COMMUNITY MAPS FOR JOINT VISUALIZATION OF IMAGES AND DESCRIPTORS

Raghavendra Singh

IBM Research India, Delhi
raghavsi@in.ibm.com

ABSTRACT

We propose a method to visualize a large image database. The key novelty is that the visualization maintains the structure often seen in clustering based visualization schemes while allowing for overlap between two clusters. Clustering is based on user-defined descriptors such as color, texture, and EXIF settings. Associations between images and descriptors are also made explicit in the visualization. Visualization of a 16000 image database downloaded from Flickr is presented.

Index Terms— Image databases, Data visualization, network theory

1. INTRODUCTION

An image database can be enriched by attaching a set of *descriptors* to an image. Descriptors include EXIF settings such as exposure and aperture that modern cameras can capture in an image header; text descriptors such as person, place, and events that users can use to annotate an image; and content descriptors like color, texture, structure that sophisticated algorithms can extract from an image.

The importance of *associating* descriptors with images, i.e., identifying the set of descriptors that best characterizes a set of images, has been well documented in literature, e.g., Qiu's work on co-clustering of images and content features [1], its extension by Gao et.al. to include textual descriptors [2], and our recent work on browsing image databases [3]. It has also been explored in text mining literature [4], where a structured 2-D matrix is used to visualize the associations between the documents and their keywords.

Systematic visualization of images in large databases is important for user interaction [5, 6, 7]. Heesch provides a classification of visualization models that broadly corresponds to *hierarchies* based on clustering, *networks* based on nearest neighbor, and *maps* based on preserving similarity measures [7]. In hierarchies such as [1, 2] images within a cluster are typically visualized together, separate from images of other clusters. The advantage of hierarchies is that typically they are structured – partitioned visualization space clearly identifies the clusters. The disadvantage is that its difficult to find perfect clusters – typically an image could belong to two or more clusters. This is overcome in maps such as [5], where similar images are visualized closer than non-similar images, with no clear partitioning of the visualization space. The disadvantage of maps is that in the low dimensional visualization space preserving similarity leads to overlap of some images while other images are laid out in isolation – typically there is a poor utilization of visualization space. Additionally, dimensionality reduction implies that associations between images and their descriptors are transformed and are hence not easily apparent.

The visualization proposed in this paper displays images in a 2D grid, the *community map*, such that,

- Images that are clustered together are clearly identified by placing them on the on-diagonal sub-grids.
- But it allows for overlap of up-to two clusters by innovative usage of off-diagonal sub-grids where images that lie in the overlap are placed.
- Moreover the sub-grids are labeled by the descriptors that are clustered together, thus clearly identifying the associations.

In our proposed method images and descriptors of a database are modeled as nodes of a bipartite network. An edge, between a descriptor node and an image node, has a weight that is the normalized value of the descriptor for the image. Community detection [8] on the network yields communities of images and descriptor, i.e., clusters of strongly associated images and descriptors. By perturbing the weights on the edges of the networks, slightly modified communities are detected, and the process unravels nodes that either belong to the core of a cluster, or lie on its border [9, 10]. These are then laid out in a 2-D grid at the appropriate location.

The paper is organized as follows – section 2 relates this work with prior work in this area. Section 3 reviews the community detection algorithm and presents a novel method for creating a visualization given a set of images and descriptors. Section 4 presents results using a database of around 16000 images downloaded from Flickr.

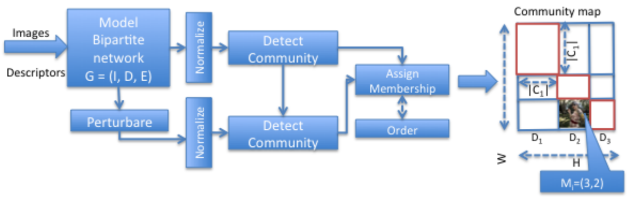

Fig. 1. Method to create proposed visualization. I, D represents images and descriptor as nodes in a bipartite network G. E is set of weighted edges between I, D nodes. W, H is user defined grid size. On-diagonal sub-grids are shown with an orange outline – their sizes are determined by the number of elements in community C_k. D_k are set of descriptors in community C_k – they label the horizontal axis. In the example the image has a membership value $(3, 2)$ i.e., belongs to the overlap of community C_2 and C_3 and hence is placed in the (3,2) sub-grid.

2. RELATION TO PRIOR WORK

- We extend previous work [1, 2] on clustering of images using their associations with user-defined descriptors. The main difference is that both strong intra-cluster and inter-cluster (up-to two) associations can be explored in our visualization. Qiu et.al. [11] have proposed visualization of images in order to "provide a mental picture of database content", but their algorithm and results are limited to color descriptors.
- We extend our previous work [3] substantially. The methodology of using community detection algorithm that requires only a null model and perturbation of communities to detect overlap for clustering based image visualization is novel and different from the graph-partitioning methods used in [1, 12, 3]. By novel usage of the off-diagonal grids we are able to present overlaps between two clusters. The results in this paper though not validated yet with a systematic user study are arrived at using a much larger dataset than reported in these previous works.

3. METHOD

The flowchart shown in Fig 1 illustrates the visualization method. We now describe each step in the method. Weight of an edge between nodes i_k and d_l is the value that descriptor d_l takes in image i_k. A value of zero implies that there is no edge between the nodes, else the weight could be binary, or integers/floats.

Normalize step normalizes the weights such that for each descriptor variance is one, and the mean is such that all non-zero weights are positive. Detect community step then partitions the graph into groups of nodes, with dense connections within groups and only sparser connections between them [8]. Community detection has been popular in network theoretic analysis [8, 13, 14]. Its advantage is that while typical graph partitioning algorithms such as [15] fix the sizes of the groups into which the network is divided, community detection, particularly one based on *modularity* does not fix the size of communities or number of communities [8]. Rather it uses the intuition that for an unweighted graph the number of edges between (within) communities should be smaller (greater) than *expected*. Let us denote the probability of an edge between nodes k and l by P_{kl}, and let g_k denote the community to which vertex k belongs. The adjacency matrix is denoted by A and its $(k, l)^{th}$ element by A_{kl}. If the number of edges is N_e, the expected number of edges between k and l is $A_{kl} - P_{kl}$, and modularity Q is defined as,

$$Q = \frac{1}{2N_e} \sum_{kl} [A_{kl} - P_{kl}] \delta(g_k, g_l) \tag{1}$$

$\delta(r, s) = 1$ if $r = s$ and 0 otherwise. The design choice is of P_{kl} which denotes the "null model" against which to compare our network - the strength of the modularity approach lies in that it makes this choice explicit. The simplest null model takes into account the expected degree of a node, but places the edges entirely at random. If k_i denotes the degree of the node i, then in this model [8], $P_{ij} = (k_i * k_j)/(2 * N_e)$. For bipartite graphs the definition is updated suitably [16], and for weighted graphs the degree is determined by the weights on the edges of the graph. This step yields a set of K communities $\{C_k\}$ each of which is set of strongly connected nodes $\{I_k, D_k\}$.

The ideal of perfectly separated communities may not be true for most real networks [17]. Overlapping communities can be considered to be equivalent to margins of separating hyper-planes in clustering of points in high dimensional space [18]. To detect nodes that lie in the overlap of two communities, first the Perturbare step perturbates the weights of the edges by adding noise to the weight of an edge. Noise is uniformly distributed over $[-\sigma E_{ij}, \sigma E_{ij}]$ for an edge with weight E_{ij} [9], σ is a parameter in $[0..1]$. Then the Detect community step finds communities in the perturbed network. Perturbation along with community detection is repeated multiple times. If the perturbed community set $\{\tilde{C}_{jk}\}$ of the j^{th} run is such that $\tilde{D}_{jk}$[1] $\neq D_k \; \forall k$, then the run and its communities are discarded, i.e. we only allow for small perturbations where the images can change communities but not descriptors.

To identify the two clusters that an image belongs to, the Assign membership step uses $\{C_k\} \forall k$ and retained perturbed communities $\{\tilde{C}_{jk}\} \forall j, k$. Let the top two communities by count that image i_l belong to be k_1, k_2. Then the membership value is $M_l = (k_1, k_2)$. Note that an image could belong to a single community despite of perturbations. These are the images that will be placed on the on-diagonal sub-grid.

[1] $\tilde{C}_{jk} = \{\tilde{I}_{jk}, \tilde{D}_{jk}\}$

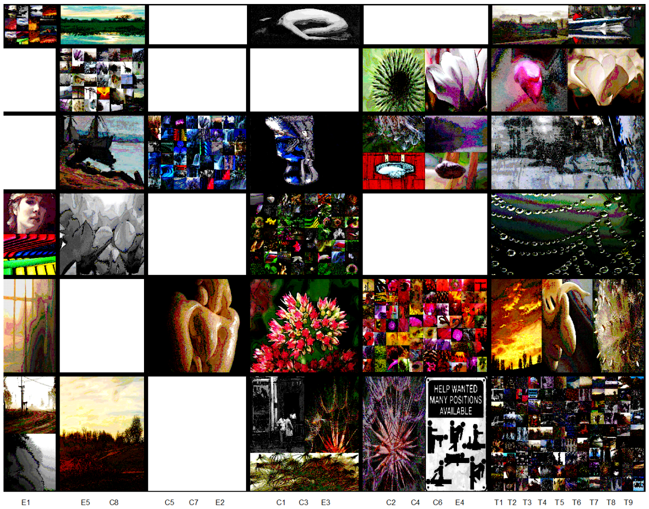

Fig. 2. Images from a single photographer. The horizontal axis is labeled from left-right on-diagonal sub-grid: $|E_1|$, $|E_5, C_8|$, $|C_5, C_7, E_2|$,$|C_1, C_3, E_3|$, $|C_2, C_4, C_6, E_4|$, $|T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9|$, where $|\ldots|$ represents communities.

The Order step is based on our previous work [3]. It reorders the communities such that those with higher overlap are close together in the order. The membership values are updated with the new order. Community map is a 2D grid which is partitioned into $K \times K$ sub-grids. The size of k on-diagonal sub-grids is determined by the size of C_k community, assuming an aspect ratio of 4/3. Images are placed in the grid corresponding to their membership value and resized according to the space available in the sub-grid – its equally divided between all images in the sub-grid. On the horizontal axis each sub-grid is labeled by the descriptors D_k corresponding to the on-diagonal community.

4. RESULTS

We downloaded, publicly available images from Flickr, a total of around 16,000 images from 24 different users. The selection of users was random among those whose images are available publicly for non commercial use. The dataset is very representative of real images ranging from portraits, natural scenes, structures to collages of photographs. The content descriptor that we have used consists of a 8 dimension RGB color histogram, a 5 dimension MPEG-7 edge orientation histogram, and a 9 dimensional vector of wavelet based texture features. In this descriptor $C1$ i.e., the first bin of color space primarily encodes dark images, while $C8$ encodes bright images. $E1$ counts the number of horizontal edges, $E2$ the number of vertical edges and $E3, E4, E5$ the diagonal and non-directional edges [20]. The texture features are ordered such that the highest decomposition level sub-bands occur first and lowest decomposition level sub-bands last. Five EXIF descriptors, ISO, aperture, exposure, focal length and flash, are used. For the first four descriptor there are three possible values, low (1), medium (2) , high (3), and flash is binary off or on.

Fig. 4 show the visualization of the 16000 images downloaded from Flicker. This visualization shows the structure that is brought about by a 2D grid layout. Due to the large number of images, it can be argued that detail of each image is lost. However the visualization does capture the aggregated associations with content descriptors for a community– observe the homogeneity of each on-diagonal sub-grid. Because there is no transformation one can clearly label the communities with the descriptors, allowing the user to understand what the descriptors “encode”, especially content based descriptors which are not extracted by a user but are often used in content based image retrieval leading to the semantic gap [5]. Additionally by using off-diagonal sub-grids we are able to show the overlaps between two communities – for example textured images with a reddish tinge or with a bluish tinge can be clearly seen in the last column off-diagonal sub-grids. A user can use this map as an initial guide that allows her to narrow down the search region for her multimedia query. In Fig 2 images of a single photographer with relatively small σ are shown. There are relatively few images in the overlap between communities. But this implies that these images can be shown at a larger resolution, emphasizing them as compared to images that strictly belong to a single community. Reddish texture images are highlighted in (4,5) sub-grid, while an image that has both horizontal edges and is dark is shown clearly in the (1,3) sub-grid. To us this is similar to emphasizing the marginal data in an active learning scenario [18].

This added feature of larger resolution images in the overlap region is immediately useful in Fig.3 where EXIF settings are used as descriptors. Here overlaps represent images where some of the settings were modified. Thus while this photographer usually shot portraits using a high focal length, (1,1) on-diagonal sub-grid, he also shot some portraits with lower focal length, and a higher aperture, (1,5) sub-grids. Both his preferences for settings for a particular type of photographs, and how the changes in settings affects the photographs is comprehensible from this visualization. In this case we did solicit user feedback from the photographer and received positive comments.

In conclusion, we present a novel visualization of an image database. Each visualization has its own merits and demerits. We have not been able to do an objective evaluation yet of our visualization. We do argue that we do not know of any work that effectively utilizes the 2D grid space. The demerits of our work is that images are small, the visualization is unable to show overlaps of more than two communities, and there is a lack of user interaction. We are working on the last two issues using interactive Venn diagrams.

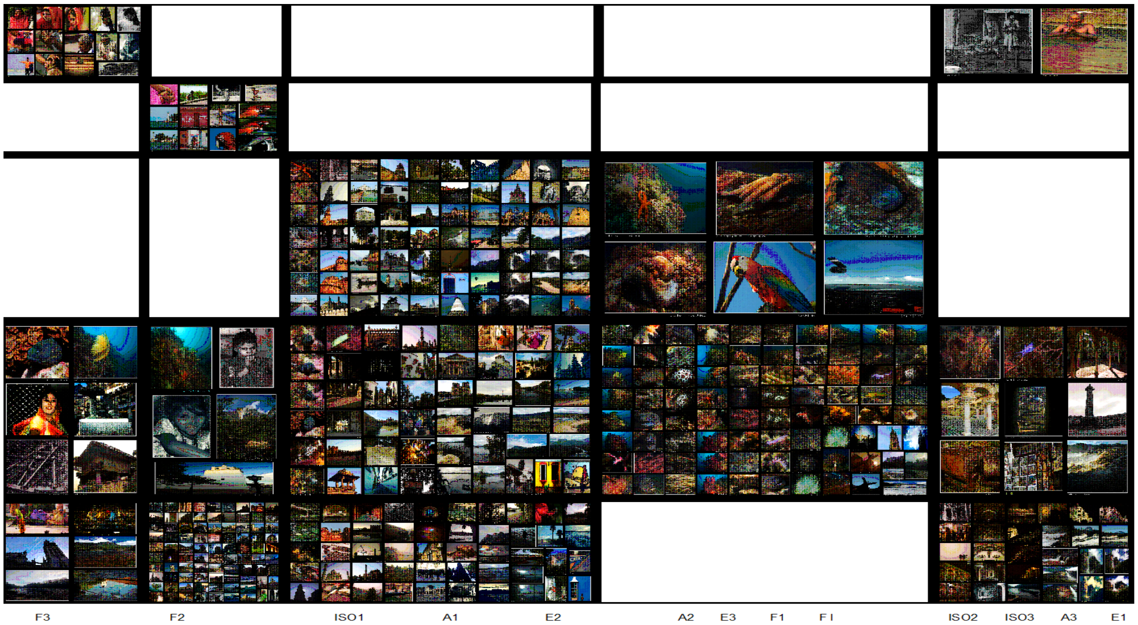

Fig. 3. Images of a different photographer whose EXIF tags are available. A stands for aperture, E for exposure, F for focal length, FL for flash. This map gives a very clear idea of what are the settings that a user prefers, and also what are the images when settings are changed slightly.

Fig. 4. All 16000 images downloaded from Flickr are organized in a community map. The horizontal axis labels are $|C8, E2, E3|$, $|C5, C7, E5|$, $|C1, C3, E1|$, $|C2, C4, C6, E4|$, and $|T1...T9|$ for each on-diagonal sub-grid from left to right. Communities are defined by bright vertical edges (top-left sub-grid), bluish tinged images, dark images with horizontal edge, reddish images, and texture images. Images that are textures and predominantly red in color are in the (4,5) sub-grid and those with bluish tinge are in the (2,5) sub-grid.

5. REFERENCES

[1] Guoping Qiu, “Image and feature co-clustering,” in *Proc. of 17th ICPR*, Washington, DC, USA, 2004, pp. 991–994, IEEE Computer Society.

[2] Bin Gao, Tie-Yan Liu, Tao Qin, Xin Zheng, Qian-Sheng Cheng, and Wei-Ying Ma, “Web image clustering by consistent utilization of visual features and surrounding texts,” in *Proceedings of the 13th annual ACM international conference on Multimedia*, New York, NY, USA, 2005, pp. 112–121, ACM.

[3] R. Singh, “Browsing an image database utilizing the associations between images and features,” in *Proceedings of ICIP 2009*, 2009.

[4] Deepayan Chakrabarti, Spiros Papadimitriou, Dharmendra S. Modha, and Christos Faloutsos, “Fully automatic cross-associations,” in *Proc. of the 10th ACM SIGKDD*, New York, NY, USA, 2004, pp. 79–88, ACM.

[5] S. Santini and R. Jain, “Integrated browsing and querying for image databases,” *IEEE Multimedia*, July-September 2000.

[6] G. P. Nguyen and M. Worring, “Interactive access to large image collections using similarity-based visualization,” *Journal of Visual Languages and Computing*, vol. 19, no. 2, pp. 203–224, 2008.

[7] Daniel Heesch, “A survey of browsing models for content based image retrieval,” *Multimedia Tools Appl.*, vol. 40, no. 2, pp. 261–284, 2008.

[8] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices,” *Physical Review E*, vol. 74, no. 3, pp. 036104+, Sep 2006.

[9] D. Gfeller, J.-C. Chappelier, and P. De Los Rios, “Finding instabilities in the community structure of complex networks,” *Physical Review E*, vol. 72, no. 5, pp. 056135, 2005.

[10] Brian Karrer, Elizaveta Levina, and M. E. J. Newman, “Robustness of community structure in networks,” *Physical Review E*, vol. 77, no. 4, pp. 046119+, Apr 2008.

[11] Guoping Qiu, Jeremy Morris, and Xunli Fan, “Visual guided navigation for image retrieval,” *Pattern Recognition*, vol. 40, no. 6, pp. 1711–1721, 2007.

[12] Yixin Chen, James Z. Wang, and Robert Krovetz, “Clue: Cluster-based retrieval of images by unsupervised learning,” *IEEE Transactions on Image Processing*, vol. 14, no. 8, pp. 1187–1201, Aug 2005.

[13] Gary William Flake, Steve Lawrence, C. Lee Giles, and Frans M. Coetzee, “Self-organization and identification of web communities,” *Computer*, vol. 35, no. 3, pp. 66–71, 2002.

[14] M. Rubinov and O. Sporns, “Weight-conserving characterization of complex functional brain networks,” *Neuroimage*, vol. 56, no. 4, pp. 2068–2079, 2011.

[15] George Karypis and Vipin Kumar, “A fast and high quality multilevel scheme for partitioning irregular graphs,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 359–392, 1998.

[16] M. J Barber, “Modularity and community detection in bipartite network,” *Phys. Rev. E*, vol. 76, no. 6, 2007.

[17] Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 7043, pp. 814–818, June 2005.

[18] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” *Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2001.

[19] Jianbo Shi and Jitendra Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[20] Dong Kwon Park, Yoon Seok Jeon, and Chee Sun Won, “Efficient use of local edge histogram descriptor,” in *Proc. of the ACM workshop on Multimedia*, New York, NY, USA, 2000, pp. 51–54, ACM.