Efficient Mobile Imaging Using Emphasis Image Selection

Jiebo Luo and Amit Singhal
Eastman Kodak Company
Rochester, New York

Andreas Savakis
Rochester Institute of Technology
Rochester, New York

Abstract

Mobile imaging helps realize "any time, anywhere" visual communication by allowing consumers to capture, review, share, and print pictures via mobile devices while "on the go". However, major technical challenges exist in bandwidth, power consumption, display and other areas. As a result, the limited resources available in mobile imaging need to be utilized in an intelligent and effective fashion. To this end, we propose to use emphasis image selection (EIS), which automatically selects the most important image, i.e., the one that should receive the most attention or special treatment, given a set of photographic images that typically belong to the same event.

Images from a film photofinishing order or a digital camera memory card are first clustered into groups related to different events based on metadata and image content. Emphasis images are selected to provide a summary of the group content and are sent to mobile devices for preview, sharing, browsing, and printing. Emphasis images also receive favorable treatment in image rendering, compression and transmission. Experimental results using consumer pictures are presented to demonstrate the efficacy of the proposed system.

Introduction

Digital imaging, wireless, and broadband communications are evolving technologies that enable new and richer communication methods. Pictures and videos, in particular, play a major role in how people share, express, and remember events in their lives. For example, friends and family members can share pictures of their vacations, children, graduation ceremonies, and wedding moments– all on the screen of a mobile device ("share life on the go"). Such scenarios are illustrated in Fig. 1.

It is well known that images and videos are effective ways of communicating commercial information. Wireless imaging is expected to capture a big share of the market on business applications. Examples include news agency photographers sending pictures from the world's hot spots out to hundreds of US newspapers, people reading news footage dramatically enhanced by the insertion of images and videos related to the events, insurance adjusters filing images of a burned house or a damaged automobile from the field, construction-company engineers sending pictures of a project back to the home office, etc.

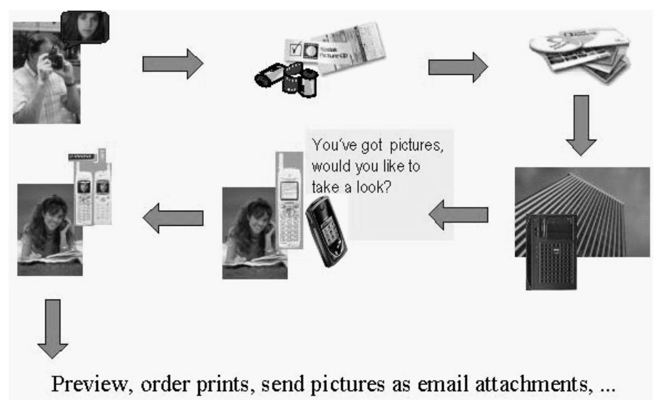


Figure 1. Mobile imaging "any time anywhere".

For all of the above to become reality, many obstacles need to be overcome in wireless imaging technology. Major technical challenges include: connection bandwidth, coverage area, error correction, power requirements, consolidation of multiple standards, image rendering for display or print, image management, and last but not least, ease-of-use.

Some of the above requirements in wireless imaging are conflicting and impose additional technological challenges, e.g. in terms of display image quality.[1] Most mobile devices with image display capabilities have small-sized display screens, a result of form-factor limitations associated with the physical devices. The smaller screen size makes it imperative that the displays have the highest color bit depth possible to achieve good image rendering results. However, the computing technology behind most of the currently popular cellular phones and PDA's does not permit color bit depth higher than 8 bits in total (instead of 8 per color channel), without significant increase in the bulk and price of the devices. A second conflict arises because of the limited bandwidth associated with the wireless transmission of data. Bandwidth constraints necessitate small image file sizes, which are difficult to achieve without imposing high levels of lossy compression, especially at higher image resolutions and color bit depth. All of these barriers affect image quality and severely hamper ease-of-use when it involves image-related applications on mobile devices.

In this paper, we focus on innovative ways of making mobile digital imaging easier and more effective. Given the limitations in bandwidth and power availability on mobile devices, it is desirable to selectively transmit images over mobile channels. Increasing the compression ratio, or transmitting only the images that are considered appealing or important, can accomplish the goal of selective transmission. The JPEG2000 standard[2] is a suitable vehicle for varying bit rates associated with particular images, as it allows progressively increasing lossy compression as needed. What is additionally needed is a way of assigning relative importance among images. In this paper, we propose to employ emphasis image selection methods for that purpose.

Emphasis Image Selection (EIS) deals with the assignment of relative emphasis or importance when processing images in a group, and has been previously used for other applications, such as automatic albuming.[3] Based on a psychophysical experiment,[4] a number of features were found to be influential in the selection of emphasis images by human observers. The following emphasis-indicating features were selected based on a compromise between importance and implementation cost: (a) features based on objective measures: sharpness, colorfulness, picture format; (b) features related to people: skin area, people present, close-up images; (c) features related to composition: main subject size, centrality, and compactness.

Within the realm of mobile imaging, images either from a film photo-finishing order or a digital camera memory card are first parsed into groups related to different events based on metadata and image content.[3,5] Emphasis images receive favorable treatment in image rendering, compression and transmission. Emphasis images may also be selected to provide a summary of the pictures and sent to a mobile device for reviewing, sharing, browsing, and printing.

Emphasis Image Selection

In a variety of applications that deal with a group of pictures, it is important to rank the images in terms of their relative appeal (emphasis), so that they can be processed or treated according to their emphasis value. Thus, emphasis image selection (EIS) may be defined as follows: given a set of photographic images that typically belong to the same event, automatically select the most important image, i.e., the one that should receive the most attention or special treatment.

Potential EIS applications include automatic albuming, image compression, image retrieval from databases, and mobile imaging. In automatic albuming, assigning emphasis values to images in the album can help distribute them among album pages or arrange them in a single page. When compressing images in a fixed memory block, the compression ratio can be adjusted according to image emphasis. In database retrieval, the order in which the images are retrieved may be adjusted, depending on image emphasis. In mobile imaging, emphasis images can be selected to provide a summary of the pictures and be sent to a mobile device to facilitate easy and efficient reviewing, sharing, browsing, and printing.

The design of an automated EIS system is a challenging task, because the estimation of image appeal is based on high-level cognitive human processes and subjective judgment. The challenges involve selection of proper features and knowledge integration through a classifier. In this work, feature selection is motivated by the results of a ground-truth study using human observers.[4] The observers' comments were compiled in terms of attributes that may contribute in a positive or negative manner toward the emphasis image score. A fundamental characteristic of the EIS feature extraction stage is that it includes both low-level and intermediate-level features. The classification stage is based on a Bayesian network classifier, which provides a flexible and easily trainable architecture for knowledge integration.

The ground-truth results showed that important features are not strongly related to objective metrics that are traditionally used for image quality evaluation.[6] In Table 1, the majority of strong positive attributes belongs to the people and composition/subject categories, and their estimation is a difficult task that, in some cases, is a subject of current research.[7] Given the current state-of-the-art of image understanding technology, the following features were selected based on a compromise between importance and implementation cost: (a) features based on objective measures: sharpness, colorfulness, format uniqueness; (b) features related to people: skin area, people present, close-up images; (c) features related to composition: saliency map size, centrality, compactness, and variation. A brief description of the features is given below.

It is possible to compute a large number of objective features; however, *colorfulness* was found to be the most important, and *sharpness* was only marginally important.

Colorfulness: A measure of colorfulness was obtained by examining for the presence of high-saturation colors along various hues.

Sharpness: The sharpness is estimated using the edge profile of the (luminance) image. The edge histogram is formed and the regions that contain the strongest edges are identified as those that are above the 90^{th} percentile of the edge histogram. Strong-edge regions are refined via median filtering, and the average of the strongest edges provides an estimate of sharpness.

Format Uniqueness: The ground-truth experiment indicated that if a picture is the only "panoramic" picture in a group, it is more likely to be selected as the emphasis image. The relative feature "format uniqueness" represents this aspect property. For each image i in the group, we define the format, based on the long and short pixel dimensions, l_i, s_i, of the image. Format uniqueness is useful with Advanced Photo System (APS) pictures that have various aspect dimensions.

People: People-related features are extremely important in determining image emphasis, but many of the positive attributes that are related to people are difficult to compute, e.g., people smiling, people facing the camera, etc. Skin detection methods[8] allow the computation of some people-related features, such as: people present, skin area, and close-up images. Some of the simpler methods involve thresholding in color space, while more complex methods include face detection.[9]

Skin Area: The percentage of skin/face area in a picture is computed as a preliminary step for determining the presence of people. Skin area is a continuous variable between 0 and 1 and correlates to a number of features related to people. For example, for pictures taken from the same distance, increasing skin area indicates that there are more people in the picture and correlates with the positive indicator of "whole group in photo." Alternatively, if two pictures contain the same number of people, larger skin area may indicate larger magnification, which correlates with the positive attribute of a "close-up" image.

People-Present: The presence of people is detected when a significant amount of skin area is present in the image. The number of skin pixels in the image is counted and people are assumed present when the skin area is above a threshold. *People-present* is a binary feature indicating the presence or absence of people.

Close-up: A simple measure of determining a close-up image is to perform skin detection and determine the percentage of skin area in the central portion of the image.

Figure 2. Saliency map example

Composition: Good composition is the most important positive attribute of picture emphasis, and bad composition is the most important negative attribute. Good composition may follow a number of general rules,[10] but these rules are often violated in order to express the photographer's perspective. The composition measures considered here are with respect to the image saliency map.[11] The features of "*borderness*" and "*centrality*" are considered in judging the saliency of image regions to construct a saliency map (see Fig. 2). The saliency map is quantized to three levels corresponding to most-salient regions (white), least-salient regions (dark), and intermediate regions (gray). The high-saliency regions in the saliency map are likely to correspond to the main subject in the picture; thus, they provide cues to how well the picture is composed. Using the saliency map, the features of main subject size, centrality, compactness, and variation are computed.

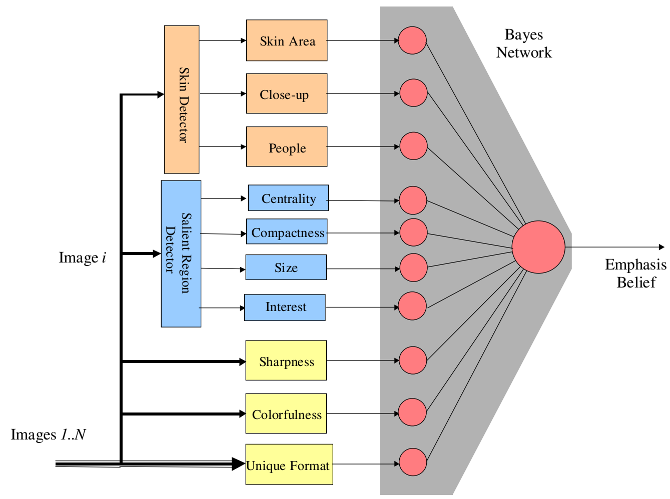


Figure 3. Emphasis value computation.

The outputs of the feature extraction stage represent statistical evidences that are integrated by a Bayesian network to compute the belief that the processed image is the emphasis image, as shown in Fig. 3. After the emphasis values have been computed for all images, the one with the highest emphasis is chosen as the emphasis image. Different evidences may compete or reinforce each other, according to knowledge derived from the results of the ground truth study. Competition and reinforcement are resolved by the inference engine of the Bayesian network.

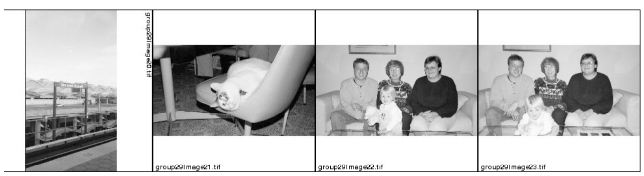
A Bayes net is a directed acyclic graph that represents causality relationships between various entities in the graph, where the direction of links represents causality. Evaluation is based on knowledge of the joint probability distribution function among various entities. The Bayes net advantages include explicit uncertainty characterization, efficient computation, easy construction and maintenance, quick training, and fast adaptation to changes in the network structure and its parameters.

A Bayes net consists of four components: (i) Priors: The initial beliefs about various nodes in the Bayes net; (ii) Conditional probability matrices (CPMs): Knowledge about the relationship between two connected nodes in the Bayes net; (iii) Evidences: Observations from feature detectors that are input to the Bayes net; (iv) Posteriors:

The final computed beliefs after the evidences have been propagated through the Bayes net. These components are discussed in more detail in.[12] The most important component for training is the set of CPMs that represent domain knowledge for the particular application at hand.

A single-level Bayes net is used in the current system, where the emphasis score is determined at the root node and all the feature detectors are at the leaf nodes. It should be noted that each link is assumed to be conditionally independent of other links at the same level, which results in convenient training of the entire net by training each link separately, i.e., deriving the CPM for a given link, independent of others. In practice, this assumption is sometimes violated; however, the independence simplification makes implementation feasible and produces reasonable results.

GT = (a) 58 (b) 70 (c) 93 (d) 94



GT = (e) 54 (f) 65 (g) 93 (h) 95

Figure 4. Example of EIS results for two groups of images with ground-truth values; image (d) was selected as the emphasis image in the group (a)-(d); image (g) was selected as the emphasis image for the group (e)-(h).

The EIS algorithm was trained using 14 image groups for which one or two statistically separable emphasis images could be identified. We have found that 77% of the time the EIS algorithm would select an image in the top-three selected by human observers (collectively).[13] Image examples are shown in Fig. 4.

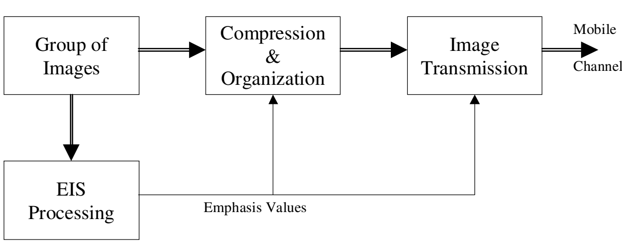


Figure 5. EIS-based mobile imaging environment.

Mobile Imaging Using Emphasis Images

Advances in digital imaging have made it possible to sort digital photographs into events using time stamps routinely embedded in digital camera or APS pictures,[3] by the scene content, or a combination of both methods.[5] Once pictures are grouped by event, an emphasis rating is assigned for each picture within a group. The highest rated images are selected as the emphasis image to facilitate image organization, compression, and transmission. Fig. 5 shows such an EIS-based mobile imaging environment. The collection of emphasis images provides a compact summary of a large number of pictures. A user can preview the photo album by browsing the emphasis images; request pictures related to a particular event; share the pictures of a particular event with someone even without previewing all related pictures; order prints of the emphasis images or related pictures; zoom in the emphasis images to check details; order enlargement of an emphasis image; and so on. All of the above tasks aim at ease-of-use and are accomplished in a cost- and time-efficient manner suitable for mobile device users.

This is applicable in the context of content-based indexing and retrieval (CBIR) of images on a mobile device as well. Unlike traditional CBIR applications, mobile devices have the additional constraint of limited bandwidth and associated expenses. When browsing through a large database of images, the information communicated between the mobile device and the server needs to be minimized. Most existing methods for hierarchical classification of images in a database rely on the use of an image-similarity metric to partition the images into sets.[14] These sets can be further clustered using a broader similarity metric to create a hierarchical scheme for image retrieval and browsing. The main drawback of this scheme is that it emphasizes low-level content over semantic content. As an example, images from a birthday party may be distributed into very different clusters, depending on whether the image is of the birthday cake or the guests. Techniques associated with EIS can be used to alleviate this problem.

Instead of using low-level features for image (and image set) clustering, we first use the event-based classification scheme to group the images in a database in terms of events. These events can be broad or narrow and can represent a specific occasion (e.g., Christmas 2000), an entire year (e.g., images captured in 2001), or an entity (e.g., images of Santa Claus) etc. Broader events can be further partitioned into narrower categories. User interaction in the form of inexpensive text messages can be used to determine the appropriate categories for classification. Depending on the application, a number of different classification hierarchies may be used to represent the images in a database. These classification hierarchies can be cross referenced to allow for more efficient retrieval. For example, a user may request a set of emphasis images related to baseball events using one hierarchy, select one of the images, and switch to a different hierarchy to request all images taken on that day, select another image taken at a different location, and, again, switch the hierarchy to get images containing one of the people in that image.

Use of emphasis images can reduce the number and size of images transmitted to the mobile device during this retrieval task. Instead of showing all of the images at device resolution, the system can transmit the emphasis image at high resolution and other related images at thumbnail resolution. The user has the option of requesting full resolution images for any of the thumbnails. Use of text in association with thumbnail resolution images can efficiently reduce the communication cost associated with mobile imaging.

Conclusions

In this paper, we describe a system for efficient and easy mobile imaging, enabled by automatic emphasis image selection. Emphasis pictures representative of events are selected to provide a photo summary that is sent to a mobile device for reviewing, browsing, sharing, and printing. We are in the process of implementing it on mobile imaging devices including phones and PDAs with (multimedia messaging service) MMS capabilities.

References

1. J. Luo, A. Singhal, G. Braun, R. T. Gray, N. Touchard, and O. Seignol, "Displaying images on mobile devices: capabilities, issues, and solutions", *Proc. ICIP* (2002).
2. David S. Taubman and Michael W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer International Series in Engineering and Computer Science, (2002).
3. A. Loui and A. Savakis, "Automatic image event segmentation and quality screening for albuming application", *Proc. ICME* (2000).
4. Andreas E. Savakis, Stephen P. Etz, and Alexander C. Loui, "Evaluation of image appeal in consumer photography", *Proc. SPIE Human Vision and Electronic Imaging* (2000).
5. J. Platt, "AutoAlbum: Clustering Digital Photographs Using Probabilistic Model Merging", *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries 2000*, pp. 96-100 (2000).
6. J. Katajamaki and H. Saarelma, "Objective Quality Potential Measures of Natural Color Images", *J. Imaging Sci. Technol*., Vol. 42, pp. 250-263 (1998).
7. A. Rozenfeld, "Image Analysis and Computer Vision: 1998", *Computer Vision and Image Understanding*, pp. 36-95, (1999).
8. M. J. Jones and J. M. Rehg, "Statistical Color Models with Applications to Skin Detection," *Proc. CVPR 99*, (1998).
9. H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network Based Face Detection", *IEEE Trans. PAMI*, (1998).
10. A. Feininger, *Principles of Composition in Photography*, American Photographic, Garden City, NY, (1973).
11. J. Luo, S. Etz, A. Singhal, and R. T. Gray, "Performance Scalable Computational Approach to Main Subject Detection in Photographs," *Proc. SPIE Human Vision and Electronic Imaging* (2001).
12. J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, 1988.
13. A. Savakis *et al*., "Emphasis image selection: Choosing the most appealing image," *IEEE Trans. Multimedia*, in preparation (2003).
14. S. Krishnamachari, M. Abdel-Mottaleb, "Image browsing using hierarchical clustering," *Proc. IEEE Int. Sym. on Comp. and Comm.*, p. 301-307, (1999).