Exploration of Large Image Corpuses in Virtual Reality

Sanket Khanwalkar
University of California, Irvine
skhanwal@uci.edu

Shonali Balakrishna
University of California, Irvine
shonalib@uci.edu

Ramesh Jain
University of California, Irvine
jain@ics.uci.edu

ABSTRACT

With the increasing capture of photos and their proliferation on social media, there is a pressing need for a more intuitive and versatile image search and exploration system. Image search systems have long been confined to the binds of the 2D legacy screens and the keyword text-box. With the recent advances in Virtual Reality (VR) technology, a move towards an immersive VR environment will redefine the image navigation experience. To this end, we propose a VR platform that gathers images from various sources, and addresses the 5 Ws of image search - what, where, when, who and why. We achieve this by providing the user with two modes of interactive exploration - (i) A mode that allows for a graph based navigation of an image dataset, using a steering wheel[1] visualization, along multiple dimensions of time, location, visual concept, people, etc. and (ii) Another mode that provides an intuitive exploration of the image dataset using a logical hierarchy of visual concepts. Our contributions include creating a VR image exploration experience that is intuitive and allows image navigation along multiple dimensions.

Keywords

Multimedia, Image Exploration, Big Data, Virtual Reality

1. INTRODUCTION

More data has been created in the 21st century than in all of human history. Estimates say that 2.5 quintillion bytes of data is generated each day [1]; So much so that 90% of the data in the world today has been created in the last two years alone. With the wide proliferation of photos and videos from phone cameras, the Web has also become increasingly visual. To put numbers in perspective for photos, in 2015 alone, for every minute on an average, Instagram users liked more than 1.7 million photos, Pinterest users pinned nearly 10,000 images and Facebook users uploaded nearly 136,000 photos [2, 4].

Information was initially represented in text documents. With the advent of the World Wide Web, it became possible to make the document corpus available to everyone, and most importantly, traverse from one document to the other using hyperlinks. The current information age needs to capitalize on the exponentially expanding image corpus by creating an interlinked visual web [18]. At present, the images cannot be interlinked with each other based on their contents or contextual metadata, making it extremely tedious and unintuitive to explore massive image datasets, whether personal or global, available on photo sharing applications like Google, Facebook, Instagram, Flickr or Pinterest. For example, Google Image Search only allows a spading-fork traversal in one direction, compelling users to query repeatedly to traverse in a different direction.

The image content and context (time, location) are becoming major sources of information for situation recognition [18, 24], with advanced techniques like deep learning being developed in computer vision for recognizing key concepts from the image content and easy availability of contextual information through the use of sensors like GPS, accelerometers etc. Therefore, these visual concepts and contextual features can now be leveraged in image search and exploration in order to implement an associative means of navigating through image datasets.

Until now, data visualization was limited to the representation of textual data, mainly on 2D interfaces. But as data evolves from text to multimedia (Eg. images), representing this data in 2D interfaces in an easily comprehensible manner becomes a challenge. Traditional visualization approaches when applied to large image datasets provide limited usability due to the constrained space on the visualization plane and the lack of interactivity for the user.

The platform on which data can be visualized is undergoing a transformation, from browsers on PCs to mobile devices to the upcoming Virtual Reality (VR) devices. VR could prove to be a powerful and engaging platform for image search and exploration, as it provides infinite environment space, multimodal user inputs (gaze, voice, head/hand motion) and distributed connectivity. VR, finally, offers an exciting possibility for applications in image exploration, because (i) with the advent of Google Cardboard and other such affordable and accessible VR devices compatible with smartphones, VR applications are now feasible for the typical user in terms of cost and size, (ii) there has been ex-

[1] In this paper, making a selection on a curved plane containing multiple options is equivalent to moving a 'steering wheel' in a particular direction.

MM '16, October 15-19, 2016, Amsterdam, Netherlands
© 2016 ACM. ISBN 978-1-4503-3603-1/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2964284.2967291

tensive research in academia and industry towards the facilitation of VR as an effective environment. In this paper, we propose a VR image exploration system that provides the user with two modes of exploration - Steering Wheel browsing and Hierarchical browsing. Steering wheel browsing allows the user to explore image datasets using a graph structure, along multiple dimensions like time, location, concepts, people, etc. Hierarchical browsing allows the user to explore image datasets using hierarchies that are based on visual concepts.

This paper is organized as follows: Section 2 describes related work in this domain. Section 3 discusses our proposed method, with the two modes of visualization, our image graph, system architecture, system implementation and evaluation. Section 4 discusses the conclusions of this paper, followed by a discussion of future work in Section 5.

2. RELATED WORK

While a lot of work has been done on content based image retrieval techniques, the existing semantic gap [27] is still a challenge and approaches like query-by-example [13] have limited application in practice. The most primitive image browsing visualizes a limited number of thumbnails on a 2D screen, requiring the user to continually go back and forth to view more images. More intuitive interfaces have been proposed for navigating image collections [25, 16], which can be broadly classified into mapping based, clustering based and graph based visualizations [22]; These visualizations are created using dimensionality reduction [27, 19], grouping based on image characteristics [11, 15] and interlinked image graphs [12, 28] respectively with similar images placed closer to each other. Visualizations represent clusters using a representative image, upon whose selection all images in the clusters are displayed to the user [8, 9, 10, 14, 23].

Traditional browsing techniques include horizontal browsing (within a particular level of the hierarchy), vertical browsing (navigation to a different level of the hierarchy) and time based browsing [21]. A spherical visualization approach to image exploration is proposed in [26], where image databases are handled in a hierarchical approach, with visually similar images placed together; Zooming operations reveal images on a deeper level of the tree structure, with the user being allowed to modify the browsing interactively.

In the past, image exploration systems like 3D MARS have been implemented in VR to leverage the infinite virtual space for displaying image query results using low-level image properties like color, texture and structure. However, such features for query-by-example render image exploration unintuitive limiting the accuracy in retrieving the semantic relationships between the images [20].

3. PROPOSED METHODS

As a first step in the direction of implementing an immersive image exploration system, we present a proof-of-concept in this section, with an overview of the image graph, the two modes, system architecture, implementation and evaluation. In our prototype, we provide the user with two exploratory modes - Steering Wheel browsing and Hierarchical browsing.

3.1 Image Graph

In recent years, with considerable research in deep learning techniques for concept extraction in images, and availability of more sensor information offering detailed context, the creation of an interlinked image graph is now a possibility. Such a visual web [18] can then provide traversal along multiple dimensions providing insights on what, where, when and who, thereby making intuitive image exploration a reality.

In our approach, we create the link structure using content and context based relationships between images, thus allowing navigation along multiple dimensions. We have used a university research dataset for image exploration in our system, which contains time and location information as part of the image EXIF metadata. Furthermore, this dataset is augmented with concepts that are mined from the images using a deep learning API (Clarifai [3]), location information that is retrieved via a location web service (Foursquare [5]) and people were extracted using pre-existing name tags, available for several images. Links are created by tagging image metadata based on time, location, concepts and people and navigation between images is based on similarity between these tags.

3.2 Steering Wheel Browsing Mode

In this approach, we arrange the image corpus as a graph with each image acting as a node, interconnected with other image nodes through edges. The edges between the graph nodes may represent connections based on time, location, concepts and people.

To implement this approach in VR, we begin with randomly generated images from the dataset. The currently viewed image is larger in size and placed exactly in front of the user, while other images are placed around it in the environment in a panoramic fashion. In this mode, the tiles on the steering wheel at the bottom represent the important entities present in the currently displayed image. These entities include time information, location information, concepts derived from the image and people present in the image. The location information is displayed via a static map and a place banner (if applicable) while people and concepts are displayed in a vertical menu.

In the example shown in Figure 1, the image identified 2 different persons viz. *Steve Jobs* and *Bill Gates* as people, *WWDC '08* as the place in *San Francisco*, which is the geographical location and dinner as a concept. If the user is interested in exploring photos involving one of the people or the city in the photo, he can simply make a selection using gaze-and-select mechanism in the direction of his choice on the steering wheel. A new set of photos are then rendered and the corresponding options on the steering wheel are updated, based on the current image. This style of exploration provides the user with valuable details present in the image and also the ability to traverse along the given dimensions (location, concept or people) in the image graph.

3.3 Hierarchical Browsing Mode

In this approach, we arrange the image corpus in logical hierarchies based on concepts present in the images. Each level in the hierarchy contains images belonging to the same concept category. Iteratively, each category at any given level, may further contain subcategories of images.

To implement this approach in VR, we begin our navigation with a predefined set of high-level categories. At each level in the hierarchy, images that belong to the same category are displayed in the environment and subcategories

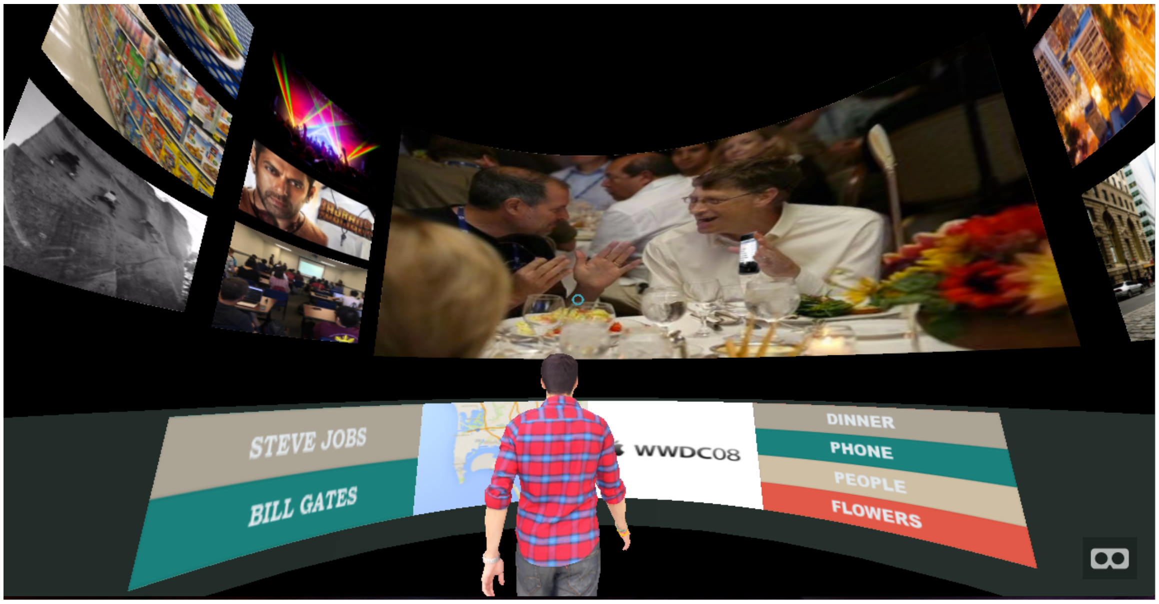

Figure 1: VR application showing graph-based image exploration

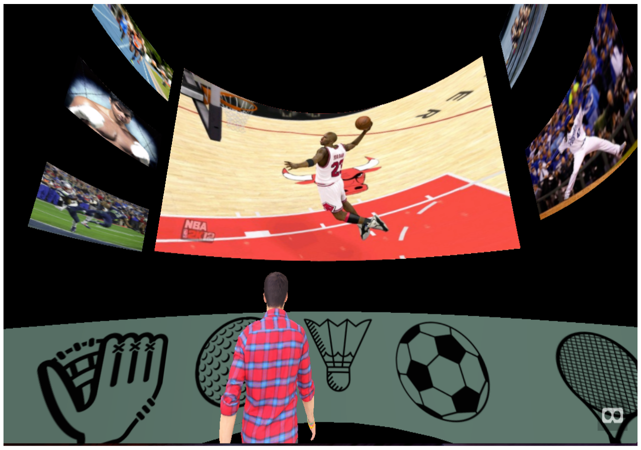

Figure 2: VR application showing hierarchy-based image exploration

are depicted by representative images on the bottom panel. With every selection on the panel, the system updates the environment with corresponding images belonging to that category and updates the panel with its subcategories. Once the user reaches the bottom of the hierarchy, the user is offered two options: (i) Traverse one level up in the hierarchy using the voice command 'back' or (ii) Jump into a conceptually related category using suggested options displayed on the right pane.

For example, we begin with popular search categories like Music, Hollywood, Sports, Politics, etc. placed as individual tiles on the bottom panel. As shown in Figure 2, selecting a representative tile at a given layer of the hierarchy, like Sports in this example, loads the subsequent options of the hierarchy like Tennis, Soccer, Baseball, etc. The user can refine his exploratory search by selecting any of these sports categories, say Soccer, by a gaze-and-select mechanism, thereby loading the next set of subcategories like players, teams, tournaments etc. on the bottom panel. Iteratively, after following a path like Sports –> Soccer –> Players –> *Lionel Messi* and reaching the end of this hierarchy, the user can either move back using a 'back' voice command or simply jump to related options like *Diego Maradona*, *Barcelona FC*, Golden Boot, etc. displayed on the right pane, and in this manner, continue his exploration infinitely.

3.4 System Architecture

Our system relies on the intricate integration between 4 components - data (image datasets, link structures), user input (voice, gaze-and-select), network (low latency, reliability) and VR display (aesthetic, intuitive, seamless rendering). Our system has a hybrid design that takes advantages of both client-side and server-side capabilities [17].

As shown in Figure 3, depending on the mode selected, the user is presented with the corresponding visualizations containing options to start his image exploration. The user provides his navigation input to the controller using voice or gaze-and-select. The controller communicates the user input to the backend and fetches the images and the corresponding link structures, for rendering in the virtual environment.

Our backend consists of servers and a data store; The data store contains the image corpus that will be explored. The servers are used to extract the link structure between images, from the image content and metadata, by querying external web services. We have utilized external web services for extracting visual concepts from image content, and possible locations from the spatial metadata.

3.5 System Implementation

3.5.1 Frontend Implementation

For developing and rendering a platform-independent application that can be deployed in the web, mobile or VR devices like Google Cardboard, we used MozVR [7]. This WebVR framework uses HTML elements to run the applica-

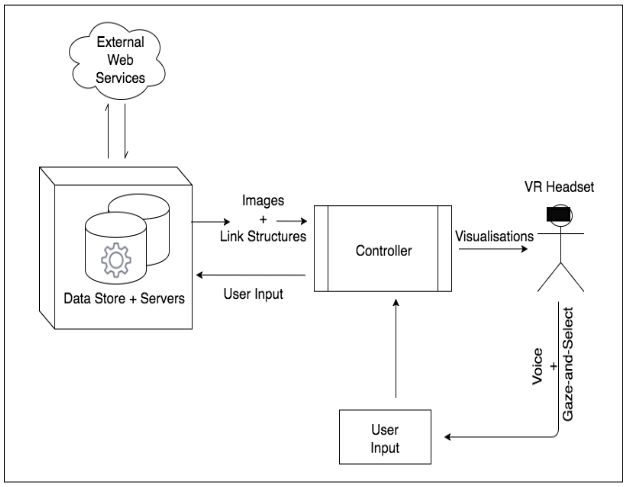

Figure 3: System Architecture

tion in a WebGL-enabled browser. This framework provides the gaze-and-select feature enabling users to choose options available in the VR space. Additionally, Google Voice API was integrated with this framework, thereby providing secondary assistance for user input.

3.5.2 Backend Implementation

We created a REST API to retrieve images from an existing university research dataset, and have further processed these images to create the link structures using external web services. The backend is designed to query external web services like ClarifAI [3] for extracting visual concepts from image content and Foursquare [5] for extracting possible locations from the spatial metadata. Our backend technology stack is Python based, and we use a Python web framework called Flask to create the REST API. For rendering static maps in our browser, we send location details in a HTTP request to Google Maps and receive static images as the response. All the other visualizations are processed in the backend and only their output is rendered in the browser as images.

3.6 Evaluation

For a general evaluation of our system, we conducted a pilot study with 11 participants (age group of 21-30). We deployed the application on a web browser in an iPhone, with a Google Cardboard as the VR device.

The following steps were performed in carrying out this study: (i) The user was initially asked to explore images from Flickr, Instagram and Pinterest feeds on a computer or smartphone. (ii) The user was then asked to explore images in our VR application rendered on Google Cardboard. (iii) During his exploration, we provide guidance to the user to help him understand his VR environment better. (iv) Post our study, with each participant taking around 10 minutes exploring our system, they were requested to answer a few questions, on a number scale of 1-5 (5 being most satisfied), related to their experience, concerns, usefulness, etc. Results of our study can be found in Table 1.

Our evaluation results suggest that there is definitely a need for a better image exploration system. The rather average score of 3.3 for the gaze-and-select mechanism can be attributed to the lack of sophistication in the UI/UX of our preliminary prototype as well as the novelty of the VR environment in general. However, the evaluation study for the remaining 4 questions clearly suggests that there is considerable potential in such a system and a willingness among participants to use this image exploration system, especially for social media and personal photos.

Questions	Score
How satisfied are you with legacy image search using keywords in textboxes?	2.7
How positive was your experience with gaze-and-select for image exploration?	3.3
How intuitive was image exploration using the steering wheel browsing?	4.4
How positive was your experience exploring images in infinite virtual space?	4.3
Would you use such a VR application in the future for image exploration?	Yes (91%)

Table 1: Evaluation Results

4. CONCLUSION

We propose an approach that provides two modes to the user - first, a mode for browsing through an interlinked image graph structure along dimensions like time, location, concepts, people; second, a mode that allows an intuitive navigation through conceptual hierarchies of images for a simpler exploratory search and browsing experience in VR. The advantage of using our proposed approach is two fold - (i) it provides interlinking between images along many dimensions by incorporating contextual and conceptual information into the navigation. (ii) it provides the user with an interactive and immersive exploration experience in VR space, where he gets to engage and set the tone of the navigation.

5. FUTURE WORK

Based on our literature survey and current work, supported by the pilot study, we strongly believe that further work on the following areas will lead to the successful convergence of image exploration and virtual reality. Firstly, a knowledge graph can be applied to the objects or concepts present in an image for more insightful exploration of large image datasets. In order to infer deeper meaning from plain text queries, Google's Knowledge Graph considers the semantic meaning behind the query and fetches improved search results. This graph is built by finding entities and relationships associated with each word with the help of sources like Wikipedia [6]. Thus, creation of a knowledge graph for images will enable efficient exploration of large image datasets by offering semantically meaningful directions of exploration. Second, the nature of VR technology creates opportunities in collaborative image viewing/exploration experience, which was not possible before. With multimodal user input, infinite environment space, remote and distributed connectivity and sophisticated sensors, collaborating and sharing your image exploration can certainly become a reality. Third, this system can be extended as a platform to incorporate social media feeds like Instagram, Pinterest, Flickr, etc. along with personal image collections to create a unified image exploration experience.

6. REFERENCES

[1] Big Data Statistics. http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/.

[2] Big Data Usage Per Minute. http://www.inc.com/larry-kim/15-mind-blowing-statistics-reveal-what-happens-on-the-internet-in-a-minute.html.

[3] ClarifAI. https://developer.clarifai.com/.

[4] Facebook Data Statistics. https://zephoria.com/top-15-valuable-facebook-statistics/.

[5] FourSquare. https://developer.foursquare.com/.

[6] Google Knowledge Graph. https://en.wikipedia.org/wiki/Knowledge_Graph.

[7] MozVR Framework. http://mozvr.com/#developers.

[8] I. Bartolini, P. Ciaccia, and M. Patella. Adaptively browsing image databases with pibe. *Multimedia Tools and Applications*, 31(3):269–286, 2006.

[9] J.-Y. Chen, C. A. Bouman, and J. C. Dalton. Hierarchical browsing and search of large image databases. *Image Processing, IEEE Transactions on*, 9(3):442–455, 2000.

[10] Y.-X. Chen and A. Butz. Photosim: Tightly integrating image analysis into a photo browsing ui. In *Smart Graphics*, pages 224–231. Springer, 2008.

[11] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 1(3):269–288, 2005.

[12] M. Dontcheva, M. Agrawala, and M. Cohen. Metadata visualization for image browsing. In *18th annual ACM symposium on user interface software and technology*, 2005.

[13] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of intelligent information systems*, 3(3-4):231–262, 1994.

[14] A. Gomi, R. Miyazaki, T. Itoh, and J. Li. Cat: A hierarchical image browser using a rectangle packing technique. In *Information Visualisation, 2008. IV'08. 12th International Conference*, pages 82–87. IEEE, 2008.

[15] B. Gong, U. Westermann, S. Agaram, and R. Jain. Event discovery in multimedia reconnaissance data using spatio-temporal clustering. In *Proc. of the AAAI Workshop on Event Extraction and Synthesis (EES'06)*, 2006.

[16] D. Heesch. A survey of browsing models for content based image retrieval. *Multimedia Tools and Applications*, 40(2):261–284, 2008.

[17] B. Huang, B. Jiang, and H. Li. An integration of gis, virtual reality and the internet for visualization, analysis and exploration of spatial data. *International Journal of Geographical Information Science*, 15(5):439–456, 2001.

[18] R. Jain. Let's weave the visual web. *MultiMedia, IEEE*, 22(3):66–72, 2015.

[19] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. S. Huang. Visualization and user-modeling for browsing personal photo libraries. *International Journal of Computer Vision*, 56(1-2):109–130, 2004.

[20] M. Nakazato and T. S. Huang. 3d mars: Immersive virtual reality for content-based image retrieval. In *ICME*, 2001.

[21] W. Plant and G. Schaefer. Navigation and browsing of image databases. In *Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of*, pages 750–755. IEEE, 2009.

[22] W. Plant and G. Schaefer. Visualising image databases. In *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*, pages 1–6. IEEE, 2009.

[23] J. C. Platt, M. Czerwinski, and B. A. Field. Phototoc: Automatic clustering for browsing personal photographs. In *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, volume 1, pages 6–10. IEEE, 2003.

[24] S. Pongpaichet, M. Tang, L. Jalali, and R. Jain. Using photos as micro-reports of events. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pages 87–94. ACM, 2016.

[25] S. D. Ruszala and G. Schaefer. Visualisation models for image databases: A comparison of six approaches. In *Irish machine vision and image processing conference*, pages 186–191. Citeseer, 2004.

[26] G. Schaefer. A next generation browsing environment for large image repositories. *Multimedia Tools and Applications*, 47(1):105–120, 2010.

[27] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380, 2000.

[28] M. Worring, O. de Rooij, and T. van Rijn. Browsing visual collections using graphs. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 307–312. ACM, 2007.