August 23, 2018

# Surfacing images of a collection based on device context

Matthew Sharifi

Katsiaryna Naliuka

Abhanshu Sharma

# Surfacing images of a collection based on device context

ABSTRACT

Users have extensive collections of digital photographs or videos on their devices. This disclosure describes techniques to select images from a user's collection that are relevant to the current context of the device and the user, and display the selected images on a device. The images are labelled with attributes relevant to the context at the time of capture of the images. At a later time, display of past images is triggered based on a variety of conditions related to current context and device activity. Past images having attributes that match the current context are selected and displayed. Displayed images are relevant to the user's current context, allowing more interesting and relevant images to automatically be presented to the user.

KEYWORDS

- surfacing images
- photo library
- anniversary images
- device context
- match context

## BACKGROUND

Users capture extensive numbers of images such as digital photographs with their mobile devices. Such images are commonly backed up to a server, e.g., using a cloud storage service. Many such images are never viewed again by users, given the size of a typical user's photos collection. Some applications and services include features that automatically find images in a user's collection that were captured on the same day of the year in the past, e.g., one year prior to the present day. Such a feature allows a user to "rediscover" images that may not have otherwise been found and viewed by the user. However, using such techniques, the relevance of the images to the user's current context is often quite limited. For example, the images may have been captured in a different context of the user, e.g., different physical location, different type of activity, etc. This can lead to such surfaced images having a reduced impact to the user's experience, and/or missed opportunities to surface images more relevant to the user's current context.

## DESCRIPTION

This disclosure describes techniques that intelligently surface images from a user's collection that are related to the current context of the user and of the device being used to view the images. Described techniques assign context attributes to images at or near the time of capture. Later, these context attributes are compared to attributes of a current context of a device to determine which stored images are relevant to the current context. Images that are found to be relevant to the current context are displayed to the user. Thus, described techniques surface image memories taking into account a user's fine-grained context, and allowing a set of images to be displayed which are highly relevant to the current moment.

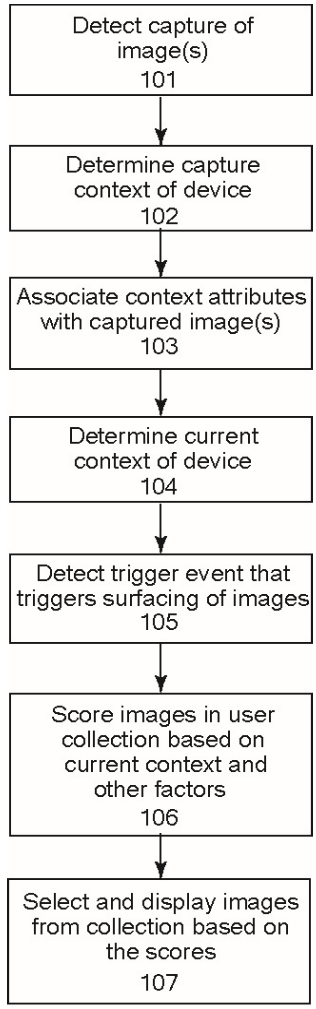


Fig. 1

Fig. 1 illustrates an example technique for associating context information with images, determining a current context of a user device, and presenting images from the user's collection that are relevant to the current context.

Capture of one or more images by a device is detected (101 in Fig. 1). For example, the user may have taken a photo using a camera of a mobile device. In some cases, multiple images may be captured.

A capture context of the device is determined (102), if user consent has been obtained. This is the context of the device when it captured the image(s). For example, the capture context can be determined based on a variety of data sensed by the device and/or accessible user data that the user has consented to be used for determining context. All of the context recognition and determination is performed locally on the device. For example, data from GPS sensors, motion sensors of the device (accelerometers, gyroscopes), microphones, sensors of light levels, etc. can be used, as permitted by the user. In some examples, objects or other characteristics depicted in the pixels of the captured image are detected. All of this data is processed, e.g., using a machine learning model (such as a deep neural network) to recognize a specific capture context of the device.

The capture context may be described by a variety of different context attributes, the collection and use of which have been consented to by the user. In some examples, the context includes attributes such as the time of capture (including particular time, time of day (morning, evening, etc.), day of the week, month of the year, year, etc.). Context attributes can include the geographic location of the device at time of capture. In some examples, the location can be a named place instead of geographical coordinates, e.g., “home,” “work,” “park,” “restaurant,” etc. as determined based on location and other user data. Context attributes can also include the weather at time of capture (e.g., obtained from accessible information sources over the Internet, etc.). Context attributes can include an activity of the user at the time of capture. Such activity can be determined using location, calendar information, and any other user data the user has consented to be used. Such activities can include hiking, running, shopping at a store, eating at a restaurant, watching a movie at a theatre, etc.

In further examples, context attributes can also include co-presence information indicating other users at the location of capture, e.g. users who are within a particular distance of the device. Such co-presence information can be derived, for example, from local network information (e.g., detecting other user devices using a local network), centralized network information (e.g., a server that has logged locations of other user devices), audio signals (voices detected and matched to users), user calendar information, etc.

Context attributes can also include image features detected in the captured image. For example, detection or recognition techniques can be utilized, if user consent has been obtained, to detect types of objects in the pixels of the image. Such objects can include persons (e.g., without identifying the identifications of such persons), objects, landscape features, etc.

Context attributes can include audio detected in the device environment at time of capture. In some examples, the type of audio can be detected, e.g., conversation based on human voices, traffic, birdsong, waterfall, etc. Another form of audio context attribute is ambient music that is playing in the device environment at the time of image capture. Context attributes describing detected music (e.g., song title, album title, artist, music genre, etc.) can be determined, e.g., based on recognizing music and accessing music database information locally or remotely over a network.

In some examples, the device can determine the capture context (e.g., determine context attributes) in response to the capture of the image(s). Further, the device can determine its current context periodically, or continuously, during operation or at particular times of device use (e.g., using background processes of the device), and can store the already-determined context at the time an image capture is detected. Various types of context attributes can be determined at the time of image capture, and can be determined before and/or after image capture.

The determined context attributes are associated with the captured image(s) (103). For example, the images are annotated with labels that describe the context attributes. Each label can include text (or other descriptor) that describes one or more context attributes, such as a time, a name or identification of a location, activity, sound / audio environment, person, music, etc. The determined labels are associated with the captured image(s) in storage, e.g., as metadata. Labels can be stored with other metadata of the image such as capture time and location.

Some labels may directly describe the determined context attributes, such as a time, location coordinates, names of detected objects in the captured image, etc. Some labels can be generated based on particular context attributes. For example, if objects such as books are detected in the captured image, a location label can be generated as “library” or “bookstore” (e.g., as determined based on location coordinates, based on a database of location names associated with particular objects, etc.). Multiple types of context attributes can be examined to reduce ambiguity in determining context labels. For example, books detected in the image indicate the context location is a bookstore and is not a restaurant located above the bookstore in the same building (which location coordinates did not distinguish).

In some examples, particular labels may or may not be assigned to the captured image based on particular context attributes, e.g., as instructed by stored rules or user preferences. For example, if a location context attribute indicates the device is in a bookstore at image capture time, then a label describing the music that is playing in the device environment can be omitted from being assigned to the image, as being not relevant. In contrast, a location or activity determined to be a wedding, where music is designated to be relevant, causes one or more labels describing the music to be assigned to the image.

Labels that are associated with an image are stored and synced to user storage on a remote server, e.g., in the cloud, if user consent is obtained. If some or all labels and attributes are private to the user, then those labels and attributes are only stored locally on the device, while the image itself is synced with remote storage if user consent is obtained. A user is provided with options to control which labels (or types of labels) and attributes are to be synced to remote storage.

To display images at a later time, a current context of the device is determined (104), if user consent has been obtained. The current context can include context attributes, which can be determined similarly to context attributes as described above for the capture context. Labels can be determined based on the context attributes for the current context, similarly as described above.

If user consent has been obtained, the device can periodically or continuously check or update the current context of the device, e.g., periodically based on time and based on events such as receiving user input or data at the device. In some examples, the current context of the device is determined at least partially in response to a trigger event being detected as described below.

For example, the current context of the device can be described by context attributes (use of which is consented to by the user), including current time, location, activity, presence of other users, audio environment, motion of the device, weather, etc. similarly as described above for the capture context. In some examples, if user consent is obtained, the current context can additionally include attributes describing current device usage by the user, such as applications or other programs being used on the device, type and source of data received by the device (e.g., messages, sender, images, map data, etc.), notifications provided by the device, or any device activity including device functions and operations initiated and/or used by the user.

A trigger event is detected (105). The trigger event is any type of event that can be detected by the device and which has been defined, once detected, to trigger the surfacing of one or more images from the user's collection. Various types of trigger events can be defined. In some examples, trigger events can be defined by the user.

In some examples, trigger events can be based on the determined current context of the device. For example, based on the current time, location, activity, device usage, or other determined context attributes of the current context, a trigger event can be detected. A trigger event can be a particular time of day or day of year (e.g., birthday or wedding anniversary), a particular location (e.g., vacation spot or friend's home), a particular type of activity (e.g., jogging), a particular title of ambient music detected, and so on, and/or can be a combination of any two or more context attributes (e.g., jogging at a particular location at a particular time of day).

Trigger events can be based on current context of the device related to use of, or activity on, the user's device by the user. For example, use of a particular application on the device, powering of the device, or display of particular data by an application, can be defined as trigger events. For example, if the user is using the device to browse a restaurant-booking application on the device and looking for a restaurant, the initiation or use of the application can be defined as a device context that is a trigger event.

A trigger event can also be defined as a change in the current context of the device, e.g., a change in any detected context attributes of the device context, a change in one or more particular context attributes, or a change in a threshold number of context attributes. For example, the change in context can include a user starting a new activity on the device relative to an old activity (e.g., from reading chat messages in a chat application to booking a holiday on their phone in a different application), the user arriving at a specific location, or a specific ambient song starting to play in the device environment, which is detected by the device.

Images in the user's collection of images are scored with respect to the current context and other factors (106). In some examples, a triggering component of the device is initiated in response to detecting the trigger event. The triggering component can be a machine-learned model or other scoring system which scores or queries over the user's images for those images which are relevant to the current context detected.

The scoring system can assign a respective total score to each image in the collection, where the total score is a combination of multiple sub-scores, and each sub-score is determined for a respective one of the context attributes. If a particular context attribute of the current context matches or is similar to a context attribute of an image, then the sub-score for that attribute can be added to the total score for that image. All the corresponding context attributes can be compared, including time, location, activity, detected audio, users present, etc. Different weights can be associated with different context attributes to adjust their sub-scores. Weights can be varied based on particular types of context attributes being present or based on other conditions. A machine-learning model can be trained based on matches of attributes.

In some examples, the scoring of images can be associated with the type of trigger event that is detected. For example, if the trigger event is a first type of device activity, then context attributes (e.g., time, location, etc.) may have sub-scores of first values, and if it is a second type of device activity, then those context attributes may have sub-scores of different, second values. For example, if the user is using a map application, then context attributes related to location can be weighted higher than if the user is using a chat application.

The desirability of interrupting the user with surfaced images can also influence the scoring. For example, particular device context attributes can be designated as interruptible, such as using particular applications (e.g., a chat application), particular locations (e.g., home), particular activities (e.g., shopping), etc. Other context attributes can be designated as uninterruptible (e.g., an activity that is a meeting at work, a location at a movie theater). If the current context is determined to be interruptible, then scoring can be performed for images to allow surfacing of images, e.g., scoring can have a normal weight. If the current context is determined to be uninterruptible, scoring can be omitted to prevent surfacing of images, or scoring can be weighted lower to make the surfacing of images unlikely, e.g., until the context has changed to include interruptible attributes.

A "novelty" characteristic can also be determined for each image, which can contribute to the total score for the image. The novelty characteristic indicates how often the image has been displayed in the past (frequency), e.g., based on context, and/or indicates the most recent display of the image automatically by the system or manually by the user. For example, if an image has not been displayed for more than a threshold amount of time and under a threshold frequency, the total score for that image can be increased compared to an image that has more recently or more frequently been displayed.

The method may also use personalization factors specific to the user to adjust triggering events and/or scoring. For example, personalization factors can take into account which images a user has previously selected for viewing, shared with other users, deleted, edited, etc., in order to adjust the trigger events and/or scoring over time based on inferred user preferences. This personalization may use a form of transfer learning or reinforcement learning.

After the scoring, particular images from the user's collection are selected and one of more of these images are displayed (107) on the device based on the scores (or similar determination), e.g., the particular images are surfaced for the user. For example, all images, or a particular number of images, that have the highest (or best) scores can be selected as the particular images. In some examples, only images having a total score that is higher than a threshold score are selected. If no images have a score higher than threshold, no image are displayed.

At the time of display, multiple images from the user's collection can be combined into a collage or other format. Other processing can also be automatically performed by the system to selected images, e.g., applying color, brightness, contrast, blurriness, or other filters or effects to images.

A user's images also can include videos that include multiple frames and/or audio, and/or can include still images that are stored with accompanying audio clips. Surfaced images from the user's collection can be such videos or images with audio. For example, audio data that is stored and associated with the particular image can be surfaced and output by the device along with the display of the particular image.

Examples

Some examples of device contexts and surfacing of user images are described below.

- When a user is using the device to book a steak restaurant, this is detected as a trigger event and the system creates a collage of other steak restaurants that the user visited in the past.

- When a user visits or books a restaurant that the user last ate at a few years ago, this is detected as a trigger event and images are displayed from the user's last visit to that restaurant.
- When a user goes to a particular place with a group of longtime friends, old images are selected and displayed that were captured at a previous time the same group was all together in the same place (user co-presence).
- When a user hears a song played on the radio that was played at his or her wedding, the user device displays set of images from the wedding.

Many other examples are possible. In each of these cases, the images that are selected and displayed are highly relevant to the user's current context and, consequently, are much more likely to be useful or interesting to the user than images selected without accounting for the context.

Machine learning models that can be used in the system are trained and implemented upon user permission to access user data that serves as input to the models. Users are provided with options to indicate their permission or denial of permission for access to various data, e.g., images, audio data, and other content in the user's image library, user-specific factors such as prior sharing history, previous device actions, etc., and contextual factors such as time, location, application in use, etc. In using the models, use is made only of user-permitted data. One or more of the models are not implemented, if users deny permission. In various implementations, training is performed based on generalized data that is not attributable to individual users, and/or training is performed only locally on a user device with user data, e.g., using a federated learning approach.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if a user device is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

## CONCLUSION

This disclosure describes techniques for surfacing and displaying images on a device from a user's collection that are relevant to a user's current context, e.g., activity, location, time, environment, etc. Images captured by the user are associated with labels describing context attributes at the time of capture of the image, and context attributes of a current context are compared to attributes of the user's images. Images in the user's collection that correspond to attributes of a current context of the device are selected and displayed on the user device. The displayed images are more relevant to the user's current context than images selected with techniques that surface images by matching only a single attribute. Further, the time that the user spends searching for particular and relevant images is reduced, improving user satisfaction.