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
{sgarg, tlberg, mueller}@cs.stonybrook.edu

Abstract. The majority of visual communication today occurs by ways of spatial groupings, plots, graphs, data renderings, photographs and video frames. However, the degree of semantics encoded in these visual representations is still quite limited. The use of icons as a form of information encoding has been explored to a much lesser extent. In this paper we describe a framework that uses a dual domain approach involving natural language text processing and global image databases to help users identify *icons* suitable to visually encode abstract semantic concepts.

Keywords: human-computer interaction, non-photorealistic rendering

1   Introduction

Much scientific evidence indicates that visual encodings, if chosen well, are more effective than textual ones. Their decoding is more time-efficient because human visual processing can rapidly and effortlessly interpret what is seen, make inferences, and explain causal histories [13,20]. Further, the retention of visually encoded content in short-term memory tasks is higher and visual encodings are more space-efficient. Experiments with a database on terrorist attacks in the 1980s have shown that iconized data occupies 10 times less space than text [16]. Thus the time for an analyst to browse, analyse and search iconized data over textual data can be potentially reduced by an order of magnitude [19]. Along these lines it has also been demonstrated that people can work with multi-dimensional icons twice as fast as with text [30]. In fact, some have called visual language a 'prosthesis' for some of the limitations of human thought since it supports and augments the severe limitations of working memory [14]. The potential of visual information encoding is undoubtedly high, but the encodings must be selected wisely to reach these reported performance rates. Our paper proposes a framework that assists in a user's effort to select or create good iconic representations for the visual encoding of concepts.

  In particular we wish to take advantage of the fact that there exist iconic or canonical views of objects, as has been demonstrated in the psychology literature for human perception. In their seminal work, Palmer et al [24] find that humans agree on canonical views of objects and that recognition is faster for these views. We propose to find representative examples for hierarchies of object classes. However, instead of having humans hand select these canonical images to use as visual encodings, we propose that these biases in preference will emerge naturally when mining large collections of images taken and posted to the internet by people.

Icons (in computing) have been around since the 1970s to make computer interfaces easier to understand for novice users, mapping concepts to standardized visual representations. The majority of these icons are symbolic representations of applications that need to be memorized by the user. Clip art, on the other hand, aims to be more descriptive and is meant for illustration. A very narrow set of clip art is used in practice, often marginally matching the situation at hand. A third option for selecting iconic representations is to use web search to gather images fitting a desired concept. However, for complicated concepts, this results in limited success because multiple queries or lengthy searching must be performed to match a concept exactly. The framework we present provides a computer-aided system that allows users to quickly and effectively design clip art that is well targeted to their concept of interest. We achieve these capabilities by extending and synthesizing techniques rooted in non-photorealistic rendering and computer graphics, image processing, web-scale content-based image retrieval and natural language processing.

The ability to design well-targeted expressive clip art in a cohesive illustrative style enables applications at a scale much grander than a singleton. We may use them for the illustrations of documents, books, manuals, and the like, and they are also applicable to visualize taxonomies of objects and even more general concepts. Further, they can replace or complement textual annotations and photographs within node-link diagrams often used in analytical reasoning tasks, making these representations much more expressive.

Our paper is structured as follows. Section 2 presents the general philosophy behind our methodology, which is rooted in a joint lexical and visual analysis. Section 3 presents previous work in the area of visual languages and icon generation as well as the background of our approach. Section 4 describes our approach in detail. Section 5 presents results and some discussion on our system, and Section 6 ends the paper with conclusions and future work.

2   Overall Motivation and Philosophy

We aim to find Visual Information Encodings (VIEs) that are intuitive, i.e., are already part of one's visual vocabulary. This avoids the need for memorization of a set of dedicated symbols for iconic communication. VIEs are relatively easy to find for most objects and actions because they can be observed in real life and are already part of one's visual vocabulary – yet their interpretation and aesthetics still leaves much room for artistic freedom in determining the best VIE design. However, as with visual languages, the greatest challenge comes from determining good VIEs for abstract concepts. Take for example, the concept 'travel'. When asked, people will offer a wide variety of possible VIEs for these concepts, and this variety is also reflected in the query results with image search engines. Hence, we desire a VIE that reaches the broadest consensus among a sufficiently wide population.

We propose an indirect approach to find this consensus, circumventing the need for an active solicitation of user responses to candidate VIEs. Instead we exploit existing public lexical databases like WordNet [7] augmented with aggregated statistical information in the form of lexical triggers [1] and couple these with public image search engines. These triggers are computed by analysing thousands of documents
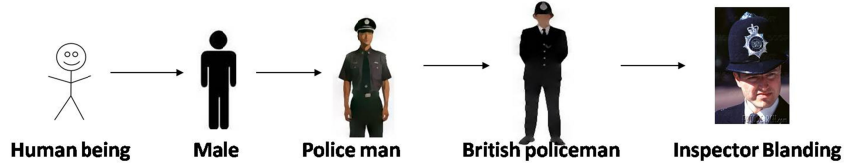
Fig. 1. Specialization from an unknown person to Inspector Blanding

and looking for words that commonly co-occur. Thus we resort to employing methods that statistically analyse data that humans have produced and which in some form represent their view of the world. Examples of lexical triggers are 'sky → blue' or 'travel → passport'. We also require a similar notion on the image side, that of visual triggers, to map concepts to suitable VIEs. The notion of a visual trigger is not readily available from any current image database. We propose a dual-domain approach to (incrementally) build and format these visual triggers. We note that this is an extremely large undertaking and all we can do here is to propose a methodology by which this could be done. Our approach uses the available lexical triggers to allow users to interactively explore the concept space, select suitable representative concepts, and finally create icons.

As mentioned, concepts can range from very specific to fairly general. Figure 1 illustrates this via an example that shows a conceptual zoom across many conceptual levels, here from 'Man' all the way to a specific person 'Inspector Blanding'. These conceptual zooms are not just multi-resolution representations obtained by low-level abstraction, i.e., by intensity or gradient domain filtering. Rather, they are semantic zooms, i.e., categorical refinements or generalizations within an object hierarchy.

3 Related Work and Background

Visual languages range from icon algebra [17] to the encoding of all information into multi-frame artist-developed cartoon-like renditions [3]. The set of icons is typically fixed, developed manually, and can be composited. Semanticons [28] is an innovative way to create new file icons by abstracting terms occurring in the file or file name along with a commercial database of images. None of these applications exploit any semantic analysis, nor do they make use of the large publicly available lexical and image databases to broaden the semantic base for abstraction and enable VIE learning. Other related work includes that of Rother et al. [27] which enable the automatic creation of collages from image collections to compose a single image with blended collection highlights. Here the user has no control over the layout of images inserted into the collage. Alternatively, Photo Clip Art [18] provides an interface that enables insertion of photo-realistic objects into new images, correctly constrained to be in a natural looking context within the resulting image. Their goals however are different than ours in that their visual objects are meant to enrich or compose graphical scenes and collages, without placing special emphasis on conveying specific semantic information. Finally, there is research targeting the automatic illustration of text via application of 3D graphics engines. The systems by Götze et al. and Götzelmann et al. [10,11] annotate their renderings according to the underlying text,
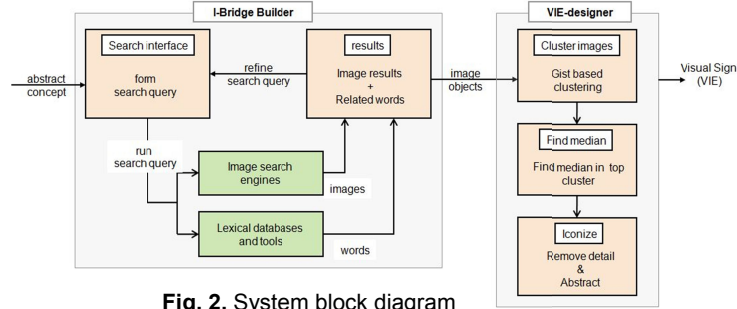
Fig. 2. System block diagram

while Word-Eye [4] analyses descriptive text according to a set of hand-coded rules to produce 3D scene renderings. These systems require a library of 3D models to which their illustrations are limited. The Story-Picturing-Engine [15] has a greater gamut, indexing image search engines with exact lexical terms occurring in the text, but no further semantic interpretation, integration, or visual abstraction is made.

One high level challenge we face is how to find representative or iconic images for a given query term. For this purpose we can exploit the vast number of images available on the Internet. The sheer number of images presents challenges not experienced in traditional well organized, labelled collections of images. Recent work has explored methods for choosing the most representative or canonical photographs for a given location [29] or monument [20] through clustering, where hard geometric constraints can be used since the object is a single instance seen from multiple viewpoints. These methods are less useful here since we apply our system to object categories that vary quite widely in appearance. Other methods work towards finding the most aesthetically pleasing search results, since returning a poor quality image is probably never aligned with a user's needs [5,26]. We propose a human-computer integrated approach that combines textual analysis with image clustering techniques and human aided image selection.

Another area with a similar high level goal to our work is the problem of image classification for content based image retrieval (CBIR). These systems utilize purely image based information for content analysis and retrieve images by measuring their similarity to a given query image (an in depth review is presented by Datta et al. [6]). The most successful approaches typically integrate a variety of colour, texture, shape, or region based cues. However, the problem of content based retrieval is extremely challenging and far from solved for most object categories, and these systems are often helped by having a human in the loop to guide the search process via relevance feedback [9,12,32]. We take such an approach here.

4   Approach

Our overall system is depicted in Figure 2. It consists of two main components, the I-bridge builder and the VIE designer. If our concept has a direct physical representation, then our job is fairly easy. Else, we require an indexical sign (I-sign). I-signs must use a good ontological metaphor by which the abstract concept is
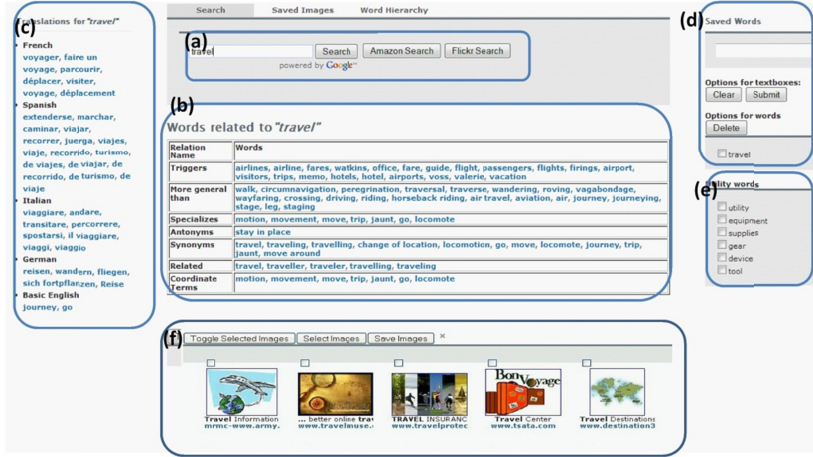
Fig. 3. The user interface. (a) Query-box (b) Related words (c) Translations to foreign languages (d) History + query builder (e) Utility words (f) Image results

represented as something concrete, such as an object, substance, container, person, or some visual action. Good mappings improve distances in conceptual space, moving the concept closer to the visual encoding. The first part of our framework is designed to build this bridge crossing conceptual space – we call it the iconicity-bridge or I-bridge. We use lexical databases and image search engines to derive potential I-bridges. Our second step takes the resulting images to create a VIE that represents the most central visual theme of a concept (the graphics-based VIE-Designer). This involves clustering the images, finding the median in each cluster, and finally abstracting the median so that it captures the common features of all objects in the cluster. In the following sections, we describe each of these components in detail.

4.1 The I-Bridge Builder

A good I-bridge is an association that is deeply rooted in our semantic understanding of the world. Such associations consist of pure lexical classifications, such as synonyms, antonyms, hyponyms (specializations), and hypernyms (generalizations) as well as statistical knowledge about co-occurrence relations between terms in documents or spoken language (so-called trigger relations). The former is captured in public lexical databases such as WordNet [7], while we use Lexical FreeNet [1] to provide the statistical knowledge component. Basic English [21] gives us the equivalent of a word in restricted English – e.g. 'bombshell' translates to 'great shock'. Utility or helper terms are also provided which combined with the original query often help narrow image results to a particular visual metaphor. For example, utilizing the 'gear' utility with the query 'travel' might direct a user towards images of suitcases or backpacks. Lastly, language translators are used to translate concepts that are polysemous (have multiple meanings), or are brand names in English.

In our web-based interface (Figure 3), all of these lexical and statistical associations are exposed to the user as tools for enabling conceptual connections and exploring the

space of an input concept. This results in a powerful interactive interface to help the user build an effective I-bridge. Figure 3a shows the query input box. Figure 3b displays semantically related words from WordNet and statistically related words from Lexical FreeNet. Figure 3c provides translations of the query concept into four languages plus Basic English, while Figure 3d maintains a history of the explored query concepts during an I-bridge building session. In Figure 3e helper terms such as 'equipment' or 'tool' are provided to enable focusing on particular visual senses. Finally, Figure 3f displays the top results from Google image search.

After the initial display of results, the user can do several things. In a perfect world, After the initial display of results, the user can do several things. In a perfect world, several relevant images would be found in the first go and the user can store them to the saved images panel. Otherwise, the user can continue browsing the lexical space until the images reflect his desired concept. For example, the user might want to explore concepts semantically or statistically related to his query (Figure 3b). Alternatively, the user could use the query builder to make complex queries. For example, for a query like 'art', the image results returned might be too 'artsy' and not represent the high level concept 'art' in a simple, concise manner. An image query of 'art' + 'supplies' (a utility word from Figure 3e) gives more concrete results such as images depicting coloured pencils, crayons or paint. The image search results themselves can suggest good I-bridges. For example, the image results for travel includes an airplane, a map and a compass, suitcases, and a person on the beach.

In order to construct a good VIE for the user's candidate I-sign, we will require a large number of relevant images to mine for the most iconic visual representation. In our experience with the I-bridge builder we have observed that more specialized queries tend to return image sets that are more coherent. For example, the query 'man' returns a diverse set of images with many depictions, while the image set for 'police man' is more homogeneous. This reveals a powerful strategy for I-sign learning: Join all the image instances obtained with specializations of the target term, obtained through our semantic analysis, and then use this collection to build the VIE.

In order to get a diverse and comprehensive set of images for a query, we download the top 200 results from Google Image Search. We ignore the later results as they tend to become much more noisy and unreliable. Instead, to increase the size of our data set while maintaining high quality, we translate our queries into 4 other languages and collect the top 200 results from each translated term.

4.2 The VIE-Designer

Constructing visual equivalents of model-based abstractions requires a semantic abstraction of these I-signs, which as mentioned above goes much beyond the image-based abstraction methods available today. More concretely, we seek a picture of the given concept that unifies all of the concept's known facts, but abstracts away the unknown facts. In images, a 'fact' is expressed as a visual feature, or a collection of features. An image set that bears feature 'noise' is a set of images that share some features (facts), but also contain a wide selection of other random features (unknown facts). We can construct an average or exemplar image for a category by looking for features in common across a set of queries.
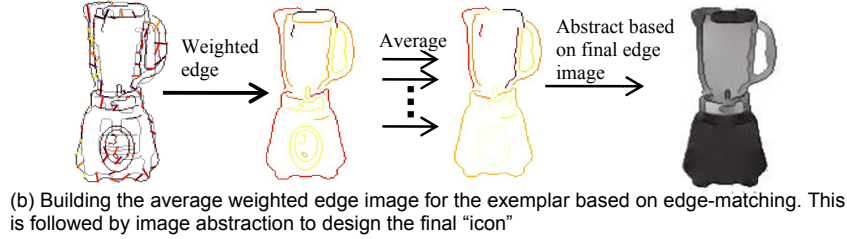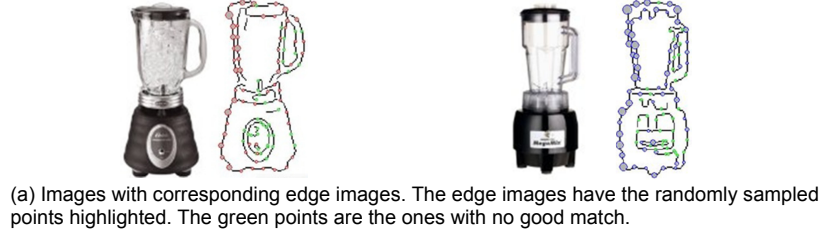
(a) Images with corresponding edge images. The edge images have the randomly sampled points highlighted. The green points are the ones with no good match.

(b) Building the average weighted edge image for the exemplar based on edge-matching. This is followed by image abstraction to design the final "icon"

Fig. 4. Transformation of a median image to the final *icon*

In this section, we present our algorithm to extract the basic icon for a set of images. Given a set of images belonging to the same category, we can find the most common shape features among them, and produce an icon. This part of our system is implemented in MATLAB.

4.3   Exemplar finding

In our system, since the images we use are results of queries to image search engines, most images will have a simple layout with the object of interest covering a large part the image and a relatively clean background. Hence, we can use a global scene descriptor like gist [22] that captures the general layout and shape of the image subject, to cluster our images. Gist is an image descriptor commonly used in computer vision and graphics applications. Gist provides a whole image feature descriptor encoding a coarse representation of the oriented edges at a range of frequencies present in an image. Since the size of the descriptor depends on the size of the original, we resize all images to a fixed size of 160x160 to compute equivalent descriptors. Next, we cluster the images and select cluster exemplars using affinity propagation (AP) [8]. This is a state of the art clustering method that takes as input measures of similarity between data points and 'preference' values indicating preference for a particular data point to be an exemplar i.e. cluster center. The algorithm then selects a set of good exemplars and corresponding clusters through a message passing algorithm. In our case we measure similarity between images using Euclidean distance between their Gist descriptors, and set the 'preference' values to be Google Image Search rank of each image (since images appearing earlier in the ranking tend to better reflect the search term). Given the resulting clusters, we order the images within each cluster based on their similarity to the exemplar. Finally, we select the cluster used to build our final VIE. This choice is based on two criteria – the size of the cluster and the average distance from the exemplar.

4.3.1 Image abstraction

At this stage we have a cluster representing a query, and its exemplar. Instead of presenting the exemplar as the icon, we abstract it so that only the relevant details present in all the images in the cluster are maintained, while removing the details specific to them. This works well only in cases where the whole object can be cleanly separated from the background. In cases where all the images have a dense background (for e.g. in case of animals), we simply use a non-photorealistic (NPR) version of the exemplar as the VIE. The NPR version works better than the original image when multiple VIEs share screen space – the abstraction process modifies images such that they look like they have come from the same source.

To distil an exemplar down to its VIE we use a shape representation to find edges in the exemplar that are also present in the other images within the cluster – *relevant details* – and remove those edges that are not present in other cluster images – *details specific* to only the exemplar. We first randomly sample feature points (locations at which to compute local features) from the edges present in each image. Then we extract shape-context [2] descriptors for each of these feature points. Shape-context is a computer vision feature descriptor that describes object shape with respect to a given feature point by computing a log-polar histogram of the edges surrounding that feature point. Next we perform a shape based alignment between the exemplar and the remaining images using the Hungarian algorithm to find the optimal one-to-one matching between the point sets (Figure 4). The Hungarian algorithm is a well-known combinatorial optimization algorithm which solves the assignment problem in polynomial time. Similarity is measured between two points using the Euclidean distance between their normalized shape-context vectors. We then remove bad feature point matches based on: a) low shape-context feature similarity, and b) large distances in image space (indicating false matches between points in very different parts of the object). This method works well at matching objects which are misaligned due to rotation or translation, but cannot handle objects that are flipped. Since the Hungarian algorithm is cubic in complexity, the running time grows fast with the number of
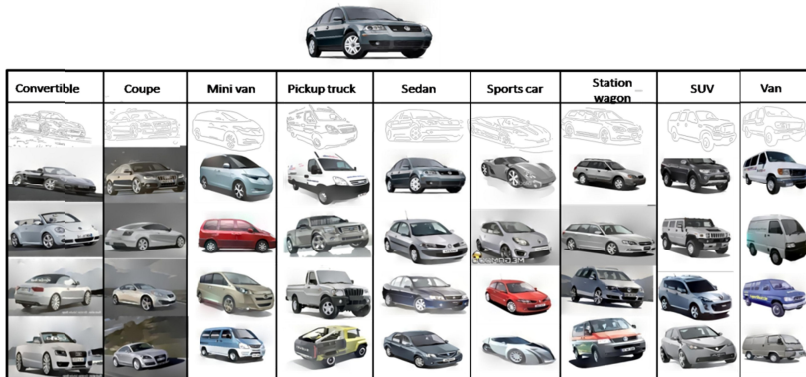

Fig. 5. Icon for average car. It is built by finding the combined median of the 9 sub-categories shown here. The images under each column are the top results for that category. The top row shows edges at the scale at which the cars look most similar.

8

Fig. 6. Icons showing the taxonomy under the category "Small Appliances". On the right, we zoom in to see the icons for the sub-category "Coffee tea espresso"

points sampled on the edges. We typically use 100 points to give us a good balance between speed and accuracy. At this complexity level, the shape matching takes a few seconds per image, with the total time dependent on the size of the cluster.

At this stage, we are left with point pairs that are highly similar. We assign scores to the exemplar points based on their similarity to the matching points in the other images. Further, the remaining points on the exemplar edges are assigned scores equal to the nearest exemplar point on a connected edge. This gives us a complete weighted edge image for the exemplar. We show an example for blender in Figure 4 - displaying the original edge image, matched edge points, and final weighted edge image. In this figure, the colour map goes from white to black via yellow and orange. We can see here that the outer edges are red/orange and black, indicating edges present in both the exemplar and the non-exemplar image. We repeat this step by matching the exemplar to all the other images within the exemplar's cluster. In the end, the final weighted edge image is the mean of the weighted edge images calculated to each non-exemplar blender.

The final weighted edge image helps us design the abstracted version of the object. We use Poisson-blending [25] guided by the noise-free edges to create an abstracted illustration [23,31] by removing the features at higher levels. Finally, we add back the edges calculated in the previous step to give it a more defined and iconic look. The final icon for the blender shows an image which is almost a silhouette, but shows important details like the base, the jar, and the cap (Figure 4b).

5   Results and Evaluation

Figure 5 demonstrates the exemplar-finding algorithm. Here we use the problem of finding the average car within a collection of many cars. We find the following nine subcategories for a car, since these represent car categorization by shape: sedan, coupe, convertible, sports car, SUV, van, pickup truck, station wagon, and minivan. For each category we use *gist*-based affinity clustering to get a set of representative visual triggers. The user selects the best cluster(s) from each category, and finally we cluster together the top ranked images using our shape-context based clustering. This

Fig. 7. Icons for different concepts discussed in the evaluation section

gives us the average car and shows the utility of clustering at multiple scales. At the end of clustering, a sedan emerges as the average car.

Next, in Figure 6, we demonstrate our entire algorithm (exemplar-finding and abstraction) for the construction of taxonomy visualization. Here we seek to assign an icon to each level, instead of just to the leaves. We first calculate the icons for the leaf nodes using the algorithm outlined above. For the inner nodes with only leaf nodes as children, we form a collage of at most the top four children under it. This order can simply be calculated based on the popularity – for example, on Amazon.com, the sub-categories always appear in the order of popularity. For example, when someone selects the category "Small Appliances", they are probably looking for a Coffee maker. Further, when forming the collage, the more important categories are allocated more space. As we move further up the taxonomy tree, one node will have many subcategory trees underneath it. In order to keep the icons compact and representative, we just percolate the icon for the top sub-category upwards.

Our system requires user input and interaction at various stages of the VIE design. To get insight and feedback from multiple people, we had members of our lab interact with the system. The users identified different concepts they were interested in visualizing, and we built the I-Bridge using our system. Some interesting results (as shown in Figure 7) were:

- Success: Person climbing a ladder (via Google images)
- Oil spill: Images of the duck in the oil spill. This is an example where a current event highly modifies the most relevant icon.
- Renaissance: Triggers artists like Shakespeare and Michelangelo. We can represent the concept using the people, or the art created by them.
- Gothic: We can represent this concept by using an icon with a person in gothic attire (via Google Images), or by using examples of gothic architecture (via Synonym relation)
- Affinity: Its synonym *kinship* gives us images of family trees
- Countries: The associated terms and images give us the political map, flag, and landmarks (Taj Mahal for India, Great Wall for China) of the country.
- Thrill: Rollercoaster. This indeed represents a good icon for representing the experience of a *thrill.*

6   Conclusions and Future Work

In this paper, we presented an approach which can accomplish the goal of finding good visual information encodings for concepts we are interested in. This requires the integration of many fields – linguistics, vision, computer graphics, and user interfaces, with a human in the loop. Our framework has great prospects in the design of clip art for various applications, such as taxonomies, book illustrations, and the expressive augmentation of graphical node/link diagrams to make these much more engaging and informative. In future we plan to fully integrate our framework into a graph drawing engine, use abstractions more freely to summarize certain facts and attributes, and use compositions for compact visual story telling with context and 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 give further insight into personal preferences. We plan to evaluate both usability and performance, in a conjoint manner (using the approach in [9]) using three types of experiments – determining I-signs given textual concepts, choosing between two I-signs for a textual concept, and finally given an I-sign, choosing among two alternative textual concepts.

Nevertheless, it goes undisputed that not all concepts have good visual representations and encodings. This is particularly true for difficult non-object concepts such as 'worship'. For those concepts that do lend themselves well to visual encodings, we believe that the power of our approach is its ability to communicate possibly quite subtle differences much more efficiently than textual descriptions.

References

1. Beeferman, D.: Lexical discovery with an enriched semantic network. In: Proceedings of the ACL/COLING Workshop on Applications of WordNet in Natural Language Processing Systems, pp. 358–364 (1998)
2. Belongie, S., Malik, J., and Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4). 509-522 (2002)
3. Chuah, M.C. and Eick, S.G.: Glyphs for software visualization. In: 5th International Workshop on Program Comprehension (IWPC'97) Proceedings, pp. 183–191 (1997)
4. Coyne, B. and Sproat, R.: Wordseye: An automatic text-to-scene conversion system. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 487–496 (2001)
5. Datta, R., Joshi, D., Li, J., and Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. Lecture Notes in Computer Science 3953. 288-301 (2006)
6. Datta, R., Joshi, D., Li, J., and Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (CSUR) 40(2). 5:1-5:60 (2008)
7. Fellbaum, C. and others: WordNet: An electronic lexical database. MIT press Cambridge, MA (1998)
8. Frey, B.J. and Dueck, D.: Clustering by passing messages between data points. Science 315(5814). 972-977 (2007)
9. Giesen, J., Mueller, K., Schuberth, E., Wang, L., and Zolliker, P.: Conjoint analysis to

measure the perceived quality in volume rendering. IEEE Transactions on Visualization and Computer Graphics 13(6). 1664-1671 (2007)

10. Götze, M., Neumann, P., and Isenberg, T.: User-Supported Interactive Illustration of Text. In: Simulation und Visualisierung, pp. 195–206 (2005)
11. Götzelmann, T., Götze, M., Ali, K., Hartmann, K., and Strothotte, T.: Annotating images through adaptation: an integrated text authoring and illustration framework. Journal of WSCG 15(1-3). 115–122 (2007)
12. He, J., Tong, H., Li, M., Zhang, H.J., and Zhang, C.: Mean version space: a new active learning method for content-based image retrieval. In: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pp. 15–22 (2004)
13. Hoffman, D.D.: Visual intelligence: How we create what we see. WW Norton & Company (2000)
14. Horn, R.E.: To Think Bigger Thoughts: Why the Human Cognome Project Requires Visual Language Tools to Address Social Messes. New York Academy Sciences Annals 1013. 212–220 (2004)
15. Joshi, D., Wang, J.Z., and Li, J.: The Story Picturing Engine—a system for automatic text illustration. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) 2(1). 68-89 (2006)
16. Kovalerchuk, B., Brown, J., and Kovalerchuk, M.: Bruegel iconic correlation system. Visual and Spatial Analysis. 231–262 (2004)
17. Kovalerchuk, B.: Iconic reasoning architecture for analysis and decision making. In *Visual and Spatial Analysis*. Springer Netherlands. 129-152 (2004)
18. Lalonde, J., Hoiem, D., Efros, A.A., Rother, C., Winn, J., and Criminisi, A.: Photo clip art. In: ACM Transactions on Graphics (TOG), ACM (2007)
19. Leyton, M.: Symmetry, causality, mind. The MIT Press (1992)
20. Li, X., Wu, C., Zach, C., Lazebnik, S., and Frahm, J.M.: Modeling and recognition of landmark image collections using iconic scene graphs. ECCV .427–440 (2008)
21. Ogden, C.K.: Basic English: a general introduction with rules and grammar. K. Paul, Trench, Trubner (1944)
22. Oliva, A. and Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. International Journal of Computer Vision 42(3). 145-175, (2001).
23. Orzan, A., Bousseau, A., Barla, P., and Thollot, J.: Structure-preserving manipulation of photographs. Proceedings of the 5th international symposium on Non-photorealistic animation and rendering . 103-110 (2007)
24. Palmer, S., Rosch, E., and Chase, P.: Canonical perspective and the perception of objects. In: Attention and Performance IX, pp. 135-151 (1981)
25. Pérez, P., Gangnet, M., and Blake, A.: Poisson image editing. ACM Trans. Graph. 22(3). 313-318 (2003)
26. Raguram, R. and Lazebnik, S.: Computing iconic summaries of general visual concepts. In: Proc. of IEEE CVPR Workshop on Internet Vision, pp. 1-8, (2008)
27. Rother, C., Bordeaux, L., Hamadi, Y., and Blake, A.: AutoCollage. In: ACM Transactions on Graphics (TOG), ACM, pp. 847–852, (2006)
28. Setlur, V., Albrecht-Buehler, C., Gooch, A., Rossoff, S., and Gooch, B.: Semanticons: Visual metaphors as file icons. In: Computer Graphics Forum, pp. 647–656 (2005)
29. Simon, I., Snavely, N., and Seitz, S.M.: Scene summarization for online image collections. In: Proc. of ICCV, pp. 1–8, (2007)
30. Strothotte, C. and Strothotte, T.: Seeing between the pixels: pictures in interactive systems. Springer-Verlag New York, Inc. New York, NY, USA (1997)
31. Strothotte, T. and Schlechtweg, S.: Non-photorealistic computer graphics: modeling, rendering, and animation. Morgan Kaufmann Pub.(2002)
32. Zhou, X.S. and Huang, T.S.: Relevance feedback in image retrieval: A comprehensive review. Multimedia systems 8(6). 536–544, (2003)