Iconizer: A Framework to Identify and Create Effective Representations for Visual Information Encoding

Supriya Garg, Tamara Berg, Klaus Mueller

Computer Science Department, Stony Brook University

Figure 1: Overview of the system

ABSTRACT

The power of visualization is driven by the fact that humans are able to detect patterns much better with visual representations than with textual ones, enables more complex reasoning processes. Currently the degree of semantics encoded in these visual representations is quite limited. The use of icons as a form of information encoding has not been sufficiently explored. Typically graphs in visual analytics use standard icons from a fixed icon gallery or an actual picture. Here we describe a framework that uses a dual domain approach involving natural language processing and global image databases to help users identify metaphors suitable to visually encode abstract semantic concepts. The visual representations resulting from these visual encodings can then be used as stand-alone expressive icons or as clip art in visualization and visual analytics applications.

KEYWORDS: Visual knowledge representation, icon design, illustrative visualization

INDEX TERMS: H.1.2 [Information Interfaces and Presentation]: Human Factors; I.3.3 [Computer Graphics]: Display Algorithms; I.4.9 [Image Processing and Computer Vision]: Applications

1 INTRODUCTION

Scientific evidence shows that visual encodings, if chosen well, are more efficient than textual ones. Decoding them is more time-efficient because visual intelligence can rapidly and effortlessly interpret what is seen, quickly making inferences and explaining causal histories. Further, having visual encodings with semantic information can highly enhance the readability of social network graphs, as commonly used in Visual Analytics. Our paper proposes a framework and infrastructure which seeks to assist in this effort of *visual information encoding* (VIE).

In particular we wish to take advantage of the fact that in human perception there exist iconic views of objects, as has been demonstrated in the psychology literature. We propose to find representative examples for hierarchies of object classes. Instead of having humans hand select these canonical images to use as VIEs, we propose that these biases in preference will emerge naturally when mining large image collections on the internet. Presently, clipart, a form of VIE, consists of descriptive icons meant for illustration. In practice, a narrow range of clipart is used, often marginally matching the situation at hand. Users instead resort to searching the web to gather more relevant illustrations. The framework we present here provides a computer
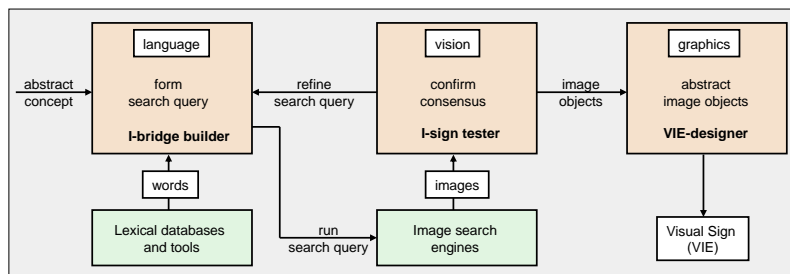
Stony Brook University, Stony Brook, NY 11794
{sgarg, tlberg, mueller}@cs.sunysb.edu

aided system that allows users to produce clip art for their concept of interest, in a style matching their personal preference.

2 APPROACH

We aim to find VIEs that are intuitive, i.e., are already part of one's visual vocabulary. VIEs are relatively easy to find for most objects and some actions because they can be directly observed in real life and are already part of one's visual vocabulary. However, abstract concepts pose a challenge. For e.g., when people are asked about concepts like 'travel' or 'admire', they offer a wide variety of possible VIEs. This variety is also reflected in the query results with image search engines, such as *Google Images*. We hence desire a VIE that reaches the broadest consensus among a given population. Lexically, we find such information in Lexical FreeNet [1] – a lexical database augmented with statistical information in the form of lexical triggers. These triggers are computed by analyzing word distribution in thousands of documents and looking for word pairs that frequently co-occur. On the image side, this notion of a visual trigger to map abstract concepts to VIEs is indirectly available via search engines – querying for a concept like 'travel' gives us semantically related images like suitcase, passport, beaches, flights etc. We propose a dual-domain approach to incrementally build and format these visual triggers. Our approach uses the available lexical triggers to form suitable search terms to query public image search engines and then iteratively tests the returned images for consensus.

3 IMPLEMENTATION

Figure 1 gives an overview of our pipeline to design VIEs. In the case of abstract concepts with no direct icons, we need indexical-signs (I-sign). I-signs must use a good ontological metaphor by which the abstract concept is represented as something concrete.

Figure 2: Semantic zoom from an unknown person to Bill

Figure 3: Left to right: Five instances of a large image set of cars, followed by the average car

We call this metaphorical mapping the iconicity-bridge (or I-bridge). Such associations consist of lexical classifications, such as synonyms, antonyms, and trigger relations. Google image search is used to see what images people commonly associate with a given query – our visual triggers. To get a diverse and comprehensive set of images for a query, we download the top 200 Google Image search results in 5 different languages. Results from foreign language queries also provide a fallback method in the case when the English word for a concept has multiple meanings. We merge these knowledge bases in our user interface to help the user build an effective I-bridge. The user is provided with some *utility* or helper terms, such as 'gear', 'device' or 'supplies', which when combined with the original term give useful image results.
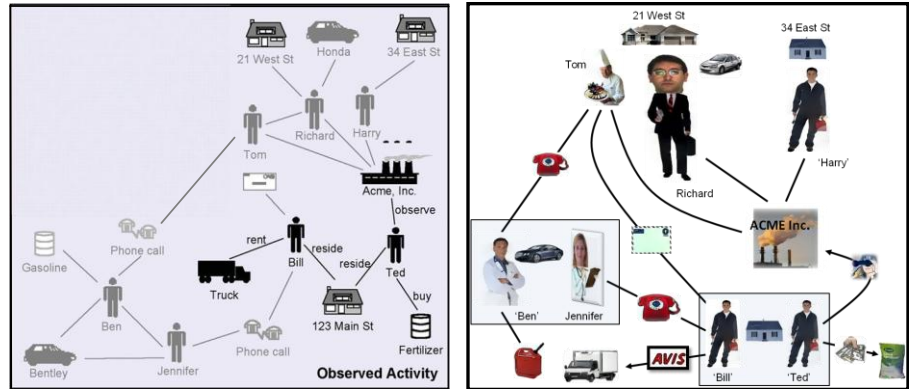

Figure 4: On the left – a commonly seen social network graph embellished with text and icons. On the right – the same graph made more expressive by adding relevant information

Consensus among I-signs is tested by clustering the images returned during the I-bridge building process. The top images from each cluster are shown to the user. Clusters which represent the concept well, and contain semantically similar images have good consensus. We use the scene based feature vector *gist* [3] during the clustering process.

Constructing visual equivalents of model-based abstractions requires a semantic abstraction of these I-signs, which goes much beyond the image-based abstraction methods available today. We seek a picture of the given concept that unifies all of the concept's known facts, while abstracting away the unknown facts. In images, a 'fact' is expressed as a visual feature, or collection of features. We can construct an average image for a category by looking for common features across a set of queries. For e.g., we may only know that our suspect is a 'male, blue-collar worker', so we could formulate a corresponding search query of this term, and in addition also use our lexical resources to retrieve images of various instances of the term, here mechanics, plumbers, electricians, and so on. The noise, i.e., the difference in detail across images can then be removed by finding the level of image-based abstractions at which all images appear similar. We implement this method using multi-scale edge detection to compute shape context histograms [2] at multiple scales. We then pack these into high-dimensional feature vectors and cluster them for all images in the set. At some level the similarity between the image feature vectors is minimized. This corresponds to the scale at which the different expressions for a concept are the most similar. The cluster exemplar (an image closest to the cluster center) then represents the average instance of this concept. Given the exemplar and the scale at which the distance is minimized, we abstract away all details at higher scales to design the final icon. For this, we use Poisson-blending guided by the noise-free edges to create an abstracted illustration that only contains features of known facts. The mechanic in Figure 2 is an e.g. of this process. Figure 3 shows some representative cars, and the average car.

S*ymbolic processes* can encode what a participant means or is. It may show a participant posing for the viewer, along with an icon symbolizing the associated information. Here, we use context to modify the meaning of an icon. For e.g., given that a person is a mechanic, the icon can be a man in overalls. As we get to know more details – e.g. the man is a car mechanic, or a sports car mechanic, the appropriate object is added to the background (see Figure 2).

4 RESULTS

The direct impact of our proposed system is best explained by presenting social network graphs. Graphs have low visual expressiveness, with generic icons or text to annotate nodes and links. In Figure 4 we show one of such graphs on the left, and a more expressive version of it on the right. The new graph shows the professions of the people, and replaces text and generic icons by relevant icons. Further, we use symbolic representation to show which people have incomplete information on files by making them face us, hence demanding our attention to inquire further. Also, Richard has been given a relatively larger size showing that he is very important to the current investigation. Co-ownership of houses etc is represented by co-locating the related entities within a box. This modified graph provides an analyst with more information, and with some leads as discovered automatically by the system, or by some other analyst before him.

5 CONCLUSIONS AND FUTURE WORK

Our framework uses basic visualization and vision concepts to form an icon given a bunch of related images. We believe that this framework can make a great tool for visual analytics, making reasoning with graphs more engaging and informative. In future work we plan to fully integrate our framework into the graph drawing engine, use abstractions more freely to summarize certain facts and attributes, and use compositions for compact visual story telling with background context and foreground key players. Given the current status of implementation, we believe that we can deploy our interface and backend processes to a wider circle of users, over the web. Such a community-driven effort will likely result in much more robust icons, and further insight into personal preferences as well.

REFERENCES

[1] D. Beeferman, "Lexical discovery with an enriched semantic network," *Workshop on Appl. of WordNet in Natural Language Processing Systems, ACL/COLING*, pp. 135–141, 1998.

[2] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE PAMI*, 24:509-522, 2002.

[3] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope" *IJCV*, 42:145-175, May, 2001.