Effective Summarization of Large Collections of Personal Photos

Pinaki Sinha
Dept of Computer Science
University of California, Irvine
psinha@ics.uci.edu

Sharad Mehrotra
Dept of Computer Science
University of California, Irvine
sharad@ics.uci.edu

Ramesh Jain
Dept of Computer Science
University of California, Irvine
jain@ics.uci.edu

Categories and Subject Descriptors

H.3.3 [**Information Storage Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Design, Experimentation

Keywords

summarization, diversity, coverage, optimization

1. INTRODUCTION

The amount of personal photos uploaded to social networks (e.g., Facebook, Myspace etc) and photo sharing sites (e.g., Flickr, Picasa etc) has been increasing rapidly. According to current estimates, three billion photos are uploaded on Facebook per month [2]. Current photo hosting systems allow users to arrange their personal photos in albums. Any information need requires the user to drill down through the entire collection of photos, using the album / directory structure. This manual browsing may be tedious and inefficient. In this research, we propose a framework for generation of overview summaries from large personal photo collections. These summaries are representative subsets of the larger corpus and try to capture the relevant information, given the size constraints. They will enable users to get an overview of the interesting information in the photo collections without skimming through the entire database. Personal photo summarization may be a very subjective process. We do not intend to create a unique summary subset from a photo corpus. Rather, our goal is to design a framework for automatic generation of *informative overview* of the collection.

Some research has been done to generate summaries of popular scenes and landmarks using web image collections. Simon et al. [3] represent a natural scene using features which correspond to 3-D points in real world. Their summarization method selects a set of canonical views based on these features. Jaffe et al. [1] use a hierarchical clustering method to generate summaries of geotagged photos at multiple resolutions. This paper differs from previous work based on the goals (personal photo summary rather than multi-user scene summary), the techniques adopted (defining properties of a photo summary), algorithms used and the evaluation process.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

2. PROBLEM FORMULATION

SUMMARIZATION: Let the photo collection $\mathbf{P}$ be a set of $\mathbf{N}$ photos, $\mathbf{P} = \{\mathbf{p_1}, \mathbf{p_2}, \ldots \mathbf{p_N}\}$. The summarization problem is to find a set $\mathbf{S}$ (with $\mathbf{S} \subset \mathbf{P}$ and $|\mathbf{S}| \ll |\mathbf{P}|$), which represents $\mathbf{P}$ in an effective manner. There are $\binom{N}{M}$ possible summaries of size M for collection of size N, which is exponentially large for any reasonable M and N. However, only a few of them will be effective summaries. In this section we propose properties which determine an effective summary and define models to compute them.

To define the properties of a summary, we make use of three basic characteristics of photos in a personal collection. First, for each photo, we associate a notion of interestingness, denoted by $Interest$. It is an inherent property of a photo which determines its attractiveness to a subject. Second, we define a notion of distance $Dist(\mathbf{p}_i, \mathbf{p}_j)$ which given a pair of photos, determines the distance between them. Finally, we assume that personal photos have a set of semantic concepts which can be extracted from the raw data. Given a photo, a collection and a set of concepts present in personal photos we define a set $Represent(\mathbf{p}, \mathbf{P})$ which denotes the set of photos and concepts in $\mathbf{P}$ which are represented by $\mathbf{p}$.

A photo summary should be interesting or attractive to the subject. We define the metric Quality which determines the aggregate interestingness of a summary as: $Qual(\mathbf{S}) = \sum_{\mathbf{p}\in\mathbf{S}} Interest(\mathbf{p})$. Diversity of a summary ensures that it contains minimum redundant information. Diversity of the summary can be modeled as an aggregation of the mutual distances of the photo pairs: $Div(\mathbf{S}) = \min_{\mathbf{p_i},\mathbf{p_j}\in\mathbf{S},\mathrm{i}\neq\mathrm{j}} Dist(\mathbf{p_i}, \mathbf{p_j})$. A summary should be a good representative of the larger corpus it is created from. Coverage of a summary is computed by the aggregating the $Represent$ values of each of its photos. $Cov(\mathbf{S}, \mathbf{P}) = \mid \bigcup_{\mathbf{p}\in\mathbf{S}} Represent(\mathbf{p}, \mathbf{P}) \mid$.

We model summarization as a multiobjective optimization function $\mathcal{F}$ which jointly maximizes the above properties.

$$\mathbf{S}^* = \arg\max_{\mathbf{S}^*\subseteq\mathbf{P}} \mathcal{F}(Qual(\mathbf{S}^*), Div(\mathbf{S}^*), Cov(\mathbf{S}^*, \mathbf{P})), \quad (1)$$

where $|\mathbf{S}^*| = M$. $\mathcal{F}$ combines the individual properties to generate a single effectiveness metric. Many such functions can be defined by combining the properties in different ways.

3. OUR FRAMEWORK

We assume that each photo in the collection contains a host of contextual data in addition to the pixels. These include timestamps, EXIF parameters, geo-tags and community induced text data. We use this multimodal data to define a semantic concept space (that includes people, event and location names) and compute the basic photo characteristics. Due to lack of space, we could not provide the extraction and computation procedure.

The summarization objective stated in Equation 1 is a classical multi-objective optimization problem. Multi-objective (MO) problems are traditionally solved by converting all objectives into a single objective (SO) function by aggregating them. We can formulate aggregation by assigning different weights to Quality, Diversity and Coverage objectives and combining them in a linear way. Thus the reformulated summarization objective is:

$$\mathbf{S}^* = \arg\max_{\mathbf{S}^* \subset \mathbf{P}} [\alpha Qual(\mathbf{S}^*) + \beta Div(\mathbf{S}^*) + \gamma Cov(\mathbf{S}^*, \mathbf{P})] \quad (2)$$

Every choice of the weights α, β, and γ will generate a different summary which may show a different overview of the collection. Note that optimization of Div and Cov is NP-Hard (they can be mapped to Max-Min Dispersion and Maximum Coverage problems respectively). Exact solution of equation 2 is computationally inefficient. Instead, we adopt a greedy heuristic which finds the subset summary (S) that produces the best aggregate $Qual$, Div and Cov at every summary size (k).

Algorithm 1 Greedy Algorithm for Summarization

1: Initialize the set $S = \emptyset$
2: Compute $Qual(\mathbf{p})$ and $Cov(\mathbf{p})$ $\forall$ $\mathbf{p} \in \mathbf{P}$
3: Find $p^* = \arg\max_{p \in \mathbf{P}} [\alpha Qual(p) + \gamma Cov(p, \mathbf{P})]$
4: $S = S \bigcup p^*$
5: Recompute Cov based on concepts covered by p^*.
6: **while** Length(S) < k **do**
7: $p^* = \arg\max_{p \in \mathbf{P} \setminus S} [\alpha Qual(p) + \beta Div(p \cup S) + \gamma Cov(p, \mathbf{P})]$
8: $S = S \bigcup p^*$
9: Recompute Cov based on concepts covered by p^*.
10: **end while**

4. EVALUATION METHODOLOGY

To evaluate the representativeness of the summary, we compare the information content of the summary with that of the larger photo collection using Jensen Shannon Divergence (JSD). We model the original photo collection $\mathbf{P}$ and a candidate summary $\mathbf{S}$ as probability distributions over the multidimensional concept space. Let the distributions be denoted by $Prob_P$ and $Prob_S$ respectively. The degree of informativeness of summary $\mathbf{S}$ can be represented as: $Inform(\mathbf{S}, \mathbf{P}) = D_{JS}(Prob_S \parallel Prob_P)$, where D_{JS} is the JSD.

Another way to evaluate summaries is to find if they satisfy the user's information need. Let us define a basic information element as nugget. Each photo is a set of nuggets. We use the concept space (both marginal and joint) to model the nuggets. Given a set of nuggets $\mathbf{N}$ is a photo collection, we define a function $NuggetGain(\mathbf{S}, \mathbf{N})$ which measures the number of nuggets that summary $\mathbf{S}$ contains. A higher value of $NuggetGain$ signifies a better summary which can satisfy more information needs. However, we note that not all nuggets are equally important to the user. We define $Prob(\mathbf{N})$ as a probability distribution over nugget space representing the importance of the nuggets. This distribution can be estimated from the user's photo collection (local model) or a community photo collection like Flickr (global model). The refined metric is computed by $\langle NuggetGain(\mathbf{S}, \mathbf{N}), Prob(\mathbf{N}) \rangle$.

5. EXPERIMENTS AND RESULTS

We collected 40K personal photos from 16 different users by crawling Flickr, Picasa and other photo archives. For every user, the dataset contains photos shot over a time span ranging from a few months to a year. For every user, we generate 10 different summaries with sizes varying from 3% to 30% of the collection size. We compare the summaries generated by our framework with two baselines: random summary (random selection without replacement) and clustering (using K-Means on the entire feature space). Table 1 compares the performance of the summaries using all three evaluation metrics. In all the cases, we find that the summary generated using our algorithm performs much better than the baselines. Figure 1 shows the performance of the three summarization algorithms using the $NuggetGain$ local and global metrics. The clustering algorithm finds exemplars by using the entire heterogeneous feature space, without leveraging on the multimodal semantic concepts. Such an approach may not be useful for a summarization objective. Hence it performs little better than random selection. In the results presented, we have chosen equal weights for $Qual$, Div and Cov in equation 2 (thus, α= β = γ = 1). However, users can generate different representative summaries by using their personal preferences to bias these parameters during the summarization process. Thus a choice of high $Qual$ and low Div may generate a summary with many attractive photos, but may have redundancies.

Table 1: Results of Evaluation

Eval Metric	Random	Clustering	Our Algorithm
$NuggetGain$ (Local)	0.49	0.53	**0.88**
$NuggetGain$ (Global)	4.41	4.5	**6.27**
JS-Divergence	0.39	0.37	**0.14**

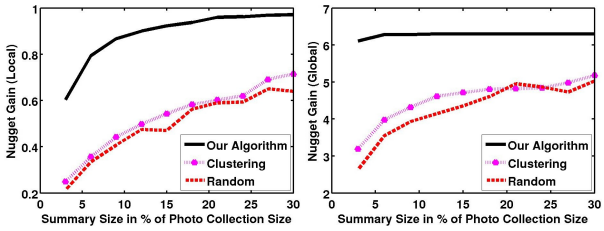

Figure 1: Evaluation using Nugget Gain (Local and Global)

6. CONCLUSIONS

In this paper we introduce a framework for summarization of archived and socially shared personal photos. We evaluate the models using 40K personal photos collected from 16 different individuals. Our results for photo archive summarization show that summaries generated using our models outperforms than baselines considerably. The performance of our algorithm monotonically increases with summary size. Future work includes investigation of incremental summarization models for dynamic photo collections (which increase in size) and a evaluation methodology which uses human feedback.

7. REFERENCES

[1] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries for large collections of geo-referenced photographs. In *Proc of World Wide Web*, 2006.

[2] D. Kirkpatrick. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. Simon & Schuster.

[3] I. Simon, N. Snavely, and S. Seitz. Scene summarization for online image collections. In *Proc. ICCV*, 2007.