Summarization of Archived and Shared Personal Photo Collections

Pinaki Sinha
Dept of Computer Science
University of California, Irvine
psinha@ics.uci.edu
SUPERVISED BY: Ramesh Jain (jain@ics.uci.edu)

ABSTRACT

The volume of personal photos hosted on photo archives and social sharing platforms has been increasing exponentially. It is difficult to get an overview of a large collection of personal photos without browsing though the entire database manually. In this research, we propose a framework to generate representative subset summaries from photo collections hosted on web archives or social networks. We define salient properties of an effective photo summary and model summarization as an optimization of these properties, given the size constraints. We also introduce metrics for evaluating photo summaries based on their information content and the ability to satisfy user's information needs. Our experiments show that our summarization framework performs better than baseline algorithms.

Categories and Subject Descriptors

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

General Terms

Algorithms, Design, Experimentation

Keywords

summarization, personal photos, social networks, 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 [7]. These photos are a rich source of information about the events taking place in a subject's life. The subject may use these photos for viewing (to evoke memories or to entertain oneself) and also for sharing with friends. With the present phenomenal popularity of smart-phones, it is expected that, in the near future the photo sharing platforms will experience an exponential increase in data traffic. Organization of such large collections of personal photos is hence an important and relevant problem.

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 or 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 important and 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.

We observe that the personal photo hosting platforms are used for two main applications: photo archival and social photo sharing. Users store the photos taken by them in web archives for accessing them at a later time. Each such photo archive typically consists of photos contributed by a single user and represents the user's life events. Table 1 shows a snapshot of 1200 photos archived by a user Joe. In addition to pixels (content), these photos have context data like time, location, event names etc. Users also share photos in their social network so that their friends are updated about the events happening in their lives. The important aspects of a photo sharing network are the users, the photos and the relationships between them. Figure 1 shows a social network of a user Joe. Each ring denotes a circular user and the boxes in it denotes the photos contributed by the user. The edges denote various relationships like user-user (friendship) and user-photo (user-tagged, user-commented). Summarization is important both for archived and socially shared personal photos. We discuss in Section 4 how the proposed summarization framework can be used in both these applications.

Personal photo summarization may be a very subjective process. Early experiments by Savakis et al. [12] show that selection of personal photos from a collection depends a lot on a users preferences and can vary across editors. Further, summary generated by the same editor at two different times may vary, reflecting their preference changes. We do not intend to create a unique summary subset from a photo corpus. Rather, our goal is to define a framework for automatic generation of a size constrained *informative overviews* of the collection. The users can select different parameters values of a summarization model to generate different overviews of the same corpus. In Section 3 we discuss these properties in detail.

We claim that an effective subset summary should satisfy some desirable properties. These properties are *Quality*, *Diversity* and *Coverage*. Quality determines the aggregate attractiveness of photos present in the summary. Diversity is a measure of non-redundancy. A good summary should not contain redundant or repeated information. Coverage ensures that the important *concepts* present in the photo collection are also represented in the size constrained summary. To clearly elucidate the effect of the proposed properties, let us consider the archived photos of user Joe (Table 1). If the goal was to generate a two element summary which ensures maximal location diversity, the summary would contain either photos from

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

Table 1: Snapshot of archived photo collection of a user Joe

ID	Location	Event Type	Time	# Photos
1	Irvine	Party	Jan 22, 09	200
2	Irvine	Office	Mar 15, 09	150
3	San Diego	Conference	Apr 16-19, 09	250
4	Beijing	Vacation	Sept 5-12, 09	600

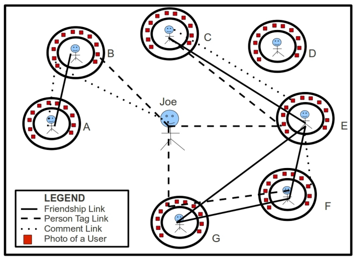

Figure 1: Snapshot of social photo collection of a user Joe

Irvine and Beijing or San Diego and Beijing (located in two different continents) and not from Irvine and San Diego (located in the same US state and hence, not maximally diverse with respect to location).

Designing an evaluation methodology for personal photo summaries is also a challenge. It is difficult to create ground truth summaries from photo collections because of the subjectivity, variance in user preferences and effort involved to skim through a large photo corpus. We take a different approach in this research. We claim that an effective photo summarization system should address two important objectives: information reuse and information discovery. The goal of *information reuse* is to find information already known to the user. The subject may want to get an overview of her own life events through a photo summary, thus evoking old memories. *Information discovery* addresses the objective of finding new information in a photo collection. This objective may be relevant for third party users like friends (in social networks) who may be interested in the subject's life events. We measure the usefulness of a summary based on its efficacy to address these twin objectives (Section 6).

In short, following are the contributions of our research:

- Proposing properties of an effective summary and defining models to compute them.
- Formulation of summarization as multiobjective optimization problem. We also propose algorithms to solve the problem.
- Modeling summarization for both archived and socially shared photo collection.
- Proposing an evaluation methodology of photo summarization without user generated ground truth.

2. RELATED WORK

The problem of summarizing web image collections has been investigated by several researchers. Simon et al. [13] address the problem of scene summarization, by selecting a set of canonical images or views web image collection of the scene. They define similarity between images based on the number of 3-D features they have in common. Jaffe et al. [4] use a hierarchical clustering method to generate summaries of geotagged photos at multiple resolutions. They also present an interface for visualizing salient photographs in a geographic region at varying zoom level. Unlike previous work, our interests lie in summarization of personal photos present in web archives or social networks. The characteristics of personal photo sets and multi-user web image collections differ significantly. Hence the modeling process in this research (including definition of summary properties), algorithms used and the evaluation methodology is different from previous work.

3. PROBLEM FORMULATION

Definition 1. 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 an effective and representative summary. In this section and the next, 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. We refer to it as $Interest$. Interest is an inherent property of a photo which determines its attractiveness to a subject. Thus for two photos, $\mathbf{p}_1$ and $\mathbf{p}_2$, if $Interest(\mathbf{p}_1) > Interest(\mathbf{p}_2)$, then the former will be preferred over the latter for inclusion in a summary. 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 the 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 $\mathbf{p}$. In this section, we define the properties of a summary based on these basic characteristics of photos. Detailed discussion and modeling of the basic characteristics is presented in Section 4.

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 follows:

$$Qual(\mathbf{S}) = \sum_{\mathbf{p} \in \mathbf{S}} Interest(\mathbf{p}) \tag{1}$$

A size constrained summary should avoid repetitions and should not contain redundant information. To achieve these goals, the photos in the summary should be diverse. Diversity of the summary can be modeled as an aggregation of the mutual distances of the photo pairs. We use minimum of the pairwise distances of the summary photos as the summary diversity (mean of pairwise distances can also be chosen as summary diversity).

$$Div(\mathbf{S}) = \operatorname*{Min}_{\mathbf{p_i}, \mathbf{p_j} \in \mathbf{S}, \mathrm{i} \neq \mathrm{j}} Dist(\mathbf{p_i}, \mathbf{p_j}) \tag{2}$$

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 individual photos in $\mathbf{S}$.

$$Cov(\mathbf{S}, \mathbf{P}) = \left| \bigcup_{\mathbf{p} \in \mathbf{S}} Represent(\mathbf{p}, \mathbf{P}) \right| \tag{3}$$

We model summarization as a multiobjective optimization function $\mathcal{F}$ which jointly maximizes these properties. A good summary $\mathbf{S}^*$, is a subset which maximizes $\mathcal{F}$ given the size constraint $|\mathbf{S}^*| = M$

$$\mathbf{S}^* = \operatorname*{arg\,max}_{\mathbf{S}^* \subseteq \mathbf{P}} \mathcal{F}(Qual(\mathbf{S}^*), Div(\mathbf{S}^*), Cov(\mathbf{S}^*, \mathbf{P})) \tag{4}$$

$\mathcal{F}$ combines the individual properties to generate a single effectiveness metric. Many such functions can be defined by combining the properties in different ways. We will discuss this in more detail when solving the optimization function in Section 5.

We should mention that similar concepts have been used in to research related to search result diversification and disambiguation [1] [3]. Agrawal et al [1] assume both text queries and retrieved documents belong to a taxonomy of information. Diversification is modeled as maximization of likelihood of finding a relevant document in the top-k position given this taxonomy. In [3], the authors propose properties which an ideal diversification function should satisfy. These properties are tested on semantic and product disambiguation applications. Kennedy et al. [6], propose a method to generate representative views of landmarks by diversifying image features and user tags. Van Leuken et al. [14] explore dynamically weighted clustering methods to generate diversified image search results using visual features.

4. SUMMARIZATION OF ARCHIVED AND SHARED COLLECTIONS

In this section, we define the data models for archived and socially shared personal photo collections. We then formulate the computation of $Represent$, $Dist$ and $Interest$ using the data models.

ARCHIVED PHOTO COLLECTIONS: Archived personal photos contain a host of contextual data in addition to the content (pixels). Some of them are captured by various sensors on the camera and some are user or community contributed. A photo $\mathbf{p}$ in an archive is represented by the tuple $(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ is a set of real valued quantitative attributes and $\mathbf{y}$ is a set of discrete categorical attributes or concepts. $\mathbf{x}$ is composed of pixel features, time and EXIF-based camera parameters (e.g., exposure time, focal length). The set $\mathbf{y}$ contains five concepts: location, event type, visual, temporal and face. The concepts can be generated from the community contributed textual data (e.g., tags, album names, descriptions etc), the image metadata (e.g., GPS induced geotags) or can be predicted using machine learning algorithms on the quantitative attributes. The *visual concepts* include four different scene types: outdoor day, outdoor night, indoor and sunset. A discrete set of *temporal concepts* is obtained by clustering the time stamps of the photos in a collection. Each temporal concept may signify a particular event that took place in a user's life. *Event types* denote a set of popular event categories that are present in consumer photo collections, e.g., birthday, trip, party etc. We leverage on the personal event ontology benchmark proposed by researchers at Kodak [8], to define these event categories. *Location concepts* are discrete city names denoting the geographical region where the photo was shot. We use a publicly available geo-database (`Geonames.org`) and the geotags present in photos to define the location concepts. *Face concepts* are set of unique faces present in a photo collection. We assume that faces are either manually tagged (e.g., Facebook's tagging feature) or are predicted by a face recognition system (e.g., Picasa or IPhoto). Given this data model, we define the basic characteristics of archived personal photos.

The $Interest$ value of a photo is a measure of its attractiveness which depends on image appeal and quality. Experiments by Savakis et al. [12] show that portraits, group photos and panoramas are some positive attributes that appeal to users of personal photo archives. Image quality depends on color distribution, hue, absence of blur and coarseness [11]. Building on these results, we formulate the $Interest$ value of a photo as: $Interest(\mathbf{p}) = \langle \mathbf{a}, \mathbf{w} \rangle$, where $\mathbf{a}$ is binary vector which denotes the presence or absence of these appealing and quality attributes in a photo. $\mathbf{w}$ are the weights associated with each of these attributes which represent their importance in the summarization process.

Distance between digital photos is formulated using both quantitative and categorical attributes: $Dist(\mathbf{p_i}, \mathbf{p_j}) = \lambda_q Dist_q(\mathbf{p_i}, \mathbf{p_j}) + \lambda_c Dist_c(\mathbf{p_i}, \mathbf{p_j})$ where, $Dist_q$ and $Dist_c$ denote the distances in quantitative and categorical feature spaces respectively. $Dist_q$ can be formulated as an euclidean distance. $Dist_c$ is dependent on the structure of the concept space. If a concept space is flat, each of its instances are disjoint and equidistant from each other. However, some concept spaces may have a hierarchical tree structure. E.g., *Geonames.org* provides a tree of location concepts and [8] provides an ontology of event types. To model the distance in a tree, we use the Jiang-Conrath Distance [5] (defined for concepts in semantic taxonomies).

The degree of representativeness of a photo $\mathbf{p}$ is denoted by the set $Represent(\mathbf{p}, \mathbf{P})$. We use the multimodal concept space to determines this set. Let $Con(\mathbf{p})$ denote the concepts present in a photo. Given a photo collection $\mathbf{P}$, we define a notion of coverage by a concept τ as a set of photo-concept pairs: $CovByCon(\tau, \mathbf{P}) = \{(\mathbf{p}_i, \tau) \mid \mathbf{p}_i \in \mathbf{P} \, and \, \tau \in Con(\mathbf{p}_i)\}$. Given this formulation, we define: $Represent(\mathbf{p}, \mathbf{P}) = \bigcup_{\tau \in \mathrm{Con}(\mathbf{p})} CovByCon(\tau, \mathbf{P})$.

SOCIALLY SHARED PHOTOS COLLECTIONS: The entities present in a social photo sharing platform are users and photos. The links or edges connecting these entities are of two types: user-user (e.g., friendship) and photo-user (e.g., if a photo belongs to a user). We skip the attributes of the user entity and discuss those of the photo entity. We represent a photo using the tuple $(\mathbf{x}, \mathbf{y})$ as before. However the individual attributes are a little different. The attribute $\mathbf{x}$, in addition to the previous values includes the features: number of *likes* and number of *comments*. The concept space $\mathbf{y}$, in addition to the previous values have the discrete features: owner, people-commented, people-liked. Using this extended data model, we redefine the characteristics of photo. $Interest$ of a photo now depends on *number of likes and comments*, and *number of friends tagged* in addition to the quality and appeal attributes defined before. $Dist$ includes a distance based on the ownership of a photo: two photos of the same owner or friends are less distance apart than two photos belonging to different people who are not related. We define a edge distance measure $Dist_e$ based on the friendship and ownership links discussed before. The definition of $Represent$ is same as before.

5. GENERATING SUMMARIES BY OPTIMIZATION

Before proposing models for generating optimized summaries, we make the following observations about properties of a summary (we skip the proof of these observations because of space):

- Maximizing Diversity (Div) is an NP-Hard Problem (since it can be mapped to the Max-Min Dispersion Problem [10]).
- Maximizing Coverage (Cov) is an NP-Hard Problem (since it can be mapped to the Maximum Set Cover Problem) [10].

Thus generating a summary with optimized Div or Cov will be computationally inefficient. However, greedy heuristics for these families of optimization problems are known to generate solutions which are a constant fraction of the optimal [10] [2]. Based on these observations, we propose our models for generating effective summaries in an efficient manner.

The summarization objective stated in Equation 4 is a classical multi-objective optimization problem. Multi-objective (MO) problems are traditionally solved by converting all objectives into a single objective (SO) function [9]. The SO problem can then be solved by traditional scalar valued optimization techniques. Below we discuss some of the classical methods and how they can be used in our application.

WEIGHTED AGGREGATION: Conversion of the MO function into an SO function can be carried out by aggregating all objectives in a single weighted function. We can formulate aggregation by assigning different weights to Quality, Diversity and Coverage objectives and combining them in a linear way. Thus the summarization objective (Eqn 4) can be reformulated as follows:

$$\mathbf{S}^* = \arg\max_{\mathbf{S}^* \subset \mathbf{P}} [\alpha Qual(\mathbf{S}^*) + \beta Div(\mathbf{S}^*) + \gamma Cov(\mathbf{S}^*, \mathbf{P})] \quad (5)$$

Every choice of the weights α, β and γ will generate a different summary which may show a different overview of the collection. Since, optimization of Div and Cov is NP-Hard, exact solution of equation 5 is inefficient. Instead, we adopt a greedy heuristic which finds the subset summary that produces the best aggregate $Qual$, Div and Cov at every summary size (Algorithm 1).

CONSTRAINT BASED APPROACH: An MO problem with n objectives can also be solved by transforming n-1 objectives into constraints and optimizing only one objective subject to the constraints. In our application, we can set a tolerance threshold on two of the objectives and try to optimize the third. The summarization is repeatedly done using different thresholds to generate the entire pareto optimal set [9]. Thus we get different summaries which lie on the pareto front. The users intending to get an overview of the photo collection may go through these small set of summaries instead of browsing through the entire corpus.

Algorithm 1 Greedy Algorithm for Summarization

```
Initialize the summary set S = ∅
Compute Qual(p) and Cov(p) ∀ p ∈ P
Find p* = argmax_{p∈P} [αQual(p) + γCov(p, P)]
S = S ⋃ p*
Recompute Cov based on concepts covered by p*.
while Length(S) < k do
    p* = argmax_{p∈P\S} [αQual(p) + βDiv(p ∪ S) + γCov(p, P)]
    S = S ⋃ p*
    Recompute Cov based on concepts covered by p*.
end while
```

6. EVALUATION METHODOLOGY

A straightforward way of evaluating automated photo summaries is to compare them with human generated ground truth. However, it is very expensive to generate ground truth summaries by asking editors to create a subset from thousands of personal photos. Also, summarization by selection of photos may be a very subjective task and is likely to vary across editors. Moreover, we are interested in evaluating the informativeness the summary overview rather than evaluating the photos present in them. In designing the evaluation process of the automated summaries, our goal was to answer the following questions:

- Is the summary representative of the larger corpus in information content?
- Does the summary address the objectives of information discovery and reuse?
- How does the information content of the summary change with its size?

To evaluate the representativeness of the summary, we compare the information content of the summary with that of the larger photo collection. Jensen Shannon Divergence (JSD) is a measure which compares the information conveyed by two probability distributions, P and Q. JSD is defined as: $D_{JS}(P \parallel Q) = \frac{1}{2}\{D_{KL}(P \parallel M) + D_{KL}(Q \parallel M)\}$ where, $M = \frac{1}{2}(P + Q)$ is the mean distribution and $D_{KL}(P \parallel Q) = \sum_i P(i) \log(\frac{P(i)}{Q(i)})$ is the KL-Divergence between P and Q.

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:

$$Inform(\mathbf{S}, \mathbf{P}) = D_{JS}(Prob_S \parallel Prob_P). \quad (6)$$

The information need of a user browsing through a photo collection can be broadly categorized into two types: information reuse and discovery. Information reuse relates to the fact of exploring information already known to the user. In our summarization application, a user digging through her own photo archive would want to evoke memories by viewing *interesting* information already known to her. An effective photo summary should satisfy this information need, thus saving the user from the tedious browsing process. Information discovery relates to the fact of finding new information in a corpus. Users exploring photos shared on a social photo sharing network would want to find *interesting* information which were previously unknown to them.

We now propose methods to define interesting information in a photo collection for both the above scenarios. First, we define a notion of information units or nuggets present in personal photos. In our data model, we define the nuggets using the multidimensional concept space. A nugget can be a marginal concept (e.g., locations like New York or events like party) or a joint concept like person-event-location (John at a party in New York), or person-visual concept-event-location (Jane in a group photo during conference trip to Italy). Each photo in a collection can be modeled as a set of nuggets composed of the marginal and joint concepts represented by it. Given a nugget n and a candidate summary $\mathbf{S}$, we define a binary function $NuggetGain(\mathbf{S}, n)$ which represents if the summary contains the nugget:

$$\begin{aligned} NuggetGain(\mathbf{S}, n) &= 1,\ if\ \exists\, \mathbf{p} \in \mathbf{S}\ s.t.\ n \in Con(p) \\ &= 0,\ otherwise \end{aligned}$$

We now propose two models for defining interestingness of information nuggets in personal photo collections. The **local model** assumes that the frequency of a nugget in a collection signifies its interestingness to a user. For instance, if the person-event-location nugget "Joe at wedding in Mumbai" is very frequent in Joe's photo collection, Joe will find the nugget interesting. A summary, should be ranked higher if it contains this nugget. The local model uses the frequency of nuggets to generate a probability distribution ($Prob_L$) over the nugget space ($\mathbf{N}$). We weight the $NuggetGain$ of a summary with $Prob_L$ to generate a measure which evaluates the information reuse objective:

$$NuggetGain_L(\mathbf{S}, \mathbf{N}) = \sum_{n_i \in \mathbf{N}} NuggetGain(\mathbf{S}, n_i) Prob_L(n_i)$$

We next propose a **global model** which assumes that there exists a "global knowledge" of relative importance of various information nuggets present in personal photos. This models reflects the preferences of an average user, who, without prior knowledge of a subject's life-events intends to get an overview through the photo

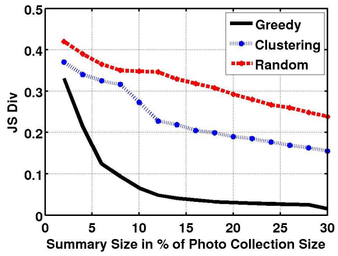

Figure 2: Evaluation using JS-Divergence

summary. We use the frequency of tags present on the community photo sharing site Flickr to generate a global probability distribution over the nugget space ($Prob_G(n)$). The assumption is, if Flickr has more photos with tags or nuggets "Trip in Paris" than "Wedding in Mumbai", an average user will be more interested to see the former than the latter. We weight $NuggetGain$ with $Prob_G(n)$ to define a measure which evaluates a summary for the information discovery objective.

$$NuggetGain_G(\mathbf{S}, \mathbf{N}) = \sum_{n_i \in \mathbf{N}} NuggetGain(\mathbf{S}, n_i) Prob_G(n_i)$$

7. EXPERIMENTS AND RESULTS

At the present state, our experiments show the results for the photo archive summarization problem. We leave the experiments on socially shared photo summarization as future work. We collected 40K personal photos from 16 different users by crawling Flickr, Picasa and other personal photo archives. For every user, the archive contains photos shot over a time span from a few months to a year. The quantitative attributes and concepts were extracted from the photos. We ran the summarization algorithms independently for every user. We generate ten different summaries of sizes varying from 3% to 30% of the original collection size.

In addition to summarization using our algorithm in equation 5, we also generate two baseline summaries using K-Means clustering and random selection (without replacement). Figure 2 shows the JS-Divergence between the summaries and the original collection in the People-Location-Event (PLE) joint concept space. The summary generated by our algorithm (solid line) has minimum divergence. Figure 3 shows the performance for the $NuggetGain_L$ and $NuggetGain_G$ models in the PLE concept space. In both the cases our summary outperforms the baselines. The monotonic performance curves of our summary (solid lines in the Figures 2, 3) prove that the information content increases with summary size. 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 5 (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.

8. CONCLUSIONS AND FUTURE WORK

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. The results show that summaries generated using our models outperforms than baselines considerably. Future directions include investigation of graph based models for summarization of photos shared on social networks and incremental summarization algorithms for dynamic photo collections (which increase in size). Developing an interactive summarization system whish uses human feedback to generate summaries may also be a future work. We may also need to develop a ground truth dataset to test photo summarization algorithms.

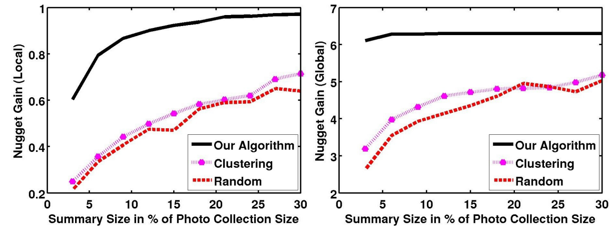

Figure 3: Evaluation using $NuggetGain_L$ and $NuggetGain_G$

9. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proc ACM WSDM*, 2009.

[2] R. Cohen and L. Katzir. The generalized maximum coverage problem. *Information Processing Letters*, 2008.

[3] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *Proc ACM WWW*, 2009.

[4] 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.

[5] J. Jiang and D. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc of Intl Conf on Research in Computational Linguistics*, 1997.

[6] L. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *Proc of World Wide Web*, pages 297–306. ACM, 2008.

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

[8] A. Loui, J. Luo, S. Chang, D. Ellis, W. Jiang, L. Kennedy, K. Lee, and A. Yanagawa. Kodak's consumer video benchmark data set: concept definition and annotation. In *Proc ACM Workshop on MIR*, 2007.

[9] P. Ngatchou, A. Zarei, and M. El-Sharkawi. Pareto multi objective optimization. In *Proc of Intl Conf on Intelligent Systems Application to Power Systems*, 2005.

[10] S. Ravi, D. Rosenkrantz, and G. Tayi. Heuristic and special case algorithms for dispersion problems. *Operations Research*, pages 299–310, 1994.

[11] J. San Pedro and S. Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *Proc of World wide web*. ACM, 2009.

[12] A. Savakis, S. Etz, A. Loui, et al. Evaluation of image appeal in consumer photography. In *Proc of SPIE*, 2000.

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

[14] R. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual diversification of image search results. In *Proc Intl Conf. on WWW*. ACM, 2009.