International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

Generating Visual Abstract of Topological Areas using Neighborhood given Images- A Survey

Priya C[1], Supriya R[2], Jayashree L K[3]

[1,2]*PG Scholar, Dept. of Computer Science, Vemana Institute of Technology, Bangalore, Karnataka.*
[3]*Assistant Professor (Selection Grade), Dept. of Computer Science, Vemana Institute of Technology, Bangalore, karnataka*

[1]priyachowdaiah92@gmail.com, [2]rsupriya53@gmail.com, [3]jayashree_nov26@yahoo.com

***Abstract*— Here an innovative approach for automatic viewable abstraction of a topological area that accomplishes user-given images and related precise and inherent metadata collected from popular websites are presented. A limited number of illustrative but distinct images are searched to represent the area within a certain radius around a specific location. The approach is based on the arbitrary walk with starts over a graph that portrays the relation between images, optical features extracted from the images, related text, as well as the data from the uploaders. A simple but effective scheme for selecting the most ideal and distinct set of images based on the data derived from the graph is proposed.**

***Keywords*—Multimodal, optical abstraction of topological areas.**

I. INTRODUCTION

Availability of video capturing devices and economical image and as well as brisk advancement of content sharing websites and social networking has led to the creation of social media. In such situation, multimedia content is usually guided by user-generated metadata, such as title, summary, tags and comments. While the types of metadata can be indicated to as definite ones, inherent metadata can be derived as well, like for example those containing the data on the uploader and user relations derived from users' interactions with the images and their actions in a social network related to these images. An approach to automated formation of optical abstraction of topological areas using neighborhood-given images and related definite and inherent metadata is presented. The goal is to yield an optical abstraction of the area surrounding a given location, e.g., a landmark, restaurant or a mall, where the location is specified by its geo-coordinates (geotags). The abstraction should be as descriptive about the location, but at the same time as compressed as possible.

The approach is persuaded by the inference that a person deciding to visit a distinct location may be governed to a large extent by his influence about the surroundings of that location (e.g., when choosing a restaurant).Compressed and descriptive optical abstraction can help to improve time capability and effectiveness in communicating with typical interfaces for location suggestion and in using communicative map investigation tools to generate an influence about the location. Two examples of optical abstraction for the area around a location at Bangalore are shown in Fig. 1.(a) shows images of a single landmark (b), which includes various feature of the area, which includes the popular landmarks and also the non-mainstream ones, such as surroundings of the location. In the text that follows the locations, objects and events to be mainstream are taken into count. If they are found interesting by many users they appear often in the image collection.

(a)

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

(b)

Fig 1. Examples of optical abstraction of the area around the location at Bangalore, VidhanaSoudha a)photographs showing the landmark only and b) photographs showing various areas nearby the location.

Correspondingly, non-mainstream locations, objects and events that are found interesting by a smaller group of users they appear less frequent in the collection. The available social media assets are integrated using a graph-based model and to guide the process of optical abstraction creation by the data derived from the investigation of the model is chosen. Here a novel method is proposed for automatic weighing of the graph edges to give the contribution of each modality, namely optical features, text, and user relations, to the overall performance of the optical summarization algorithm.

In Section II related previous work is addressed and the rationale behind the approach is presented in Section III. The proposed algorithm for generating optical abstraction is discussed in Section IV. Section V describes the baseline approaches to produce experimental results that are presented in Section VI. The discussion in Section VII concludes the paper.

II. Related Work

Increasing uniqueness without related decline poses a huge problem in the diversity in photo retrieval [1]. The various approaches can be divided into several categories according to whether only, text associated with the images, visual content, users' activity statistics, automatically generated metadata or a combination of these resources are exploited.

One way is to perform image clustering in optical area of expertise and then select a representative of each cluster to be presented to the user which is called as optical diversification of image search results [3]. In diversifying the image retrieval results [4], two concepts are used one is current abundant count and diversity count. The current abundant count measures the data given by each image which are added to the result list. The diversity count measures the current area of the final image results list. The images tagged with topics that are rarely included in the image set are used more over regular topics widely spread throughout the collection. To calculate the current abundant score, first the similarities between output of a particular image search engine and images in the candidate set are calculated. Further it is calculated iteratively, by combining the current abundant count of its neighboring images. Furthermore, the image diversification based on the associated text only has been performed, while the optical content of the images is not taken into consideration which is suboptimal.

In order to encourage uniqueness, recall of the cluster was used, which measures the ratio of retrieved clusters in the top N results and the total number of possible clusters associated with a given search topic. However in the given approach, annotated images and user given images available are used. The quality and type of data in the web is often very different from the professional content. The optical abstraction target visual presentation of topological area within a given radius (e.g., 1 km from a given location) and the photos captured there might show a large number of points of interests, so the tag sets of relevant images might not even intersect.

In generating diverse and representative image search results for landmarks [2] a multimodal approach to selecting representative and unique image search results for landmarks are presented. Both the user given images and visual information in the images are used. The main objective is to show the surroundings of a given (e.g., landmark) location, but focus is on selecting the best views of the landmark itself.

In mining tourist information from user-supplied collections [6], an approach that influences the images in Flickr and the related metadata for discovery and investigation of a huge number of tourist trips is proposed.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

However, the approach does not speak about the problem of optical abstraction of a given topological area, it is mentioned because it relies on neighborhood given data. Finally the ideas used by implementing different approaches takes 10 tourist cities as destinations, it is not clear that the approach will be as reliable for smaller topological areas that are not often covered.

III. Rationale Approach

In the current section, the set of reasons underlying the given approach for creating an optical abstraction of a topological area is discussed in detail.

Firstly, the input into the optical abstraction algorithm is a location, e.g., a landmark which is given by its geographical location to a photograph as shown in Fig 2,. For the implementation purpose, the selection for the range of radius is restricted to around 1 km around the input position and chose the existing set of photographs based on their photographic location. In addition to many existing graph-based approaches for solving problems in social media retrieval [7]–[10], a graph-based model is chosen as the basis for the summarization algorithm.

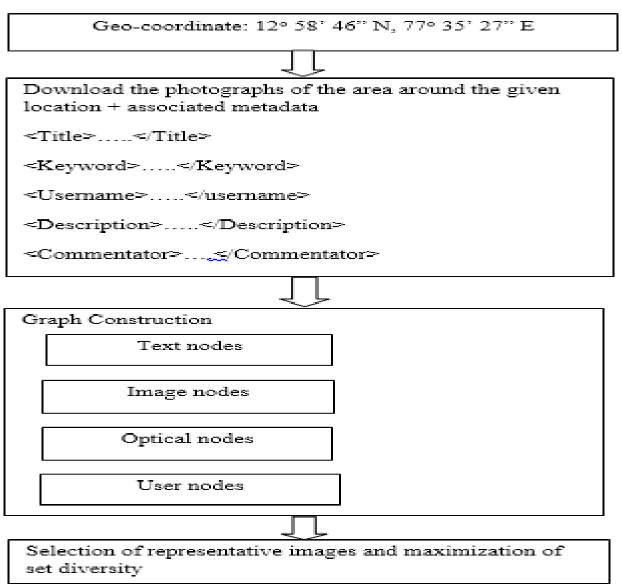

Fig 2. Illustration of location abstraction. The approach consists of three main steps (a) collecting the initial image set and related metadata (b)multi-modal graph construction and (c) using the graph to filter the initial image set for representative and diverse images.

A multi-modal graph consisting of several types of edges and nodes, is constructed in order to model the relation between associated explicit and implicit metadata as well as images captured within a certain radius from a given topological location. The initial stage of the approach, does not depend on the possible availability of geotags in neighborhood given photograph collections, the summarization algorithm is kept as general as possible and do not depend on geotags as a data when designing the graph model. Instead, the input depends on user-generated annotations, visual content of images and user interaction with the images and their activity in a social websites related to the images.

Although parameters like the image view count, can also be considered as an information resource to be used in the graph model, but it is not used when designing the summarization algorithm.

IV. Algorithm For Summarization

The four main algorithmic modules that are related to steps (b) and (c) in Fig. 2. are discussed here.

A. *Graph Construction*

Let $G = (V, E)$ be the undirected graph with the set of edges E and set of nodes V. The reason to choose undirected graph is that the relation between nodes are symmetric. The graph is illustrated as shown in the Fig 3.

1) *Nodes*: The graph consists of the following sets of nodes:

- *Image nodesI* = $\{i_1, i_2, \ldots i_n\}$: For the N images of a particular location an image node is given.

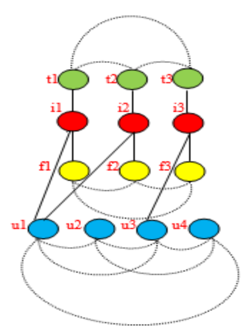

Fig. 3. Illustration of the proposed four-layer graph structure.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

- *Optical feature nodes* $F = \{f_1, f_2 \ldots f_N\}$: A visual feature node is added for each of N images in the initial set. The visual content of an image is represented with a vector composed of two low-level feature components, local color moments and Gabor texture features extracted over the 5 x 5 regular rectangular grid. It is also sensible to expand the feature vectors in the existing optical feature nodes with new feature components depending on the matching strategy and the similarity metrics used.
- *Term nodes* $T = \{ t_1, t_2, \ldots t_n\}$: For each image in the set a term node is added. To indicate images we accomplish description, user generated title and tags. The weights of those fields are not treated individually, but rather considered all the text related to an image to be a single document.
- *User nodes* $U = \{ u_1, u_2 \ldots, \ldots u_n \}$: For each of N_u users uploading a photograph related to a given location or commenting somebody else's image, a node is added called as user node.

The final set of nodes in our graph is equal to:

$$V = I \cup F \cup T \cup U \qquad (1)$$

2) *Edges:* There are two types of edges in the graph namely, similarity edges and attribute edges.

- *Similarity edges:* Similarity edges are represented with dashed lines in Fig. 3. This set of similarity edges link nodes of the same type.
- *Attribute edges:* Attribute edges are represented with solid lines in Fig. 3. An attribute edge is added between an image and each of its attributes—text nodes and optical feature. Note that a single user may upload multiple photographs, while some users do not have any uploads and their nodes are not linked to any image.

The edges linking optical feature nodes are weighted by the optical similarity count is calculated by

$$W_f(l, j) = sim(f_l, f_j) = \exp-\left(\frac{||fi - fj||^2}{2\sigma^2}\right) \qquad (2)$$

Where W_f is the weight of $N \times N$ weights of matrix.

To weigh the edges between user nodes the similarity count is taken, i.e., an image must be related to both the users for e.g., one of them must have uploaded the image and the other one must have commented on it. The corresponding user similarity measure is calculated as

$$W_u(l, j) = sim(u_l, u_j) = \frac{|I_l \cap I_j|}{|I_l \cup I_j|} \qquad (3)$$

Where I_l, $I_j \subset I$ are the set of images commented/ uploaded by the users u_i, u_j and W_u is the N_u x N_u user similarity matrix.

3) *Adjacency Matrix of the Graph*: The adjacency matrix A of the graph is illustrated in Fig. 4. It consists of the following submatrices:

- **II**$_{N \times N}$ = $[0]_{N \times N}$: Since the multi-modal similarities between image nodes i_l, $l = 1, \ldots N$ are not known, **II** matrix is filled with zeros.
- **IT** $_{N \times N} = \beta t\, I_N$: Weights on attribute edges linking text node t_l, $l = 1,\ldots N$ with the corresponding image nodes I_l are multiplied with the overall modality dependent weight βt.
- **IF**$_{N \times N} = \beta_f$ **I**$_N$: Weights on attribute edges linking optical feature nodes f_l, $l = 1,\ldots.N$ with the corresponding image nodes i_l are multiplied with the overall modality dependent weight βf.
- **IU**$_{N \times Nu}$ $(l, j) = \begin{cases} \beta u, & uj = uploader(il) \\ 0, & otherwise \end{cases}$: Attribute edges connecting uploader nodes uj , $j = 1,\ldots Nu$ with their corresponding image nodes il , $l = 1,\ldots.N$ are assigned the overall modality-dependent weight β_u.
- **TT** $_{N \times N} = \beta t$ W t: Weights of the edges linking text nodes
- .**FF**$_{N \times N} = \beta_f$**W**$_f$: Weights of the edges linking visual feature nodes.

	Image	Text	Visual features	Users
Image	II	IT	IF	IU
Text	IT'	TT	FT	TU
Visual features	IF'	FT	FF	FU
Users	IU'	TU'	FU'	UU

Fig. 4. The adjacency matrix A of the graph G.

- **FT**$_{N \times N}$ = $[0]_{N \times N}$: Since the multimodal similarities between text nodes and optical feature nodes are not known, the corresponding matrix is filled with zeroes.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

- **TU** $_{N \times Nu}$ = [0] N × N_u: Since the multimodal similarities between user nodes and text nodes are not known, the corresponding matrix is filled with zeroes.
- **FU**$_{N \times Nu}$ = [0] $N \times N_u$: Since the multimodal similarities between user nodes and optical feature nodes are not known, the corresponding matrix is filled with zeroes.
- **UU**$_{N u \times N u}$ = $\beta_u \mathbf{W}_u$: Weights of the edges linking user nodes.

The adjacency matrix A is column-normalized such that the values in each column sum to 1.

B. Choosing of Representative Images

Select the representative images in a specified location. First compute similarity that is multi-modal affinities between the items in graph, using plethora of methods. Here comparison and analysis of methods are not considered so, optioned random walk with restarts (RWR). RWR a well-known concept in retrieval of information with best application in the Google Page Rank algorithm which is being successfully used in tagging and image retrieval and for recommending collaboration. After this design choice will not affect the method of generality. So other state-of-the-art approach can be used for computing the similarity between items. Algorithm for selected image representation is to visualize a given geographical area the outline is as follows. First, initiate RWR through each image node, one at a time. To each step, a random walker selects randomly the outgoing edge or else decides to restart random walk with probability. Initially reporting the optimal value, but the later works by RWR in tagging image and retrieving the applications which is found and optimal value to be significantly higher. The stationary probabilities of node that are obtained after equation solving.

If the graph is large the inversion of matrix is practically infeasible or computationally intensive, efficiently which can be solved in an iterative manner. RWR is repeated for each image node present in graph by setting position in the restart vector to 1 and stationary probabilities are stored in image nodes and in the matrix column. Each pair of images, representing a multimodal similarity between them. The representative images must be salient, or it can be similar with other images captured in the vicinity of a specified location. Thereby, an arbitrary image compute the similarity summation of all other images in the graph.

While calculating do not take account image that is self-similarities, so that relatively high self-similarity values will enable and visual outliers to negatively affect the results. In the next steps sorting images according to the increasing value and defining representativeness score RS. Every image must be equal to the rank position of the image in the list which is sorted, then set the target image set that element must be optimal set then the image with the highest representativeness score is selected. The result of sorted images is according to their representativeness.

C. Maximization of Set Diversity

Next is to select most representative image for iterative fashion, The key point is that to make the resulting set of images as representative and diverse in the following step: To enforce these representativeness and diversity, the next selected image must be as different with that of the previously selected image(s) and with the same time a high representativeness score. Thereby, initializing RWR setting values in the restart vector at the position of already selected images or it is 0 otherwise. Stationary probabilities are computed with all nodes in the graph and the first value is stored. The stationary probabilities which reflect the similarity between every image in the graph. Further sort the elements in the decreasing order. For each and every image define diversity score which is equal to the position in the sorted list, such that the image must be similar to selected images and the image is least similar. Finally, select the image with the highest value among all the images that have not been included. The procedure in selection of images is repeated until the desired number that is diverse and representative images have been selected.

D. Baseline Approaches

In order to get the effectiveness of the approach with respect to the considered work, there are six baseline methods to compare the represented algorithm:

View Count: Selecting the images with the highest "view count"

Random: Images are selected randomly per location.

Visual Clustering: For visual feature representations K-means clustering is applied and to cluster the images into clusters. Centroid is selected to represent each cluster and the closest image is taken in the cluster.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

K-means is adopted as clustering mechanism which should be consistent with similar work on the basis of cluster approaches.Other clustering approaches are also experimented, that is affinity propagation clustering and hierarchical clustering. Due to the high dimensionality of the feature space and high visual diversity of user generated content the number of images per location is relatively small that is 100, on k-means clustering recently taken clustering algorithms do not provide appreciable improvement.

Number of Comments: In Flickr, for image appeal the number of comments is a useful indicator. In order to summarize the represented location, select the images with largest number of comments.

Assembling the Cluster: Multiple modalities may improve clustering performance. Initially, Visual Clustering is described then use K-means clustering to group images together into exactly clusters. Images are clustered independently on the basis of text and visual features. To merging individual cluster results single consolidated clustering, apply cluster ensemble framework. Ghosh and strehl demonstrated various scenarios on cluster ensemble also known as consensus clustering which yields results that are at least as good as the results of any individual clustering being merged.

After cluster creation, create the final visual summary by collecting images with respect to clusters, visual centroids.

MA Clustering: On basis of clustering, affinities that is which uses multimodal similarities between computed image nodes described and stored in matrix. Cluster the images with a well-known affinity propagation clustering approach. Affinity propagation clustering was proven effective in tasks and use here because it does not require feature vectors as input, but rather data point similarities between them should be there. Then image clusters are created, so now sort them in the descending order in terms of the number of elements. If that is larger than or equal to the size of the visual summary to be created, create the visual summary by simply sampling the exemplars of top-ranked clusters. Else, first select the exemplars for the detected clusters and then in an iterative fashion select remaining images. Now start with the top ranked cluster and centroid is selected, then proceed it by sampling the cluster which is next ranked in the same manner until images are selected.

Approach is intended to confirm the effectiveness of diversification strategy not only serve as the general visual summarization baseline. Selection of cluster representative and image clustering is the final results set for common diversification strategy.

V. DISCUSSION

A. Performance Security across Locations

Measuring the average performance over multiple queries is common, which is misleading, especially when exceptionally large improvement must be obtained for a limited queries but for other queries the performance decreases. So experiment chose to compare the performance for the proposed RWR-RD approach to the baselines. For each individual location, particular method yields the best results.

Additional advantage is evaluating the algorithm instead of using the values of MNPMF explicitly, which corresponds to the way users were asked to compare the quality of image sets, by saying which one is better suited among all the options. RWR-RD method clearly outperforms the baseline approaches in terms of diversity and representativeness.

B. Non-Mainstream Image Selection

This particular experiment evaluates whether the given approach is capable of selecting non-mainstream images that belong to the "long tail" of less popular, but potentially interesting part of image collection in the considered area. Most commercial systems fail, to present these images to the users and for this reason it is needed to expect smaller number of comments. To investigate the effectiveness can use the average number of comments per image in the generated description for evaluation criteria and method is compared with random, visual clustering, cluster ensemble and MA clustering baselines. With respect to location specified a particular approach selects the image set with lower average number of comments for particular image than the other approaches. View count baseline is not used because of high correlation between nr of comments and view count. RWR-RD-W clearly selects largest number of non-mainstream images in more locations than the baselines.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), Volume 5, Special Issue 2, May 2015)
International Conference on Advances in Computer and Communication Engineering (ACCE-2015)

VI. CONCLUSION

The effectiveness of the approach in generating a compact set of diverse and representative images of an area is discussed. Baseline approach gives average performance as well as the percentage of locations. Multiple modalities proposes RWR-RD-U for the optical abstraction task for computational complexity which is a critical parameter. The approach is capable of selecting not only mainstream images, but also the less popular images.

REFERENCES

[1] M. L. Paramita, M. Sanderson, and P. Clough, Diversity in photo retrieval: Overview of the ImageCLEFPhoto task 2009," in Proc. 10thInt. Conf. Cross-Language Evaluation Forum: Multimedia Experiments, 2010, pp. 45–59.

[2] L. S. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in Proc. 17th Int. Conf. World Wide Web, 2008, pp. 297-306.

[3] R. H. van Leuken, L. Garcia, X. Olivares, and R. Zwol, "Visual diversification of image search results," in Proc. 18th Int. Conf. World Wide Web, 2009, pp. 341-350.

[4] K. Song, Y. Tian, W. Gao, and T. Huang, "Diversifying the image retrieval results," in Proc. 14thAnnu. ACM Int. Conf. Multimedia, 2006, pp. 707–710.

[5] A. Popescu, G. Grefenstette, and P.-A. Moëllic, "Mining tourist information from user-supplied collections," in Proc. 18th ACM Conf. Inf.,Knowl. Manage., 2009, pp. 1713–1716.

[6] J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu, "Automatic multimedia cross-modal correlation discovery," in Proc. 10th ACMSIGKDD Int. Conf. Knowl. Discovery Data Mining, 2004, pp. 653–658.

[7] M. Clements, A. P. De Vries, and M. J. T. Reinders, "The task-dependent effect of tags and ratings on social media access," Trans. Inf. Syst., vol. 28, pp. 21:1–21:42, Nov. 2010.

[8] L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, and T. Huang, "A worldwide tourism recommendation system based on geotagged web photos," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2010, pp. 2274–2277.

[9] B. J. Frey and D.Dueck, "Clustering by passing messages between data points," Science, vol. 315, pp. 972–976, 2007.

[10] P.Over, G. Awad,M.Michel, J. Fiscus,W.Kraaij, and A. F. Smeaton, "TRECVID 2009—goals, tasks, data, evaluation mechanisms and metrics,"in Proc. TRECVID, 2010.