on his plane near Mount Rainier in Washington. Inspired by this phenomenon, the National UFO Reporting Center (NUFORC) [3], founded by Robert J. Gribble, has been recording this phenomenon (called UFOs sightings) since 1974 based on user reports. Currently, there are over 100,000 UFOs sightings that have been recorded. Despite over 70 years of observations and 40 years of data collections, the reported evidences are still uncorroborated. Therefore, the rumors of UFO sightings are still interesting over the years: Does the UFO really exist?

Observing UFO occurrences and the rumors about them are valuable for reporting but synthesizing what people are talking about UFOs at specific locations is highly desirable. There is a need to have such a graphical tool that allows scientists to visualize, analyze and get insight of UFOs in a certain geographic area to reveal some of the interesting questions such as: Where are the most popular places where UFO sightings often occur? Which events are mostly spotted by viewers? Is there any correlation between different phenomenon? What is the most highlighted event in a given period of time? or is there any misleading sightings of UFO?

We are currently lack tools that allow users to get more insight about the data in the literature. Existing techniques and application on the internet so far help users answer only portion of the curious question. In response, we present the design of *UFO_Tracker*, a visualization application of UFO sightings. Our system is motivated by the literature concerns. We attempts to answer these above questions that have not revealed completely until now. The main contributions of this paper are:

- We provide an interactive data analytics tool for visualizing and analyzing a large number of UFOs sightings distributed in certain geographic areas.
- We integrate multiple visualization and data mining techniques, such as parallel coordinates, dot plots, word clouds, and k-means for detecting top key words concerning UFO reports and the correlations between them. Though the resulting application leverages existing techniques, the requirements for designing end up with a number of novel customization. This helps UFO scientist to quickly synthesize a large amount of information on UFO sightings.
- We demonstrate our application on the U.S. national UFO reporting center dataset, which has 25, 559 sightings from 2001 to 2010. We also describe interesting UFO reports in the data.

The rest of this paper is organized as follows: We first summarize existing techniques on geospatial temporal visualization and share relevant finding from UFO sighting applications. Then provide an overview of visualization tasks and describe the design of the *UFO_Tracker* in detail. We discuss our implementation and availability of our tool in Section IV. Finally,we describe an informal study of the systems usage and conclude our paper with future plans.

II. RELATED WORK

This section does not intend to survey all visualization tools for geospatial temporal visualizations [4]. Instead, we discuss some related tools.

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

A. Geospatial temporal visualizations

Visualization of geospatial temporal data has a rich history. One of the earliest spatial temporal illustration is the Napoleon's march towards Moscow as depicted in Figure 1. The illustration uses position, color, size, and text annotations to captures different states of the Napoleon's army during the Russian campaign in 1812 [5] within a single snapshot.

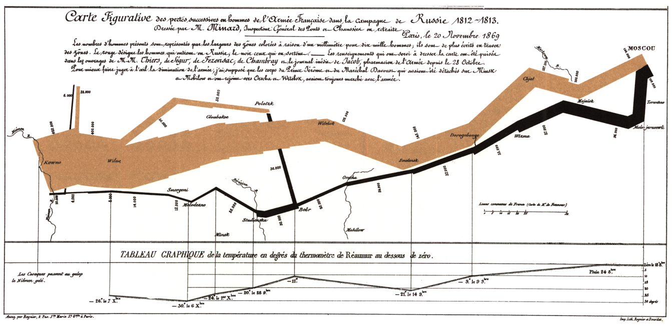


Fig. 1. A famous spatial temporal illustration of time flattening [5].

A simple approach to visualize temporal data on a 2D map is to attach the time series graphs directly on top of each geolocation [6]. This easily becomes too clustered for larger time series, especially when geolocations are not equally distributed. Showing summary statistics [7], such as average, standard deviation, or trends [8] on a choropleth map resolves this concern. However, this comes back to the trade-off between details and simplifications.

Shneiderman [9] designed task taxonomies based on seven data types and tasks are divided into seven categories (overview, zoom, filter, details-on-demand, relate, history, and extract). Using Shneiderman's metaphor, CrimeViz [10] combines with mashup techniques to integrate additional visual elements on top of google map for crime analysis. Similarly, Ramakrishna *et al.* [11] present another software mashup, which can handle a large spatio-temporal data set with 2.5 million records using hexagon binning.

B. UFOs sightings visualizations

There have been various available tools for visualizing the occurrences of UFOs sightings over the last decades. Sam Monfort [12] creates an online tool for visualization from the NUFORC data. This graphical tool renders aggregated data since 1905 through line chart and heat map that allow users to see the distribution of UFO spotted throughout the world. But this tool is limited to static images and users have no other option to view the detail of the specific event.

Mutual UFO Network (MUFO) [13], located in CA, is the oldest non-profit organization that investigate the reported UFO sightings, its tool, UFO Stalker - a live UFO map, uses real time Google Maps and overlay the the sightings on top of it, a heat map is introduced to show the density of sightings in the same location, each sighting event is annotated by a shape based on the users' report. Besides from over-viewing the distribution, Stalker also shows the number of the events increasing or decreasing compared to the previous weeks, months and years. Filtering options in a wide range of criteria allows viewers to narrow down the desired results.

Max Galka [14] also uses Google Map for UFO sighting distribution, sightings that occur on the same night within about 50 miles of each other will be grouped together into a circle, the size of each circle indicates the number of the events that it contains. Sighting reports are shown from different perspectives such as by population, near airport or military bases. However, lack of necessary interactions and and get insight into data make this application hard for detail analytics.

John Nelson [15] provides a per-capita approach for UFO sightings views. This tool shows the overview of the ratio of the sightings by population, a bi-variate mapping of sightings in the color dimension, and population density in the opacity dimension. The most interesting part of this paper is the trending area where different shapes of the sightings are compared to each other in terms of time. Another interesting feature of this application is that non-geographic data is plotted into a cell chart (horizontal axis is the time of day and night while the vertical axis splits up the months of the year) and other bar charts embedded to it. This allows users to see data from multiple perspectives.

A more comprehensive approach is created by Francisco [16], he collects data from different sources from all over the world with more than 176,000 records. Aggregate data map allows users to see the statistical results by continent or by country, users are also be able to search the events based on time scale and desire locations. A heat map is overlaid on top of the Google map, letting users spot the number of sighting occurrences in the specific areas. Although this site both collects and combines real time data, its visualization tool is limited into some of the very basic functions such as viewing and searching. Drilling and brushing are the two other functions for users that need to be implemented in this application.

Darrent et al. [17] use Python programming language to create a visualization report of UFO sightings distribution in the US to understand the correlation between demographic parameters and the quality of UFO sightings. The input data is also obtained from the NUFORC. Each sighting is represented by one dot and is colored according to the shape. Darrent uses statistical methods to explain some trends and patterns of the sighting data. His reporting results is based on the hypothesis of the increasing number of internet connections, working hours of the day, seasons and location of the events. However, too many colors for sightings makes it difficult to distinguish between each events, especially where there many sightings reported in the same or very close locations because of overlapping. this interesting report shows only some general patterns and lack the level of details in each event.

Despite this prior work on UFO sightings, none of these available tools are able to synthesize information and allow users to get insight into data upon users' request.

Unlike other applications,in this paper, we provide an interactive data analytics tool for visualizing and analyzing a large number of UFOs sightings which are distributed on certain geographic areas and integrate parallel coordinate along with topic modelling for detecting highlighted sightings based on users' report.

III. *UFO_Tracker* VISUALIZATIONS

Our goal with *UFO_Tracker* is to build a visual analytics tool that gives the user a higher level view of where different types of sightings occur, investigates whether sightings are increasing or decreasing over time [7], and discovers the correlations between different events. This tool also supports filtering the popular terms of reports across regions based on user selections. Finally, it allows analysts to compare report cases across different areas and over time. These goals position *UFO_Tracker* different from other UFO sightings applications used as analysis tool by both expert and non-expert in ufology. Overall, there are three main components in our system: Processing input dataset, visualization components and user interactions.

This section explains *UFO_Tracker* components in detail.

1) **Processing input datasets:** We obtain the data from the NUFORC [3]. Then we calculate the distribution of UFO sighting organized by years, dates in a month, and hours in a day. (see Section III-A)
2) **Visualization components:** The visualization contains five main components: the control panel, the Google map, the time series sliders, the text cloud, and the parallel coordinates. (see Section III-B)
3) **User interactions:** This part summarizes all possible user interactions using *UFO_Tracker*. (see Section III-C)

Keim *et al.* [18] suggest three step processes for information visualization: overview first, zoom and filter, and then details-on-demand. Moreover, information visualization systems should allow users to perform analysis tasks that largely capture people's activities while employing visual analytic tools for understanding data [19], [20]. One of the major challenges when designing an application is the engagement of domain scientists due to the limited accessibility to well-trained study participants and the difference in participants prior knowledge. Therefore, we divide visualization tasks into different categories so that novice users will be able to participate in the empirical studies while ensuring that the results are also applicable to ufology domain experts.

Thus, the *UFO_Tracker* implements five low-level visualization tasks,consisting of mouse clicks and movement, based on the interesting research questions posed in the Section 1.

- **T1**: Display overview sighting distributions.
- **T2**: Retrieve and present sighting details on demand.
- **T3**: Cluster sightings based on their geolocations.
- **T4**: Show trends or patterns of sightings over time.
- **T5**: Filter and sort sightings by user interests [21].
- **T6**: Find the correlations between different sightings.
- **T7**: Detect anomalies or suspicious sightings.

A. Processing input datasets

We will demonstrate *UFO_Tracker* on UFO sighting data from the NUFORC [3] in the contiguous US from January 2001 through December 2010, which is about 35 MB of data (with 25,559 sighting reports) in the form of TSV file. We also try to find the correlations between UFO sightings with the airport locations available at [22], military bases available at [23], and population density available at [24].

B. UFO_Tracker overview

The input data is expanded in the following dimensions: time, geolocation, and text. Exploring the connections between theses dimensions is challenging. In this application, we apply multiple visualization techniques to highlight individual dimensions as well as the relationship between them. Figure 2 shows the main GUI of our visual analytic tool which is implemented Processing [25] and Java. Box A contains the control panel (or list of options provided to users). Box B displays the Google map, where each UFO report is displayed as a circle. Google-map navigation capabilities, such as zooming in/out and panning are supported. Box C shows the top keywords extracted from user descriptions of the UF0 reports selected on the map. Finally, Box D contains the time series filters (range sliders).

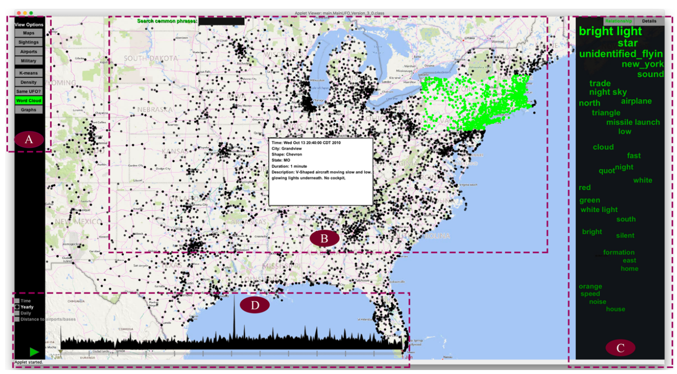


Fig. 2. Main interface of *UFO_Tracker* visualization: a) Control panel b) Google map c) Text cloud d) Time series slider.

The Google map We use Googe API to retrieve real time mapping of the world. Then UFO sighting data will be overlaid on top of the map. Within this map, each black circle represents an UFO sighting (visualization task **T1**). Mousing over the circles, users can view the report details (on a pop-up window), such as timestamp of the sighting, city, state, shape, duration, and description (as shown in Box B of Figure 2). This meets the visualization task **T2** requirement.

The heat map When there are too many sightings on the map, overlapping is unavoidable. The heat map represents individual values contained in the same area as colors. Users have an option to switch to the density (heat) map. In particular, we apply a kernel smooth density estimation in order to produce an high-level overview of the sighting distributions (visualization task **T1**). Brighter color illustrates areas with more sightings. As depicted in Figure 3, there are many UFO sightings reported in San Diego, located on the Pacific coast of California.

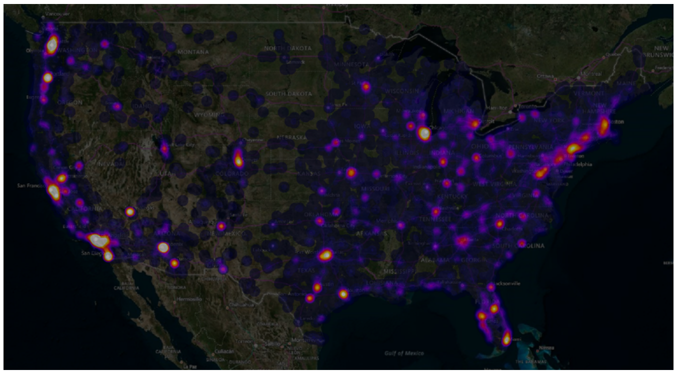


Fig. 3. Kernel density heat map in *UFO_Tracker*.

The airport and military bases Airplanes and military experiments are the most mistaken for UFOs [26]. All US airports and military bases can be plotted on the map on demand so that users can visualize and analyze the correlations of them and UFO sightings. Figure 4 shows only sightings reported near airports and military bases. From the distribution information of the UFO sightings in the heat map combined with this figure, it is shown that most of the reports are near airports and military bases and *UFO_Tracker* is able to detect the misleading sightings in the use case section.

Fig. 4. UFO sightings within 50 miles to the nearest airport and military bases.

The k-mean clustering *UFO_Tracker* clusters UFO sightings based on their geographic locations using k-means [27] (visualization task **T3**). This simple unsupervised learning algorithm will partition the incidents into k clusters based on the nearest mean. Since clustering based on geolocations is straight forward, k-mean algorithm is chosen because of its simplicity, fast and efficient in terms of computational cost, and easy to implement. Figure 5 shows an example of k-means clustering for 20 clusters (encoded in colors). The white circles are centers of these 20 clusters. As depicted, more clusters are formed on the East half of US.

The time series graphs Time series visualizations have been commonly used to show chronological distributions of geospatial data in which time is represented in horizontal

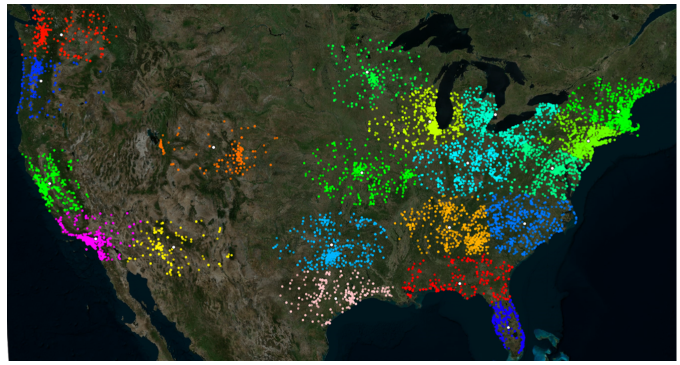


Fig. 5. K-means clusters of UFO sightings based on their geographical locations.

axis and a statistical variable in vertical axis [28]–[32]. In *UFO_Tracker*, we use area graphs to show the sighting frequency over years, days in a year, and hours in a day. These graphs show patterns and trends of the number of user reports over time (visualization task **T4**).

A range slider is implemented for each distribution graph which allows users to easily narrow down the time interval of interests (visualization task **T5**). The selection within the slider is set to white while the unselected area is dark grey. This will let users easily distinguish the difference between the two areas. In Figure 6(a), we filter the sightings which are at least 50 miles from the nearest airport/military base, from 9pm to midnight, and from September 2005 to July 2009. In the top (time) filter, we create double-layer sliders. The lower slider shows monthly frequency of sightings. The upper two sliders show daily frequency of September 2005 (yellow, on the left) and July 2009 (red, on the right).

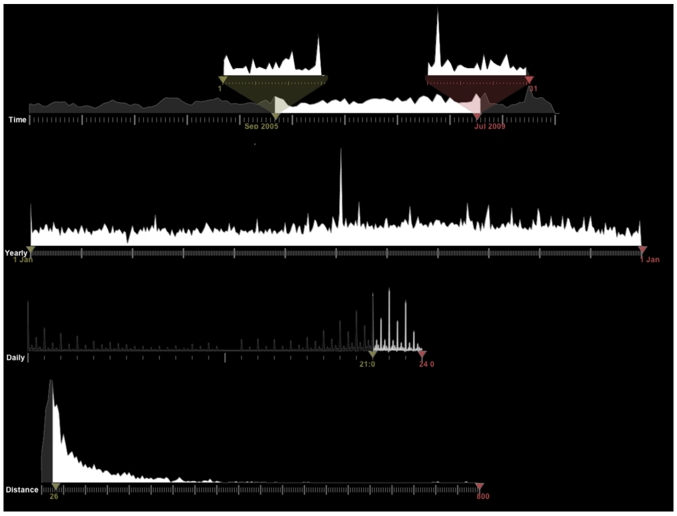
As depicted in Figure 6(b), most of sightings are reported in rural areas close to the big US highways, as an indication that these UFOs are probably sighted and reported by truck drivers (at night). White bands are the filtered areas.

Besides filtering reports by time intervals, *UFO_Tracker* also supports three other types of filtering:

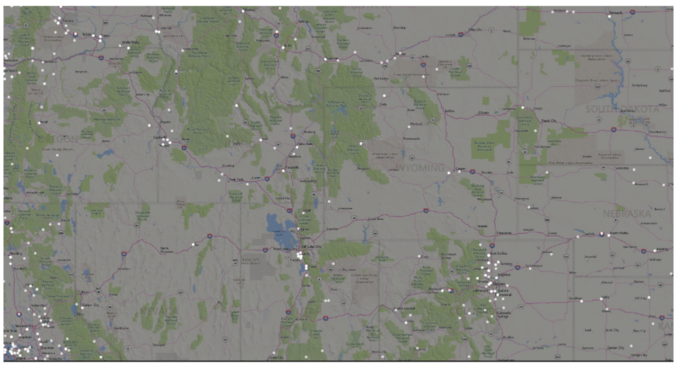
- By inputting term into a search box, such as 'aliens' or 'abduction', relevant sighting (contain the search terms in their report descriptions) are highlighted.
- By rectangle selection on the map using mouse drags.
- By nearest sightings: if there are other sightings close in locations and times.

Further text analytic inspections on the selected sighting are described next. As our color convention in this paper, white is used for available sightings on the map while green highlights selected sightings based on one of the three approaches listed above.

The text analytics: We are usually curious in not only the sighting locations, times, or shapes of the sightings but also what people said about the scenes they observed. Our text analytics window provides users a summary of topics in extracted from the large corpus of sighting descriptions. Two options for text visualization include popular terms and



(a)



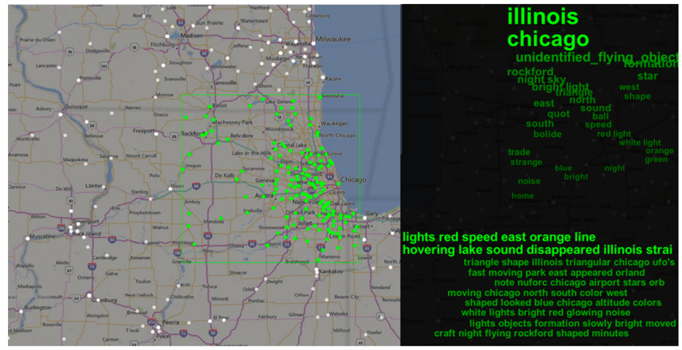
(b)

Fig. 6. The time series graphs show the trends of sightings. Data can be filtered by moving the range sliders along the charts. White bands in the middle of the range sliders are the filtered areas.

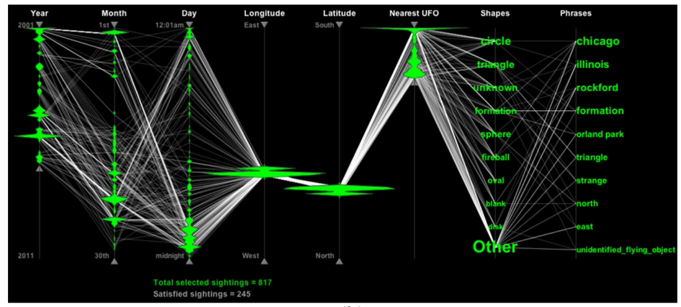
popular topics. Popular topics are extracted using Latent Dirichlet Allocation (LDA) [33] model, the most common topic model currently in use, along with Java-based topic modeling package [35]. The output of LDA method will give a bag of words where the words in the same bag that frequently occur together. In our tool, we limit the number of topics to 10 since increasing the number of latent topics will lead to more junk.

The right panel of Figure 7(a) shows the top 30 keywords (top right) and the top 10 topics (bottom right) for the selected sightings in Chicago land using rectangle selection (the green box). The more frequent keywords/topics are larger, brighter, and appearing on top the word/topic clouds. The color of these words/topics is set to the same color with the selected events in the rectangle region. Each LDA topic contains the 6 most frequent terms (not a complete sentence). When users mouse over a keyword/topic, the subset of (green) sightings containing this keyword/topic is highlighted on the Google map (visualization task **T6**). Users can now explore correlations between sightings in a customized parallel coordinates.

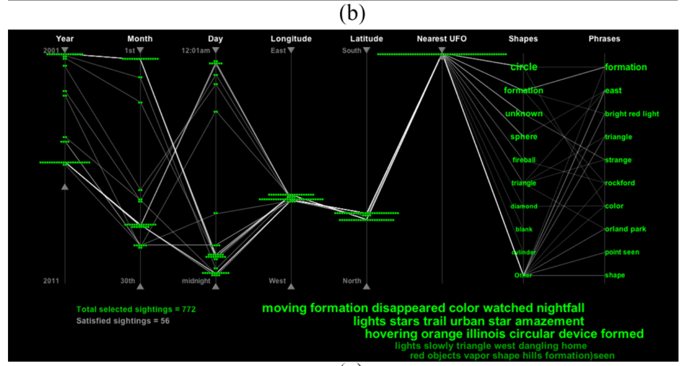
Parallel coordinates: This view categorizes sight features



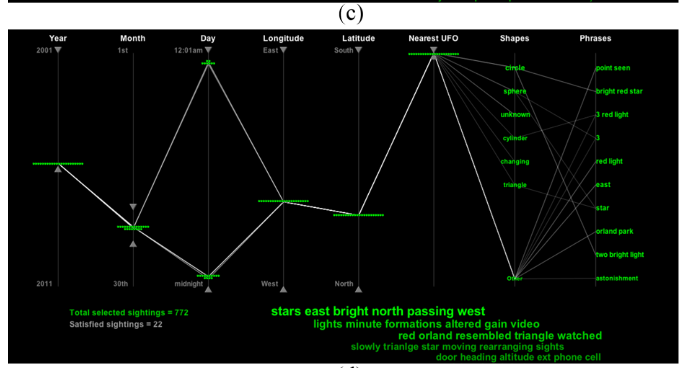
(a)



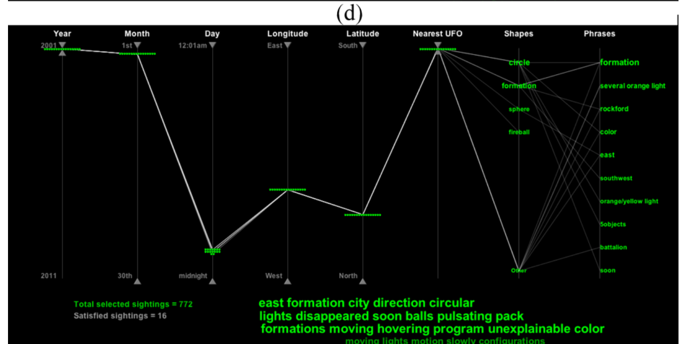
(b)



(c)



(d)



(e)

Fig. 7. Visualizing texts in *UFO_Tracker*: (a) The most frequent keywords/topics (extracted from LDA algorithm [33]) (b) parallel coordinates (c) dot plots [34] on each dimension (d) UFO sightings at Rockford, Illinois in January 2001 (e) UFO sightings at Tinley park, Illinois in October 2004.

into parallel coordinates which is often used for visualizing high-dimensional geometry and analyzing multivariate data. As shown in Figure 7(b), the distributions of selected sightings are displayed in smooth area graphs along the coordinates.

Range sliders are available on each dimension to allow users to narrow down their search. In Figure 7(c), we apply filtering on the “Nearest UFO” coordinate. In other words, we want to highlight only UFO scenes which are sighted and reported by different people in the same area and at almost the same time (visualization task **T6**). The time and space dimensions have been standardized and aggregated into a single score, called “Nearest UFO” for each UFO sighting. Notice that the area charts are now replaced by dot plots [34] (every dot is a UFO report). This option is automatically enabled to display a smaller number of selected sightings (less than 100). Consequently, brushing and linking can be done via individual dots (sightings).

Also within this parallel coordinates view, each sighting is represented as a white polyline with lower opacity. Multiple overlapped (similar) sightings create a strong profile which might come from the same UFO scene. Therefore, users can easily spot the correlations between the selected sightings (visualization task **T7**). As depicted in Figure 7(c), we can see two strong cases of UFO sightings: one in January 2001 and another one in October 2004.

Users can narrow down these events individually (using sliders along each dimension) as depicted in Figure 7(d) and (e). The five most important topics extracted from the UFO scene descriptions are also plotted at the bottom. The stories about these UFO sightings in Rockford and Tinley park, Illinois can be found and verified online [36].

C. User interactions

Interactions on mouse over, mouse clicks, and mouse wheels are provided in *UFO_Tracker*. There are nine options on the left that allows users to view sightings with different aspects as shown in Figure 2 (a). Google map views are available in Aerial, Hybrid, and Road. **Airports** option toggles airport locations on the map. **Military** option toggle all military base positions. **K-means** button clusters UFO sightings based on similarly geographical locations as depicted in Figure 5. **Density** renders density sightings in an area by colours. Figure 3 illustrates the density of sightings through heat map. **Same UFO?** gives information about similar observed sightings spanning throughout the nation. **Word Cloud** [37] option allows users to select multiple sightings and text analytics window highlights the top keywords/topics corresponding to the chosen sightings. **Graphs** option displays the correlations between the number of sightings to different attributes such as population density, distance from airports, distance from military area, year, month, day, and hour.

Details: This option provides detailed information about the selected sightings such as time of occurrence, city, shape and description. It also highlights keywords corresponding to each sighting as shown in Figure 8(b).

Brushing and Linking: In our application, we use brushing and linking technique to coordinate the contents of the Google map with the data shown in the Text cloud window. This is the main mechanism in the outer layout of the screen to help users discover sighting locations and grasp the popular topics/terms quickly.

Filtering: *UFO_Tracker* provides four basic filtering interactions based on the distance, day, year, and time of the events.

IV. Implementation

UFO_Tracker is implemented in Processing [25] and Java. The open source code, video, and project documentation are available on our GitHub repository at [38].

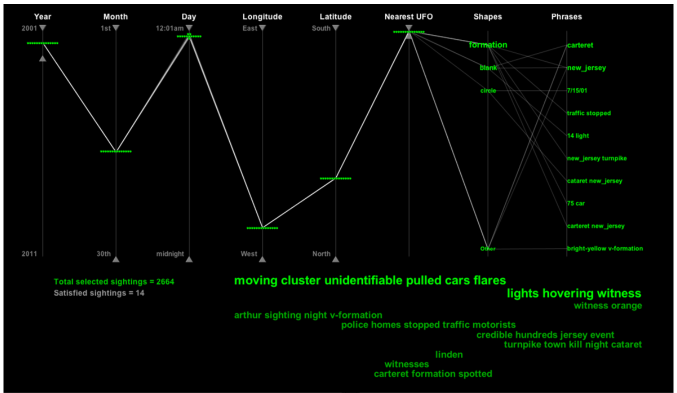
V. Case study

To demonstrate the usefulness and effectiveness of the *UFO_Tracker* visualization. We conducted a case study with a user who is not involved in the development process of this application. The expected results the ability to detect patterns, trends, anomalies of the UFO sighting events. We observed a non-expert in ufology to use this tool to explore sighting data and observed his actions. Starting from main screen view, the user first explored the view options to familiar himself with different graph layouts. Then he played with some data on the Google map, such as zooming, panning, and selecting data. He then read detailed information about specific sighting by mouse over some sightings in New York, Washington, California, and rare cases in the Dakotas. The user looked at the time series graph and moved sliders to left/right directions to see how the distribution of sightings changes. The user found that many UFO sightings were reported around 10pm to 1am, on Saturday and Sunday, and in July.

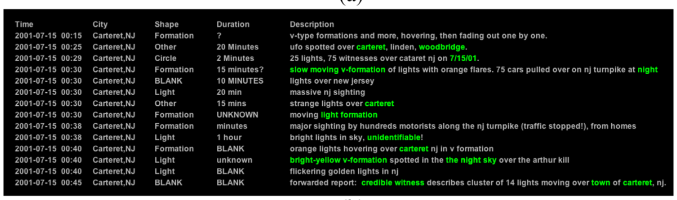
He turned his attention to the word cloud. He selected different areas on Google map to see which keywords were popular in each area. Then he further investigated sighting details by clicking on the relationship option to see the parallel coordinates. From there he noticed that some profiles are very dense (thicker white polylines), so he was curious to know more about this phenomenon by using the range sliders. Based on the phrases or top keywords (carteret, new_jersey, 7/15/01, traffic stopped, 14 light, new_jersey turnpike, carteret new_jersey, 75 car, bright-yellow v-formation) along with the shapes information as depicted in Figure 8(a).

He discovered that these keywords can be generally interpreted as “There were 14 bright-yellow lights on the sky, mostly in unknown shape, at Turnpike road, Carteret, New Jersey on 15 July 2001. All traffic stopped, around 75 cars at this location to witness this event”. In order to validate the inferred sentence, he made a Google search to find this sighting events (as depicted in Figure 9). It was very interesting that this inferred sentence not only provides accurate information but also gives additional detail about the event [39].

Another similar story can be found in Minneapolis on 19 August 2004 where the top keywords can be rephrased as “a video of the bright object was taken in the Minneapolis metro area on 19 August 2004”. This



(a)



Time	City	Shape	Duration	Description
2001-07-15 00:15	Carteret,NJ	Formation	?	v-type formations and more, hovering, then fading out one by one.
2001-07-15 00:25	Carteret,NJ	Other	20 Minutes	ufo spotted over carteret, linden, woodbridge.
2001-07-15 00:29	Carteret,NJ	Circle	2 Minutes	25 lights, 75 witnesses over cataret nj on 7/15/01.
2001-07-15 00:30	Carteret,NJ	Formation	15 minutes?	slow moving v-formation of lights with orange flares. 75 cars pulled over on nj turnpike at night
2001-07-15 00:30	Carteret,NJ	BLANK	10 MINUTES	lights over new jersey
2001-07-15 00:30	Carteret,NJ	Light	20 min	massive nj sighting
2001-07-15 00:30	Carteret,NJ	Other	15 mins	strange lights over carteret
2001-07-15 00:30	Carteret,NJ	Formation	UNKNOWN	moving light formation
2001-07-15 00:38	Carteret,NJ	Formation	minutes	major sighting by hundreds motorists along the nj turnpike (traffic stopped!), from homes
2001-07-15 00:38	Carteret,NJ	Light	1 hour	bright lights in sky, unidentifiable!
2001-07-15 00:40	Carteret,NJ	Formation	BLANK	orange lights hovering over carteret nj in v formation
2001-07-15 00:40	Carteret,NJ	Light	unknown	bright-yellow v-formation spotted in the the night sky over the arthur kill
2001-07-15 00:40	Carteret,NJ	Light	BLANK	flickering golden lights in nj
2001-07-15 00:45	Carteret,NJ	BLANK	BLANK	forwarded report: credible witness describes cluster of 14 lights moving over town of carteret, nj.

(b)

Fig. 8. Filtering on parallel coordinate to uncover the UFO scenes happened at 12:15 AM, 15 July 2001 at Turnpike road, Carteret, New Jersey: (a) There were 14 reports (green dots on each dimension) recorded for this event (b) The details of these reports are displayed on demand.

Fig. 9. A YouTube screen-shot of the UFO formation on July 15, 2001 at Turnpike road, Carteret, New Jersey [40]. This picture is captured when the user tried to find relevant resources on the internet based on the finding in Figure 8(a).

special sighting can also be easily found on YouTube at https://www.youtube.com/watch?v=O0S9tN3BrbY.

It is also very interesting that *UFO_Tracker* also allows user to detect misleading UFO sightings due to special event. For example, through the thickness of the link in Parallel coordinate graph, in September 19, 2009, there are six UFO cases reported around Pennsylvania state. However, top index terms indicate that this is a missile launch experiment.

Besides useful functions of the application, there are some limitations are founded during a case study that we need to improve. First, recommended by user, the application should support to automated detect anomalies and highlight them on the screen as a widget. Second, data should be updated in real time so that users are able to catch up with the latest events. And third is the topic modeling, semantic content extraction from several events should be applied so that users can understand the whole context without guessing from ”bag of words”.

Due to the limited space of the paper, we advise viewers to see more UFO stories and findings using *UFO_Tracker* on our GitHub repository available at [38].

VI. Conclusion and Future Work

Investigating UFOs sightings is an ongoing challenge and ultimately an interesting topic for researchers for discovering the myth of the universe. In this paper, we introduced *UFO_Tracker* as a graphical tool that allows researchers, especially in the field of ufology, to gain a better understanding about UFO sightings based on geographic locations, and aggregated top keywords in a given area. The usefulness of the application is evaluated through a case study where a non-expert in ufology is able to find some typical interesting sightings. Our application can also be able to detect some misleading events such as missile launch or fireworks on a specific day through keywords and topic extraction. One limitation of our application is the data which is not up-to-date when new sightings are posted since the application pulled and processed data locally. One possible approach to overcome this drawback in the future is to consolidate data into database and update records in the database at regular time interval.

This tool is not limited to UFO sighting discovery but have more general applications on many different areas (which require to analyze the data on the following dimensions: time, geospatial, and text analysis), such as disease spreads, rumor tweets, and political blogs. This is one of the future direction of this work.

In another future work, we are planning to introduce topic recommendations into our application (the complete inferred sentence extracted from top topics). This new direction is very promising since it helps ufologists and scientists grasp deeper level of knowledge on UFO sightings.

References

[1] C. Sagan, “Unidentified flying objects,” *Bulletin of the Atomic Scientists*, vol. 23, no. 6, pp. 43–44, 1967. [Online]. Available: http://dx.doi.org/10.1080/00963402.1967.11455092

[2] K. ARNOLD, “Kenneth arnold - facts summaries,” 2017, http://www.history.com/topics/kenneth-arnold[Accessed date: June 27, 2017].

[3] R. J. Gribble, “National ufo report center,” 2017, http://www.nuforc.org[Accessed date: Mar 2, 2017].

[4] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and C. Sheelagh, “A review of temporal data visualizations based on space-time cube operations,” in *Eurographics Conference on Visualization*. Eurographics, 2014.

[5] E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT, USA: Graphics Press, 1986.

[6] N. Andrienko and G. Andrienko, “Interactive visual tools to explore spatio-temporal variation,” in *Proceedings of the Working Conference on Advanced Visual Interfaces*, ser. AVI ’04. New York, NY, USA: ACM, 2004, pp. 417–420. [Online]. Available: http://doi.acm.org/10.1145/989863.989940

[7] T. Dang and L. Wilkinson, “TimeExplorer: Similarity search time series by their signatures,” in *Proc. International Symp. on Visual Computing*, 2013, pp. 280–289.

[8] P. Shanbhag, P. Rheingans, and M. desJardins, "Temporal visualization of planning polygons for efficient partitioning of geo-spatial data," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, Oct 2005, pp. 211–218.

[9] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings of the 1996 IEEE Symposium on Visual Languages*, ser. VL '96. Washington, DC, USA: IEEE Computer Society, 1996, pp. 336–. [Online]. Available: http://dl.acm.org/citation.cfm?id=832277.834354

[10] R. E. Roth, K. S. Ross, B. G. Finch, W. Luo, and A. M. MacEachren, "A user-centered approach for designing and developing spatiotemporal crime analysis tools," in *Proceedings of GIScience*, vol. 15, 2010.

[11] A. Ramakrishna, Y.-H. Chang, and R. Maheswaran, "An interactive web based spatio-temporal visualization system," in *International Symposium on Visual Computing*. Springer, 2013, pp. 673–680.

[12] S. Monfort, "I want to believe, ufo sightings around the world," 1999, https://vizthis.wordpress.com/2017/02/21/i-want-to-believe-ufo-sightings-around-the-world/[Accessed date: Jan 21, 2017].

[13] J. F. SCHUESSLER, "Mutual ufo network," 2017, http://www.mufon.com[Accessed date: Feb 27, 2017].

[14] M. Galka, "Ufo sightings map," 2017, http://metrocosm.com/ufo-sightings-map.html[Accessed date: Mar 2, 2017].

[15] J. Nelson, "Ufo sightings per capita," 2001, http://uxblog.idvsolutions.com/2015/06/sightings.html[Accessed date: June 16, 2017].

[16] F. J. F. de Pedro, "Ufo hunters," 2007, http://www.ufo-hunters.com/[Accessed date: June 16, 2017].

[17] M. J. Darren Stevenson, JP Obley, "Project uppgifter: Us national distributions of reported ufo sightings," 2013, http://jpobley.com/wp-content/uploads/2013/09/uppgifter.pdf[Accessed date: June 17, 2017].

[18] D. A. Keim, C. Panse, and M. Sips, "Information visualization: Scope, techniques and opportunities for geovisualization," in *Exploring Geovisualization*, J. Dykes, Ed. Oxford: Elsevier, 2004, pp. 1–17.

[19] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *Proc. of the IEEE Symposium on Information Visualization*, 2005, pp. 15–24.

[20] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages & Computing*, vol. 14, no. 6, pp. 503–541, 2003.

[21] H. Hochheiser and B. Shneiderman, "Dynamic query tools for time series data sets: Timebox widgets for interactive exploration," *Information Visualization*, vol. 3, no. 1, pp. 1–18, Mar. 2004. [Online]. Available: http://dx.doi.org/10.1145/993176.993177

[22] A. Partow, "The global airport database," 2017, http://www.partow.net/miscellaneous/airportdatabase/[Accessed date: Mar 2, 2017].

[23] CoBases, "Cobases - us military bases," 2017, http://www.cobases.com/[Accessed date: Mar 2, 2017].

[24] U. S. C. Bereau, "American factfinder," 2017, https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml[Accessed date: Mar 2, 2017].

[25] B. Fry and C. Reas, "Processing.org," *Processing. org*, 2010.

[26] L. Science. (2017) 7 things most often mistaken for ufos. [Online]. Available: http://www.livescience.com/32849-7-ways-to-generate-a-great-space-hoax.html

[27] J. A. Hartigan and J. Hartigan, *Clustering algorithms*. Wiley New York, 1975, vol. 209.

[28] J. L. Klein, *Statistical visions in time: a history of time series analysis, 1662-1938*. Cambridge University Press, 1997.

[29] H. G. Funkhouser, "Historical development of the graphical representation of statistical data," *Osiris*, vol. 3, pp. 269–404, 1937.

[30] T. Dang and V. T. Nguyen, "ComModeler: Topic Modeling Using Community Detection," in *EuroVis Workshop on Visual Analytics (EuroVA)*, C. Tominski and T. von Landesberger, Eds. The Eurographics Association, 2018.

[31] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[32] V. T. Nguyen, A. S. Namin, and T. Dang, "Malviz: An interactive visualization tool for tracing malware," in *Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis*, ser. ISSTA 2018. New York, NY, USA: ACM, 2018, pp. 376–379. [Online]. Available: http://doi.acm.org/10.1145/3213846.3229501

[33] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[34] T. N. Dang, L. Wilkinson, and A. Anand, "Stacking graphic elements to avoid over-plotting," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1044–1052, Nov 2010.

[35] Mallet, "Topic modeling for java developers," 2017. [Online]. Available: http://mallet.cs.umass.edu/topics-devel.php

[36] M. Charvat, "Top 6 spots for ufo sightings in illinois," 2015, http://q985online.com/top-6-spots-for-ufo-sightings-in-illinois-list/[Accessed date: Feb 27, 2017].

[37] C. McNaught and P. Lam, "Using wordle as a supplementary research tool," *The qualitative report*, vol. 15, no. 3, p. 630, 2010.

[38] iDVL. (2017) Ufo tracker. [Online]. Available: https://github.com/iDataVisualizationLab/UFO

[39] G. A. Filer, "Radar confirms ufos over carteret, nj - vivid us sighting continue," 2001, http://www.rense.com/general13/filers999.htm[Accessed date: Feb 27, 2017].

[40] A. Gottfridsson, "Ufo - carteret, new jersey, july 15, 2001," 2008, https://www.youtube.com/watch?v=Yhzf9Lbfv1I[Accessed date: Mar 2, 2017].