Selecting Representative Images from Larger Photo Sets

Author Information Withheld for Blind Review

ABSTRACT

Selecting a single photograph to represent a set of photographs is a useful approach when creating interfaces to large photo collections. Unfortunately, many different methods for selecting photographs are used in practice today, and none have been well studied or shown to be particularly effective. In this paper we look at several different common methods for selecting a single photograph in order to represent a larger collection. Our results show that the human selected images of most representative greatly outperforms the automated methods that are often employed. Further, the human selected images of least representative greatly under performs compared to the automated method. This implies that some images may not be appropriate for selection, and care should be given when selecting a representative image, as to not inadvertently select a “bad” image. Further study of this problem is needed to improve automatic methods in this domain.

INTRODUCTION

As digital photograph collections grow [3], new applications are constantly being developed to help users organize and navigate through their collections [5, 7, 8, 14]. When dealing with a large collection of images it can be difficult and confusing for the user to be presented with the full collection. Therefore, a large part of these applications is to select a single image, or small number of images, to represent parts of the entire collection (i.e. a single photo to represent a days worth of pictures, or the photographs in one album). However, there is very little research concerning the actual image selection methods, and very little is known which methods work the best. In this paper we investigate different methods of selecting a single representative photograph from a larger set. For the purposes of this paper, by representative image, we mean the image that gives the viewer the best idea of what is in the overall set. We do not mean “best” as in a personal favorite. Our findings show that a human selection for “most representative” image significantly out performs the other methods commonly in use today, indicating that there are better methods and a lot of room for improvement.

Researchers have shown that photographs tend to be taken in groups, which are naturally related [1, 4, 6]. This relationship implies that most photo collections have a common internal theme, making it possible to select a single image which illustrates that theme for a viewer. If an appropriate image can be selected then applications can make use of this by giving the viewer a lot of information in a small amount of space. However, if the selection is done poorly, then the image will not convey the theme of the set and may confuse the viewer rather than be useful. In other words, a good selection can be very beneficial to photo applications, while a bad selection will hurt the results.

The problem of image selection can be formally stated as follows. Given a set of photographs $P = (p_1, p_2, ..., p_n)$, and a test for how well a picture represents the set, $R(p_i)$, the goal is to select the p_i for which R returns the highest value, as shown in Equation 1. The problem is that R is completely unknown. The methods that we test in this paper are simple approximations of R.

$$max(R(p_i), R(p_2), ...R(p_n)) \tag{1}$$

We studied several common methods for selecting an image to represent a larger set. Those methods are: the first image in the set, the middle image in the set, the image with the largest number of distinguishable faces, the image whose histogram is closest the overall average histogram of all images, a randomly selected image, a human selected “most representative” image for the set, and a human selected “least representative” image for the set. Each of these methods (with the exception of the last two) were chosen because they are simple heuristics that have either already been employed in previous research or could easily be implemented in such an endeavor. We used a human selected best image as a control to see how the other methods fair in our tests, and a worst image as a baseline to see if some images clearly are inappropriate.

The results of our study shows that the human selection of most representative outperforms the other methods. This ‘implies that current methods can be improved. A popular comment from our study participants was that given the choice they would have chosen a different photograph. This implies that there is no single answer for the problem, but rather several images may be appropriate as representative images. In other words, a good image selection method needs only to select one of the representative images. Likewise, we found that the human selection of worst image had significantly lower performance than any other method. This implies that as there is a set of appropriate representative images, there is also a set of inappropriate representative images. At the very least, a selection method should avoid picking from this latter set, as it will produce undesirable and confusing results.

RELATED WORK

Most commercial application approach the image selection problem by choosing the first image in the set. Perhaps most noticeable is the Windows operating system. When viewing folders in “thumbnail mode,” the first four images in a folder are overlaid on top of the folder icon. Figure 1 shows an example of a folder icon in this view. The photo sharing web site Flickr also selects the first image in an album as the default representative image for that set of images. Also, several digital cameras will display a calendar view, where a thumbnail of the first image of the day is shown in each calendar box.

AutoAlbum [10] is a system that is designed to automatically create photo albums from a collection of photographs. In the original design of the system, the middle image of a set was taken to be the representative image. This method was changed during testing when a photograph of a nondescript ceiling was selected to represent a set of images of a room full of people.

A common approach to image selection is to use some type of color analysis of the images [12, 10]. In such methods the color properties can be used to determine if the picture is appropriate for selection or if the color scheme is close to the average of the other images in the set. In our study we investigate the effectiveness of color histogram analysis as a means of selecting a photograph.

Recently researchers have begun using collages as a means of summarizing large collections of photographs [2, 13]. In these approaches, a composite of many or all of the photos in the set is made into one image. Alternatively, AutoCollage [11] is a system which takes pieces of images to create a new unique summary image of the set. We do not investigate these methods, as we are interested in single image selection rather than creating new images.

SELECTION METHODS

We investigated several different methods for selecting a single image from a subset of larger images. In this section we briefly describe each of the methods and reasons for testing it. We describe our implementation of each method in the Experiment Section of the paper.

First Image In Set: It is common to select the first image in a set as the representative image. This is done in the Windows operating system, when the user is in thumbnail view, a folder that contains images will be shown with thumbnails of the first four images in the folder overlaid on top of the folder icon. Figure 1 This is also the method used by the photo sharing web site Flickr when displaying a thumbnail for a photo album.

Middle Image In Set: This is similar to first image in the set, and researchers have previous used the image in the middle of the set as a representative image.

Average Histogram: The image in the set whose histogram is closest to the overall average has also been used as the most representative image. The color distribution of the selection image should be similar to most of the images in the set.

Figure 1: Example of a folder containing images when in Thumbnail view in the Windows Operating System.

Internal Contrast: The human eye is often drawn to sharp changes, or contrast in the scene. Such contrast may be indicative of scene context and other information.

Faces: The appearance of faces tells of the people who were at the event being photographed. Usually a photograph of people will also be framed in such a way that it provides additional context about what is being photographed.

Human Selection of Best: It is our expectation that a human can do better at selecting the most representative image with higher accuracy than any of these methods alone.

Human Selection of Worst: It is our expectation that a human can select an image that is not representative.

Random Selection: Randomly selected images should have approximately average results.

EXPERIMENT

For our experiment we used twenty-one sets of twenty images each. Six of the 21 sets were donated explicitly for use in this research project. No one person donated more than two image sets, so if a donor participated in the study, his or her familiarity with the photographs should have minimal impact on the final results. The remaining fifteen sets were albums acquired from the Flickr web site, and are under a Creative Commons license, allowing for redistribution and modification of the original images. Only the first twenty images in each selected album was used in the experiment. For each set, six of the 20 images were selected as being potentially the most representative image in the set, using the first six methods described in the previous section. If less than six unique images were selected, either because of a lack of faces, or a single image qualified under two methods, then a random and/or worst image was also selected. In all 17 of the sets had a faces image, 11 had a worst image, and 9 had a random image. Every set was represented by one image selected by each of the other methods. We describe how we implemented each method later in this section.

We invited participants to take part in the study over the World Wide Web. Initial invitations were sent to mailing lists for computer science and education graduate students.

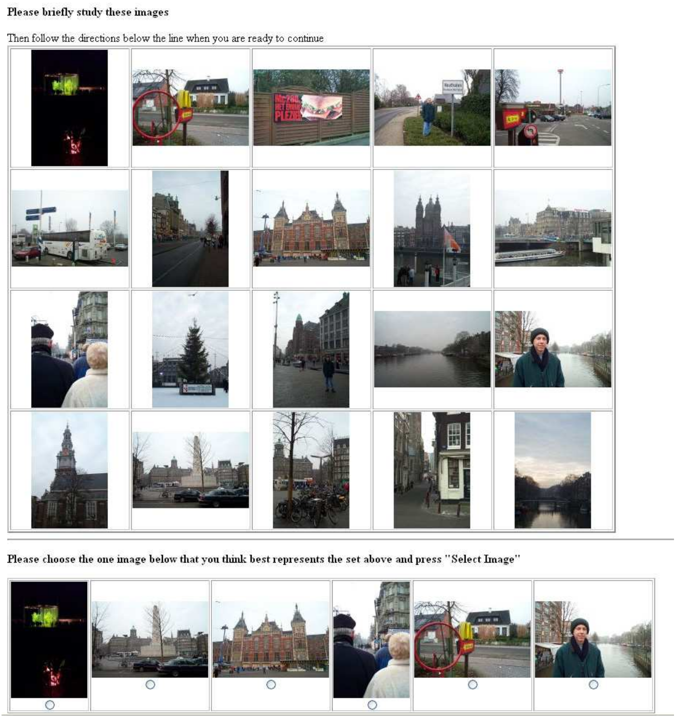

Figure 2: Screen shot of our user study.

The invitation encouraged participants to forward the invitation to friends and family who they thought may be interested in participating. Our human subjects approval prohibited us from collecting any demographic or geographic information about the participants. After agreeing to participate in the study, each user was shown a set of 20 images. Below that they were shown the 6 candidate images to choose from and asked to select the one image that they felt was most representative. This was repeated a total of 21 times. The order of the sets and order of the candidate images images were independently random for each volunteer. Incomplete surveys were not recorded. Volunteers were also given the opportunity to leave comments about their experience at the conclusion of the survey, however this information was separated from individual answers. In total we received 63 completed surveys. Figure 2 shows a screen shot of a single image trial from our user study.

For each set, the selection method was conducted in a consistent manner. All of the photographs contained EXIF metadata and were examined at the same resolution. All of the photographs were either 4×3 or 3×4 aspect ratio. The first image in the set was determined by the time stamp recorded in the photograph EXIF data by the camera. Since each photo set contained exactly 20 images, the 10^{th} image in the set was used as the middle image, again ordered by the time stamp. We used the Python Image Library (PIL) to compute the RGB histogram for each image, and the average histogram. We took the image with the smallest difference between its own and the average histogram. For internal contrast we used the algorithm described by [9] and chose the image whose sum of pixel salience was the highest. To determine the photograph with the most faces, we simply

Selection Method	Total Votes	Expected Vote
First Image	131	225.792
Middle Image	170	225.792
Average Histogram	218	225.792
Faces	194	184.32
Contrast	154	225.792
Least Representative	20	118.272
Random	107	96.768
Most Representative	542	225.792

Table 1: The total number of "votes" for each selection method and the expected number of votes.

counted them by hand. A random number generator set to return values between 0 and 19 (inclusive) was used whenever we needed to randomly select an image.

RESULTS AND DISCUSSION

Our hypothesis is that human selection can do a better job than the automatic methods that we tested for picking a most representative image, and further that human selection can be used to select a least representative image, under performing by comparison to the automatic methods. We perform a χ^2 test with a null hypothesis that each method should perform with the same results to each other. Table 1 shows number of times an image of each method was selected and the expected selections, assuming that each method should perform equally. Faces, least representative, and random selection have a lower expectation than the other methods since they were not used in all 21 sets.

For our results, $\chi^2 = 602.752$ with 7 degrees of freedom. The P value is less than 0.0001. With extreme confidence we may reject the null hypothesis that all methods perform equally, and infer that the human selection for most representative clearly out performs the other methods. This means that a human can approximate R, the test for how well an image represents the set, with higher accuracy than than the other methods. While the results have likely experienced a masking effect, we believe that the extremely low performance of the human selection of least representative indicates that there are poor choices, and again a human can also select such an image.

A common comment among participants in our study was that for some sets, they would have chosen a different image that was not one of the six choices. This comment combined with the fact that the human selection method out performs the other methods implies that there is some amount of malleability in the selection process, i.e. while the most representative image to the volunteer was not an option, some selected image was good enough. In other words, there may be more than one image that is acceptable as most representative. Given a set of photographs P, there is another set $M \subset P$ where any $p_i \in M$ is an acceptable choice for representative image.

While there may be several correct answers, the fact that the human selection for worst performs so poorly brings out an important point: there are clearly images that do not represent the set properly and should be avoided. We may say that there is a set $L \subset P$ where $p_i \in L$ is a poor choice for representing the set. Masking effects in our experiment prevents us from drawing additional conclusions about the performance of the other methods, and should be addressed in future studies.

From these above conclusions we may relax the formulation of the selection problem given in Equation 1. Rather than trying to select the single image that is the most representative image, we can say that an image selection method is successful if it selects $p_i \ni p_i \in M$. At the very least a selection method should not select an image from L.

We may extend our conclusions to say that some image processing should be applied in the selection process to make an informed select, and that random and set positional methods should be avoided as they may inadvertently make a poor choice. This was observed in [10] where they were originally using the middle image for selection until it yielded a nondescript picture of a ceiling. It is our belief that the face detection, histogram analysis, and contrast detection methods would yield better results than the random and set positional methods. However, limitations in our experimental design preclude us from being able to draw such conclusions.

We are excited by the initial results of this work. We have shown that human selection can outperform many of the common methods that are used today for selecting a representative image, meaning that better results are possible, however more study in this area is necessary. We have also been able to give a case for relaxing the formulation of the selection problem. Unfortunately limitations in our design experiment have prevented us from being able to make strong comparisons between the different automatic image selection techniques.

CONCLUSION

In this paper we investigate the problem of choosing a single image to represent a larger collection of related photographs. In total we test eight different selection methods based on: set position (first and middle images in the set), histogram analysis, face detection, contrast detection, random selection, human selection for best, and human selection for worst.

From our results and volunteer comments we can conclude that while human selection for best image out performs the other methods, often several images may serve as a representative image. A good image selection method should be able to pick at least one image from the set of representative images. Likewise, not all images serve as good representative images, and at the very least a selection method should avoid picking from the set of poor images. This means that set position and random selection methods should not be used as they may unintentionally yield bad results. Limitations in our experimental design prevented us from being able to make nay strong conclusions about the other selection methods. This challenging problem will continue to become more important as personal digital photo collections continue to grow, and further study in this area is necessary.

Acknowledgements Withheld for blind review

References

[1] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. *ACM Transactions on Multimedia Computing, Communication, and Applications*, 1(3):269–288, 2005.

[2] Nicholas Diakopoulos and Irfan Essa. Mediating photo collage authoring. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 183–186. ACM Press, 2005.

[3] Steven M. Drucker, Curtis Wong, Asta Roseway, Steven Glenner, and Steven De Mar. Mediabrowser: reclaiming the shoebox. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 433–436. ACM Press, 2004.

[4] Ullas Gargi. Modeling and clustering of photo capture streams. In *MIR '03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 47–54. ACM Press, 2003.

[5] Andreas Girgensohn, John Adcock, Matthew D. Cooper, Jonathan Foote, and Lynn Wilcox. Simplifying the management of large photo collections. In *INTERACT*, 2003.

[6] Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 326–335. ACM Press, 2002.

[7] David F. Huynh, Steven M. Drucker, Patrick Baudisch, and Curtis Wong. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1937–1940. ACM Press, 2005.

[8] Allan Kuchinsky, Celine Pering, Michael L. Creech, Dennis Freeze, Bill Serra, and Jacek Gwizdka. Fotofile: a consumer multimedia organization and retrieval system. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 496–503. ACM Press, 1999.

[9] Y.F. Ma and H.J. Zhang. Contrast-based image attention analysis by using fuzzy growing. *Proceedings of the eleventh ACM international conference on Multimedia*, pages 374–381, 2003.

[10] J Platt. "autoalbum: Clustering digital photographs using probabalistic model merging", proceedings of ieee workshop on content-based access of image and video libraries (cbaivl-2000) ieee press,". In *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL-2000)*. IEEE Press, 2000.

[11] Carsten Rother, Lucas Bordeaux, Youssef Hamadi, and Andrew Blake. Autocollage. *ACM Trans. Graph.*, 25(3):847–852, 2006.

[12] Alexander C. Loui. Andreas Savakis. Automated event clustering and quality screening of consumer pictures for digital albuming. *IEEE Transactions on Multimedia Media*, 5(3):390–402, 2003.

[13] Jingdong Wang, Jian Sun, Long Quan, Xiaoou Tang, and Heung-Yeung Shum. Picture collage. In *Computer Vision and Pattern Recognition*, 2006.

[14] Liu Wenyin, Yanfeng Sun, and Hongjiang Zhang. Mialbum - a system for home photo managemet using the semi-automatic image annotation approach. In *MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia*, pages 479–480. ACM Press, 2000. A photo organization tool.