Review on Visual Summaries of Geographic Areas Using Community-Contributed Images

Rupali T. Waghmode
Computer Engg. Department Pune University
waghmodert@rediffmail.com

Bhavana R. Kanawade
Computer Engg. Department PuneUniversity
bhavana.kanawade@zealeducation.com

Abstract— In today's age of information technology variety of multimedia data is available. We acquire multimedia information from social media accompanied by user generated metadata. In this paper we give a new technique for automatic visual summarization of geographic area. We present a new retrieval model and learning framework for automatic visual summarization of geographic area. Retrieval model incorporate context dependent feature representations. Here we represent semantic relations between multimedia objects based on user interaction. In this technique we use random walk with restarts over a graph which models relations between images, visual features, explicit and implicit metadata. This new evaluation method does not require input from human.

Keywords - Multimedia, social media, multimodal fusion, visual summarization of geographic area, image set representativeness; image set diversity, automatic evaluation of visual summaries.

I. INTRODUCTION

Visual information represents audio, video, images, graphic arts etc. Visual summarization of geographic area gives brief but useful information regarding that geographic area. In today's age of information technology a large amount of multimedia data is available from social networking and content sharing websites. Community contributed knowledge and resources are becoming commonplace and represent a significant portion of the available and viewed content on web. The information provided by users is often inaccurate and noisy; photos are of varying quality; and sheer volume alone makes content hard to retrieve and represent. We present a new approach to overcome these challenges, using community contributed media to improve quality of representative images

Our approach in this paper uses the set of geo-referrrenced images on Flicker. The exact location of flicker images was automatically captured by the camera or alternatively, specified by the user. Flicker website supports this functionality and over 40,000,000 images are available. With the arrival of location-aware camera phones and GPS-integrated cameras, we expect the number of geotagged images on Flicker and other sites to grow rapidly. This multimedia content is accompanied by user generated explicit metadata such as title, description, tags and comments. Implicit metadata contains information on uploader and user relations inferred from users.

Here we present an approach to automatic creation of visual summaries of geographic areas using community contributed images explicit and implicit metadata. The aim is to generate visual summary of the area surrounding a given location. Location represents landmarks such as hotel or a museum, where the location is specified by its geo-coordinates. We outline a method that provides precise, diverse and representative results for landmark searches.

This is summarized as follows:

- Propose a novel approach for automatic visual summarization of geographic area.
- Search for limited number of representative but diverse images to represent the area within a certain radius around a specific location.
- Propose a novel evaluation protocol which does not require input from human annotators, but only exploits geographical co-ordinates.

II. LITERATURE REVIEW

A.Diversity Based Reranking for Rreordering Documents and Producing Summaries[6]

In this system Maximal Marginal Relevance criteria is used to maximize the relevance in retrieval and summarization. It reduces redundancy and maintains query relevance in reranking retrieved documents. The major merit of this technique is it produces non-redundant multi-document summaries, where MMR results are clearly superior to non-MMR passage selection. This techniques deal with only text summarization and not for other information such as visual information.

B.Automatic Multimedia Cross-modal Correlation Discovery [8]

In this system we find correlations across the media in collection of multimedia objects. It is used for automatic image captioning. It presents multimedia object like image and its attribute as nodes in graph and converts multimedia problem into graph problem. It use "random walk with restart ("RWR") for estimating the importance/affinity of node "A" with respect to node "B". The importance of node B with respect to node A is steady state probability $u_A(B)$ of random walk with restarts. To solve auto captioning problem for image estimate steady-state probabilities for all nodes of "MMG".

C.Visual Diversification of image search results[7]

In the visual diversification of image search results three clustering methods are used. The clustering algorithms are Folding, Maxmin, and Reciprocal election. In this case approach is dynamic one, which allows to dynamically weight the importance of the visual feature such as color, shape, texture. Folding is an approach that appreciates the original ranking, by assigning a larger probability of being a representative to higher ranked images.

D.Generating diverse and representative image search results for landmarks [3]

In this system we combine image analysis, tag data and image metadata to extract meaningful patterns from community-contributed datasets. We use tags (text labels associated with images by users) and location metadata to detect tags and location that represent landmark or geographic features. We apply visual analysis of images associated with discovered landmarks to extract representive sets of images for each landmark.Using various image processing methods, we cluster the landmark images into visually similar groups, as well as generate links between those images that contain the same visual objects. Based on the clustering and on the generated link structure,we identify canonical views, as well as select the top representative images for each such view.

E.Adaptive image retrieval using a graph model for semantic feature integration [2]

This system describes how semantic relations between multimedia objects based on user interaction can be learnt and then integrated with visual and textual features into a unified framework. The framework models both feature similarities and semantic relations in a single graph. Querying in this model is implemented using the theory of random walks. In addition, we present ideas to implement short-term learning from relevance feedback. Systematic experimental results validate the effectiveness of the proposed approach for image retrieval. In our model, images, terms, and visual features are represented as nodes in an Image-Context Graph (ICG). We propose a group-based contextual feature (peer information) based on mining usage information while searching in a multimedia collection. We show how the peer information can be integrated with already existing low-level visual features and textual annotation in a graph model. We compute the most similar images to recommend to user by using theory of random walk.

III Methodology for proposed work

A. Graph Construction

Let G=(V,E) is undirected graph with the set of nodes v and set of edges E.
The Graph has four layers

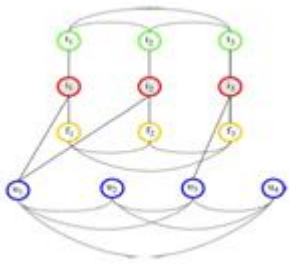

Fig.1 Proposed Four layer Graph structure[1] *Nodes:*

- Image nodes (I): I = {i1; i2; :::::i_N} For each of N images of a particular location an image node is introduced.
- Visual feature nodes (F):F = {f1; f2; :::::f_N} Visual content of image are represented by vector
- Term nodes (T): T = {t1; t2; :::::t_N} It contains title, description and tags for each image in set I and consider all text associated with an image to be an single set.Finally each image is represented by vector of TF-IDF.
- User nodes (U): U = {u1; u2; :::::U_N} users uploading an image related to a given location or commenting somebody else image.

2) Edges:

There are two types of edges:

- Attribute edges: It is an edge between image and each of its attributes such as visual feature, author,text nodes etc.
- Similarity edges: Similarity edges links nodes of same type.
  Visual similarity score computation using Gaussian Kernel:
  $W_f(l,j)=sim(f_l,f_j)= \exp(-||f\ l- fj|| / 2\sigma^2$
  Where W_f is NxN matrix of weights.
  User similarity score:
  Wu(l,j)=sim(ul,uj)=|Il∩ Ij|/|IlUIj|

3) Adjacency Matrix of Graph:

The adjacency matrix A of graph G consists following submatrices:

- $II_{NxN}=[0]_{NxN}$ Since the multi-modal similarities between image nodes are not known, II matrix is filled with zeros.
- $IF_{NxN}=B_f I_N$
  Weights on attribute edges linking image nodes, with the corresponding visual feature nodes are multiplied with the overall modality-dependent weight. Note that is an identity matrix.

	Image	Visual Features	Text	Users
Image	II	IF	IT	IU
Visual Features	IF^T	FF	FT	FU
Text	IT^T	FT^T	TT	TU
Users	IU^T	FU^T	TU^T	UU

Fig. 2 Adjacency Matrix A for Graph G

- $IT_{NxN}=B_t I_N$: Weights on attribute edges linking image nodes il, with the corresponding text node tl , l={1,……,N}are multiplied with the overall modality-dependent weight B_t.
- $IU_{NxNu}(l,j)=\{ B_u$ or 0 :Attribute edges connecting image nodes with their uploaders' nodes are assigned the overall modality-dependent weight . Remember that the total number of users in our system is and as users we consider both uploaders (authors) and commentators (users commenting an image).
- $FF_{NxN}=B_f W_f$: Weights of the edges linking visual feature nodes.
- $FT_{NxN}=[0]_{NxN}$: Since the multimodal similarities between visual feature nodes and the text nodes are not known, we fill the corresponding matrix with zeros.
- $FU_{NxN}=[0]_{NxNu}$: Since the multimodal similarities between visual feature nodes and the user nodes are not known, we fill the corresponding matrix with zeros.
- $TT_{NxN}= B_t W_t$: Weights of the edges linking text nodes.
- $TU_{NxNu}=[0]_{NxNu}$: As the multimodal similarities between text nodes and the user nodes are not known, we fill the corresponding matrix with zeros.
- $UU_{NuxNu}= B_u W_u$: Weights of the edges linking user nodes.

In this way values for all submatrices are calculated. Adjacency matrix A is column-normalized such that the values in each column sum to 1.

B. Selection of Representative Images

To select representative images for given location, multimodal similarities between items in the graph need to computed first. For this purpose we use Random Walk with Restarts (RWR) over graph. We initiate RWR from each image node in graph one at a time.It use "random walk with restart ("RWR") for estimating the importance/affinity of node with respect to other node. The importance of node with respect to other node is steady state probability of random walk with restarts. The computation of steady state probabilities is very important.

Suppose we want to do RWR for each node in graph, we need to compute the steady state probability

Let

A - Adjacency Matrix of Graph is column normalized

α- Probability of restarting the random walk from any node

v-restart vector where all the values in restart vector are set to 0 except position of starting image node.

$$p=(1-\alpha)Ap+ \alpha v \quad (1)$$

OR

$$P=\alpha (I - (1-\alpha) A)^{-1} v \quad (2)$$

We repeat RWR for each image node i_l in the graph by setting lth position in the restart vector to 1. Store the probabilities of image nodes in lth column of matrix

$S=[S_{lj}]_{N\,X\,N}$

$\{il,ij\} \in I$, S_{lj} represents the multimodal similarity between them.

For an arbitrary image il we compute sum of similarities to all other images in graph

$$ql=\sum_{\substack{j=1..N \\ j\neq l}} Slj$$

q=[q1,q2,…….,qN)

By sorting the images by increasing q value and define the representative score RS for each image.We add the image with highest representative score to OS (optimal set).

C.Maximization of set Diversity

Image selection for OS should be such that images should be dissimilar as possible with previously selected images.We initialize RWR setting values in restart vector v to 1/|OS| in the position of already selected images and otherwise.Stationaries probabilities of all nodes are computed using(2) and first N values of p are stores in

$p(OS)=[p_1^{(OS)},p_2^{(OS)}).....P_N^{(OS)}]$

Further we sort elements of p (OS) in decreasing order and define diversity score DS. Finally, we select the image with RS*DS value. This procedure is repeated until the desired number of Nr of representative and diverse images are selected.

IV.CONCLUSIONS

Thus we have overlooked various steps for generating visual summaries of geographic areas. We studied preparation of Image Context Graph (ICG) that encodes visual, textual and peer features together. The theory of random walks is employed to compute retrieval results. By using RWR over graph representative score (RS) and diversity score (DS) is calculated

REFERENCES

[1] Stevan Rudinac, Alan Hanjalic , “Generating Visual Summaries of geographic Areas Using Community-Contributed Images,” IEEE Transaction on Multimedia, val. 15. No. 4, june 2013.

[2] J. Urban and J.M. Jose, “Adaptive image retrieval using a graph model for semantic feature integration,” in *Proc. 8th ACM Int.WorkshopMultimedia Inf. Retrieval, 2006, pp. 117–126.*

[3] L. S. Kennedy andM. Naaman, “Generating diverse and representative image search results for landmarks,” in *Proc. 17th Int. Conf. World Wide Web, 2008, pp. 297– 306.*

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

[5] M. L. Paramita, M. Sanderson, and P. Clough, “Diversity in photo retrieval:Overview of the ImageCLEF Photo task 2009,”*in Proc. 10th Int. Conf.*

[6] J. Carbonell and J. Goldstein, “The use of mmr, diversity-based reranking for reordering documents and producing summaries,” in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf., Retrieval,*1998, pp. 335–336.

[7] R. H. van Leuken, L. Garcia, X. Olivares, and R. Zwol, “Visual diversification of image search results,” *InProc. 18th Int. Conf. World WideWeb,* 2009, pp. 341–350.

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

[9] M. Sanderson, J. Tang, T. Arni, and P. Clough, “What else is there? search diversity examined,” in *Proc. 31st Eur. Conf. Inf. Retrieval,* 2009, pp. 562–569.