Kioku:
A Semantic Indexing and Exploration Interface
for Digital Images

Alex Dodge
New York University - ITP
721 Broadway
New York, NY 10003
+1 212 998-1880
alex.dodge@nyu.edu

Akira Shibata, PhD
Generative
1205 Manhattan Ave, Ste. 2-4-4
Brooklyn, NY 11222
+1 718 407-0874
akira@generativeinc.com

ABSTRACT
The amount of personal data in the form of social media, email, music, video, and personal photos continues to grow at a rapid pace. While some advances have been made in the more conventional methods of indexing and searching large catalogues of images, few approaches address the need for a system that enables users to explore their personal libraries in a non-linear and dynamic way. In this paper we propose an interactive system that enables users to explore their personal photographs by dynamically organizing them based on multiple image attributes such as, but not limited to: time, location, face recognition, color, as well as social attributes. A prototype system and interface were developed. The underlying processing of image data employed a suite of algorithmic methods including Principal Component Analysis, face detection and recognition, and histogram analysis.

Categories and Subject Descriptors
H.3.3 Information Search and Retrieval: [Clustering, Information filtering]; H.5.2 User Interfaces: [Graphical user interfaces (GUI), Prototyping]

General Terms
Algorithms, Experimentation, Human Factors.

Keywords
Photo clustering, face recognition, photo tagging, annotation, cluster annotation, visualization, image browsing, image retrieval

INTRODUCTION
Personal photo albums continue to grow at an increasing rate due largely to the integration of point and shoot quality photo capabilities with mobile devices such as smartphones making it easier for individuals to take photos. The estimated number of digital photos taken by Americans in 2006 was 53 billion photos or roughly 177 per person. In 2011 those numbers jump to 80 billion and 255 respectively. By 2015 it is expected that Americans will take 105 billion digital photos or 322 per person [1]. The advent of social media and collectively sourced catalogs of photos also contribute greatly to the amount of relevant photo content available to users. These factors are resulting in an exponential growth of personal photo album size. It is estimated that in 2011 the popular social networking site Facebook [2] was home to nearly 100 billion digital photographs. With an average Facebook user having around 345 friends, each with an average of 282 shared photos; the shared network of potentially relevant photos is numbered at nearly 97,000 digital photos per user [3].

Related Work
Systems capable of managing large collections of digital images are available but as the collections grow larger the ability for users to explore and ultimately enjoy their photo albums becomes increasingly cumbersome using existing methods of indexing and annotation. Traditional methods of manually annotating photos, either one photo at a time “selective annotation” or by larger groups of photos “batch annotation” becomes time prohibitive as the number of photos and possible categories increase. A great deal of research has been devoted to fully automated approaches to photo annotation [4,5,6] that attempt to use face recognition or event detection to annotate a collection of photos with little or no intervention on the part of the user, however these approaches have been largely unsuccessful due to both the immense technical difficulty that the task presents as well as the variance among users’ qualifiers that dictate useful forms of annotation.

Existing software solutions continue to migrate toward web-based products and services including Flickr[7], Picasa[8], Facebook, among many others. The advantages of web-based services include centralized storage of personal photos that may originate from multiple devices and locations. Most of these services offer the ability to tag people in photos and when available, reference existing Exchangeable image file format data (Exif data) [9] such as time and location, however the method of browsing and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference’10, Month 1–2, 2010, City, State, Country. Copyright 2010 ACM 1-58113-000-0/00/0010...$10.00.
exploring a collection of photos is limited to sub-collections or “albums” created by the user which are typically displayed sequentially.

Existing software solutions continue to migrate toward web-based products and services including Flickr[4], Picasa[5], Facebook[6], among many others. The advantages of web-based services include centralized storage of personal photos that may originate from multiple devices and locations. Most services offer the ability to tag people in photos and when available, reference existing EXIF data such as time and location, however the method of browsing and exploring a collection of photos is limited to sub-collections or “albums” created by the user which are typically displayed sequentially.

Our Method

We propose that an alternative photo management solution should be able to facilitate the ability to efficiently search for and locate a specific photo or group of photos, what we will refer to as “linear exploration” or “linear search” characterized by a user’s preexisting knowledge of a specific photo or criteria that is being searched for, but in addition it should also facilitate the ability to freely and efficiently explore a collection of photos when a user has no predetermined or known query which we will refer to as “non-linear exploration” or “non-linear search”. Our system, Kioku, addresses the problem of non-linear search through the combination of information analytics and an interactive user interface. We evaluated our system using a test set of 3000 family photographs from taken by a single family over the five years.

1. Data Analytics and Visualization

Kioku uses a method of dimension reduction known as Principal Component Analysis (PCA) as way of displaying photographs taking into account specific user-defined data filters. PCA is an established unsupervised method of exploring multivariable data sets [10]. We chose PCA as our primary analytical method for its ability to let a selected set of digital photographs self-organize based the distributed variance of the photographs’ attributes. This means that a large set of digital photographs, each with multiple attributes like “time”, “location”, “persons”, and “average color”, can be organized visually based on the most distinguishing attributes among them.

2. UI and Selective Filtering

Kioku’s UI allows users to explore a collection of images that have been arranged according to our analytical method. In addition to general navigation throughout an arrangement of photographs Kioku offers the ability to adjust filter criteria, which results in the real-time reconfiguration of the arrangement of photographs.

Data Analytics and Visualization

Kioku performs PCA based on the following variables:

- **Face Recognition fit values** (derived from an Eigenfaces method)
- **Time** (derived from Exif data)
- **Location GPS** (derived from Exif data)
- **Color** (represented as brightness, contrast, average color, entropy)

The applied Principal Component Analysis (PCA) method takes the photo-derived multivariable data and transforms it to a new coordinate system so that the greatest variance or most discernable features in the data are projected on the first and second principal components in a two dimensional representation or even three dimensions including the third principal component. The resulting x and y coordinates are used to locate the corresponding photographs as thumbnail images on a canvas. This method of organizing photographs makes it possible to visualize complex variables and high dimensional space in a simplified two-dimensional layout.

A Mathematical explanation of PCA describes an orthogonal transformation to convert input variables: $\mathbf{X} = [x_1, x_2, ..., x_m]^T$ into new uncorrelated variables or **Principal Components**: $\mathbf{Y} = [y_1, y_2, ..., y_m]^T$. This transformation can be expressed as: $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$, in which the columns of the matrix $\mathbf{A} = [A_1, A_2, ..., A_d]^T$ are the first $d$ eigenvectors, sorted according to the largest eigenvalue, $\lambda_1$, in the covariance matrix $\mathbf{X}$, with $d < m$. Here the matrix $\mathbf{A}$ consists of the majority of the information. We can represent the original data with just $d$ values rather than $m$, giving a much smaller data set.

When a user, in Kioku, selects only two or fewer image attributes to explore a collection of photographs with PCA is not used and the single or two dimensions are represented on the canvas simply plotted on the $x$ and $y$ axes. Figure 1. shows a collection of images filtered on time of day and year.

![Figure 1. shows a collection of images filtered on time of day and year.](image)

When three or more dimensions are selected PCA is used to describe the collection of photographs. Figure 2. shows the same collection filtered on average color, contrast, and month.
Figure 2. shows the same collection filtered on average color, contrast, and month.

Faces

Our method of deriving face fit values combines a typical Haar-like feature classifier [12] to detect faces in a collection of photographs after which fit values can be established using an Eigen Faces and Fisher Faces method based again on Principal Components (PCA) and Linear Discriminate Analysis (LDA) [11].

When new digital photos are entered into Kioku information is extracted from the images and stored as image signatures. Among this information is face data. Faces are first detected using a conventional Viola–Jones object detection framework [12] that uses a cascade architecture to validate the presence of a face. Once a face is detected it is parameterized to a basis of eigenfaces [13] or the values derived by performing singular value decomposition (SVD) on a set of training faces. Again this process uses Principal Component Analysis (PCA) to reconstruct a face using eigenfaces [Fig.3], which are the principal components [14]. The eigenfaces serve as a distinct signature with which distances to other faces can be established. Faces that are similar or of the same person should have similar eigenface composition and therefore located closer to each other in Euclidean space which forms the basis of the face axis in Kioku.

Time and Location

Time and GPS location data are both extracted from the image file’s embedded Exif data.

Color

Color pixel information is derived from the RGB color channels of each image file. Kioku uses a number of different color metrics to base distances on including color histograms or frequency of colors in each image [15].

Image contrast is measured defined as RMS Contrast [16]:

$$\sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I_{ij} - \bar{I})^2}$$

Where pixel intensities $I_{ij}$ are the $i$-th $j$-th element in a pixel array such as an image converted to grayscale of size $M$ by $N$. $\bar{I}$ is the average intensity of all pixel values in the image.

Average color and color variation or color entropy are also used to establish metrics for images. Color entropy is described using Shannon entropy [17]:

$$H = k \log \left( \frac{1}{p} \right)$$

Where the information entropy in an image is $H$ and $k$ the units, in this case color pixel values.

UI and Navigation

Kioku’s browser-based user interface [Fig.4] based on the Google Maps API [18] allows users to navigate a two-dimensional space by clicking and dragging in any direction and zooming in and out using either the on-screen zoom control, mouse wheel, or track pad method. Navigation is also available for touch devices.

Figure 2. Eigenface reconstruction

Figure 4. Kioku interface
Users can define specific viewing or filtering criteria via menus [Fig.5]. The UI menus allow users to filter selected people available in a collection of images, specific time ranges, locations, and even color ranges.

After an updated set of criteria is selected, PCA is performed again and the collection of images is updated accordingly. This offers users a fast way to dynamically explore a collection of images, quickly comparing different sets of criteria.

**CONCLUSION AND FUTURE WORK**

We are excited to continue development on Kioku toward a web-based tool that people can use to explore not only their collections of personal photographs but image archives of any type.

**ACKNOWLEDGMENTS**

We would like to extend our thanks to Phillip Wagner for his Python implementation of face recognition with OpenCV and his generous advice on PCA, Jake Porway for his soft spoken guidance on Latent Dirichlet Allocation, Clay Shirky for his ability to enthusiastically see beyond, and of course Heather Dewey Hagborg for making this paper happen.

**REFERENCES**