Survey on Use of Crowd sourced User Preferences for Visual Summarization of Image Collection

[1] Rupali Tanaji Waghmode,[2] Prajakta Subhash Shinde
[1]Assistant professor,[2]Assistant professor
Computer Department
Zeal College of Engineering, Pune,India

***Abstract*—We present an algorithm for selecting suitable images for inclusion in visual summaries. This technique is designed on the basis of insights regarding how human perform visual summarization. For this purpose we use Amazon Mechanical Turk online Crowdsourcing Service. This algorithm use content and context of images, image sentiments, aesthetic features. We present a new solution based on text summarization and machine translation. We utilize pair wise ranking algorithm Rank.SVM for better performance. The studies conducted on a collection of geo-referenced Flickr image collections demonstrate the importance of our image selection approach.**

***IndexTerms*— Crowdsourcing, aesthetic features, user-informed image selection**

I. INTRODUCTION

Quick growth of social networking and content sharing websites has given rise to amount of digital data. There is need of powerful mechanism for representation, summarization, analysis and abstraction of data for more efficient browsing and retrieval. Summarization technique gives a concise representation of multimedia data document or data collection. Based on requirement summaries may consist of text, images, and video segments. In this paper we concentrate on visual summaries. Visual summaries contribute to abstract a video, set of videos, an image collection, video segments. Existing techniques of visual summarization have been guided by representativeness, relevance, diversity of information included in summary. Exact information on how humans perform visual summarization has rarely been considered while developing summarization technique. The insights obtained are insufficient to provide guide-lines for developing a strong visual summarization technique. There is a necessity of the specific criteria reflecting the user's perception of the summarization quality must be identified and applied to guide the summarization algorithm. We can say that this summarization algorithm should be created in such a way that user can easily understand it, and after that it will be successful. We use generation visual summaries of geographical places as trial use case to show the potency of. the proposed user-informed picture selection concept. The paper makes the following benefits:

- This paper present a new approach based on how humans select images for visual summaries, which was collected with a large-scale crowd sourcing study, as the basis for a novel method for automatically selecting images for visual summarization.

II. RELATED WORK

Summarizing Tourist Data Using Bidirectional Similarity

In [6] We present a new technique to summarization of visual information such as images or video based on similarity measure. This technique depends on optimization of similarity measure. Visual data can be images, videos, etc. Summary should capture essence of input data. It should not produce new visual artifacts. Bidirectional similarity measure satisfies both the properties. Two signals are measured as visually similar only if all the patches of 1st image are present in 2nd image and vice versa. We can use this methods for solving more problems like synthesis and completion of visual data, image collage, reshuffling of images and many more.

Generating diverse and representative search results for landmarks

In [9] we conjecture tag data and images explicit and implicit metadata, image analysis to extract meaningful features from community-contributed datasets such as Flickr. We use tags i.e. text information associated with images by users and location metadata to detect tags and location that represent landmark or geographic features. We perform clustering on the landmark images into visually similar groups by using various image processing methods, 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.

Summarizing tourist destinations by mining user-generated travelogues and photos

In [1] we represent a structure of summarizing tourist destinations by leveraging the rich textual and visual information in large amount of user-generated travelogues and photos on the Web. The structure first discovers location representative tags from travelogues and then selects relevant and representative photos to visualize these tags. The learnt tags and selected photos are finally arranged appropriately to provide an informative summary which describes a given destination both textually and visually. Experimental results based on a large collection of travelogues and photos shows the potency of destination summarization approach.

Generating Visual Summaries of geographic Areas Using Community-Contributed Images

In [2] we give a new technique for automatic visual summarization of geographic area. We present a new retrieval technique and framework for automatic visual summarization of geographic area. Here we take geo-coordinates of particular location as input. Based on the input data down-load images within set of 1 km radius from Flickr website. It requires metadata, textual and visual modalities of images. Then we represent semantic relations between images different modalities. In this technique System uses Multimodal Graph which combines visual, textual and other modalities together. The theory of random walks i.e. RWR over graph is used to compute multimodal similarities between nodes in graph. The representative score (RS) and diversity score (DS) is calculated.

VISUAL SUMMARIZATION

The Visual summarization system is having three important building blocks. Croswdsourcing Experiment, Feature Extraction and Ranking by RankSVM these are main blocks in user informed visual summarization system.

Crowdsourcing Experiment

In [11] we study Amazon Mechanical Truck online crowd-sourcing service. This system deals with supply of tasks that require human intelligence to do it. It is an online web service provide labor market where workers are recruited by employers for the execution of tasks called as HITs in exchange for a wage. Both workers and requesters are known to each other although responses size of the reward. Workers can read brief details and see previews of the tasks before accepting . Once a worker has finishe, a task, the employer who supplied that task pay him.

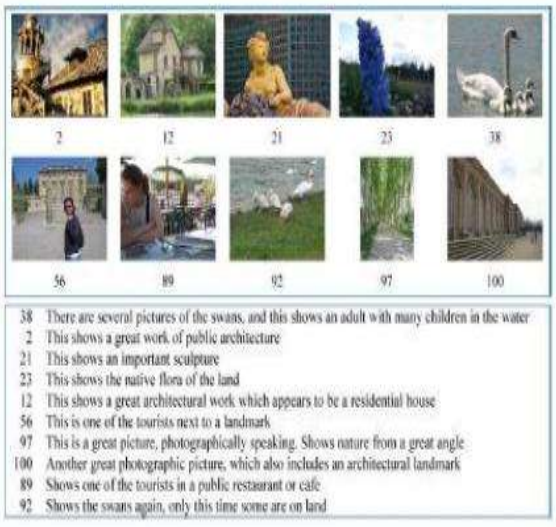

Fig.1. Example of Manually generated Reference

Summmaries Using Crowdsourcing [5]For manually creating reference summaries they recruit 20 workers for each location. We take 100 images of each location and display them in 10 rows with 10 images each. The task given is assignment of description to each image. Out of 100 images 10 images should be selected in summary. One of the example is given in the above Fig.1.

Feature Extraction

By taking guidelines derived from crowdsourcing study, we first define categories of features we extract from the images in collection. We describe each image a with feature vector x_a based on its importance, popularity, aesthetic appeal and sentiment observed.

- *Clustering based on geo-coordinates:*

  We use affinity propagation clustering method. This method takes measures of similarity between pairs of data points as input. Here we make affinity propagation clustering to cluster images into a certain number of geo-clusters. The input to this algorithm is the similarities between the images computed as follows:-

  $$S_g(a,b)=sim(g_a,g_b)=e^{-\delta(lat\,a,lona,latb,lonb)} \quad \ldots(1)$$

  Where $\delta(lat_a, lon_a, lat_b, lon_b)$ is a great circle distance.

- *Sentiment Analysis*
  - In [8] we present a methodology for determining polarity of text within a multilingual structure. The method relies on lexical resources for sentiment analysis available in English (SentiWordNet). We implement methods to automatically determine the polarity of sentences. SentiWordNet, is a.

lexical resource for opinion mining . In SentiWordNet to each synset (sets of synonyms of WordNet) a polarity scores is given i.e., a positivity, negativity and objectivity score. The addition. of these scores is always 1.A document in other than English language is translated into English using translation software. Then, the translated document is classified based on score into positive and negative. In sentiment classification, a document is searched for sentiment bearing words like adjectives. In SentiWordNet, scores for positivity and negativity are assigned to words. For example the 0, 1, 0 i.e. (positivity, negativity, objectivity) score is assigned to the synset of the term bad. After adding sum of all scores of this synset is 1.

- *Image Popularity*

  In content sharing websites such as Flickr number of comments posted on image plays important role. It is usually easy to retrieve and does not require any additional computational cost. Mean and variance of the number of comments posted on image is also calculated.

Ranking By RankSVM

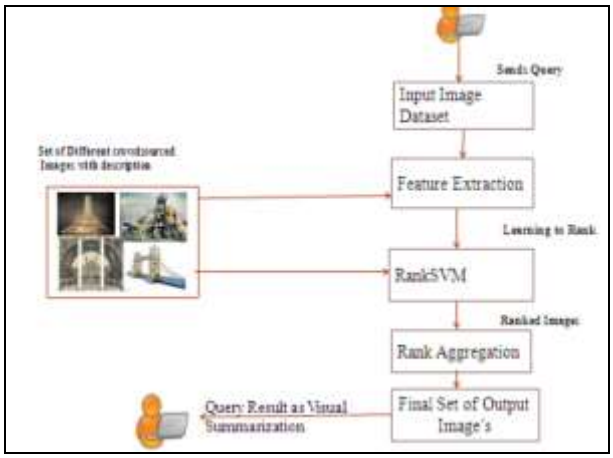

Fig.2 Visual Summarization System Block Diagram

Ranking methods have become important in machine learning and data mining in recent generation. There are number of ranking algorithms such as RankSVM, RankBoost and RankNet. We are focused on pairwise ranking accuracy that is fraction of correctly ranked pairs of objects. Biparitite means having two parts or arguments. RankSVM is pairwise ranking technique. It is used for designing ranking models. In [10] we RankSVM model is based on minimizing the following objective function. To select the images suitable for inclusion in visual summary we should calculate ranked list of images per location. We select training data set from crowdsourcing study. We select a set of image preference, (a,b) Є P, each consisting of a top ranked and bottom ranked image. Then, to learn the ranking model, a well-known RankSVM method is used. In the method proposed by Joachims in [8] the RankSVM model is based on minimizing the following objective function.

$$\mathbf{1.1}\quad \frac{1}{2}||W||^2 + c\sum_{(a,b)\in P} l(W^T x_a - W$$

...(2)

Here x_a and x_b are the feature vectors representing images, c is regularization parameter and *l* loss function. SVMLight is only publicly available software for implementing RankSVM. It requires to compute all possible difference vectors x_a, x_b.. Due to this RankSVM is slow and performace is low. Therefore there is need of fast RankSVM method. The performance of RankSVM is inferior due to incomplete training with it. In [7] we use primal Newton Method to obtain efficient performance. This method is used to speed up RankSVM training

.

CONCLUSION

We have collected information with a large scale crowd-sourcing experiment regarding how humans select images for visual summaries. We use this information to extract important features from image collection. Then we train fast RankSVM method based on Newton Optimization this reduce the training time. This ranking produce a list of images ranked by their suitability for inclusion in a summary.

REFERENCES

[1] T. Joachims, "Optimizing search engines using clickthrough data,"in Proc. 8th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, ser. KDD '02. New York, NY, USA: ACM, 2002, pp.133–142

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach", in Proc. 9th Eur. Conf. Computer Vision.Springer 2006, pp. 288-301

[3] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008, vol. 0, pp. 1-8.

[4] L. S. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in Proc. 17th Int. Conf. World Wide Web, ser. WWW08. New York, NY, USA: ACM, 2008, pp. 297-306.

[5] K. Denecke, "Using SentiWordNet for multilingual sentiment analysis", in Proc. IEEE 24th Int. Conf. Data Eng. Workshop, 2008 (ICDEW 2008), Apr. 2008, pp. 507512.

[6] O. Chapelle and S. S. Keerthi, "Efficient algorithms for ranking with SVMS,"Inf. Retriev., vol. 13, no. 3, pp. 201–215, Jun. 2010

[7] S. Siersdorfer, E. Minack, F. Deng, and J. Hare, "Analyzing and predicting sentiment of images on the social web," in Proc. Int. Conf.Multimedia, ser. MM 10. New York, NY, USA: ACM, 2010, pp.715-718.

[8] Y. Pang, Q. Hao, Y. Yuan, T. Hu, R. Cai, and L. Zhang, "Summa-rizing tourist destinations by mining user-generated travelogues and photos,"Comput. Vis. Image Understand., vol. 115, no. 3, pp. 352-363, Mar. 2011.

[9] G.Kazai, "In search of quality in crowdsourcing for search engine evaluation,"in Advances in Information Retrieval, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2011, vol. 6611, pp.165-176.

[10] S. Rudinac, A. Hanjalic, and M. Larson, "Generating visual summaries of geographic areas using community contributed images," IEEE Trans. Multimedia, IEEE 2013

[11] Stevan Rudinac, Martha Larson, and Alan Hanjalic, "Learning Crowd-sourced User Preferences for Visual Summarization of Image Collec-tions" 1520-9210 2013 IEEE