Guiding Capture of New Photos Based On Published Photos

Amy Zhang
Cliff Sze

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Zhang, Amy and Sze, Cliff, "Guiding Capture of New Photos Based On Published Photos", Technical Disclosure Commons, (July 09, 2020)
https://www.tdcommons.org/dpubs_series/3413

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
Guiding Capture of New Photos Based On Published Photos

ABSTRACT

This disclosure describes techniques for presenting recommendations that guide users to capture high-quality photos based on existing crowdsourced, popular, and/or shared photos. Machine learning models are used to find high-quality public photos that have similar attributes to a photo the user wishes to capture. These public photos are displayed as examples for the user and are used to guide the capture location, photo framing, and/or camera settings during the photo taking process. Described features inform and guide a user in a process of capturing high-quality photos and are easily accessed in a device used to capture the photos.

KEYWORDS

- photo capture
- framing recommendation
- pose recommendation
- location recommendation
- photo examples
- user guide
- crowdsourced photos

BACKGROUND

Popular personal devices such as smartphones usually include a camera that enables the user to capture digital photos. However, some photos captured by users may have poor or awkward view angle, framing, and/or poses of subjects. When posting photos to social media environments or sharing via online photo sharing tools, people are concerned as to the quality
of their photos. Capturing scenes with aesthetically pleasing attributes is important for many users capturing and posting photos.

**DESCRIPTION**

This disclosure describes features enabling a user to capture high quality photos using a camera device. Described techniques present recommendation photos obtained from the large resource of public online photos to assist users in capturing new photos, e.g., as suggestions that are relevant to a desired photo subject. The user may be guided in the various steps of a photo capture process. Described features can be provided in a separate app, or can be built into existing applications, such as a camera app, a photo library app, a map app, etc., or can be implemented in the firmware of a hardware device, e.g. cameras.

The techniques described herein are implemented upon specific user permission to access a user’s data, e.g., photo library, location, etc. Users are provided with options to grant permissions to and/or to disable features entirely. The user can enable or disable techniques discussed herein for particular subjects, sources, locations, types of photos, time periods, or other conditions.

**Obtaining and classifying photos to use in recommendations**

The recommendation system first obtains an initial set of photos from which recommendations are selected. The initial set of photos can be a large set of photos obtained from available public sources on the Internet. Information associated with these photos can also be obtained, e.g., a source of each photo (website, photographer, etc.) and other information described below.

The recommendation system identifies a set of high-quality photos from the initial set of photos using machine learning model(s) that have been trained with training data (e.g., photos,
videos, etc.). The models include both supervised and unsupervised learning. Supervised learning are trained with user-permitted photos that have been manually rated for quality (e.g., number of “likes” or “shares” in social media, or votes/recognitions in contest or discussion forums, or using natural language processing (NLP) to rate photos based on the comments), such that the models can detect high quality photos (e.g., based on brightness, blurriness, framing, etc.) in the initial set of photos. The models are retrained periodically using additional manual ratings to include new training data and to reflect changing aesthetics and trends.

Unsupervised learning such as clustering is used to identify similar images, or images about the same subject matter. The models can be implemented on a user device, e.g., a mobile device. If user permission is obtained, the models can be fully or partially implemented on a server in communication with the user device. In some cases, heuristics and rules can be used instead of or in addition to machine learning models to identify high quality photos.

Several indicators associated with photos can also be used to determine which photos in the initial set have high quality. Some examples of such indicators are a large amount of positive ratings (e.g., likes) for the photo or a large amount of shares (including re-shares) of the photo on public social media platforms. Natural language processing (NLP) can also be used to run sentiment analysis from user comments to create positive, negative or neutral ratings. Another indicator is the credibility or reputation of the photographer of the photo, which can be obtained from web rankings and lists or by using algorithms that rank web pages and photos based on references, links, shares, etc. Additional indicators can include particular usage of photos, recommendation of photos, or certifications of any type for photos. For example, photos chosen as a background in a commercial service, published in an article on a
well-known website, entered in or won photography competitions, etc., are determined as more likely to be high-quality photos.

After the set of high-quality photos is determined, the photos in this set are evaluated and classified as location-based photos and non-location-based photos using machine learning clustering techniques. The subject of a location-based photo is a feature that is tied to a particular geographic location (a location-based feature). Example features include a building, a landmark, or a natural landscape. A photo is classified as location-based if its subject is this type of feature (e.g., as detected by the machine learning models). The geographic location, as well as name, description, and other information of the location, are stored in association with such a photo. This information may be associated with the photo, or online services can be used to identify or verify locations of photos.

Non-location-based photos are directed to features that are common objects (non-location-based features), e.g., benches, tables, chairs, cars, steps, doors, trees, etc., that are not specific to a particular geographic location. A photo is classified as non-location-based if it subject is this type of feature (and if there are no prominent location-based features in the photo). Names or descriptions describing these features are stored in association with the photo. Identification of objects need not be accurate for every object in a photo, e.g., unique objects. Non-location-based photos can be used in recommendations in which the capture location is not important, e.g., photos emphasizing human poses. Note that each photo can be classified for multiple features including both location-based and non-location-based features at the same time. For example, a sculpture that is part of the permanent exhibit at a museum is both location-based (the location of the museum) and non-location-based (the sculpture can be seen as a standalone object).
Other features of the high-quality photos in the set may also be detected and stored for use in some types of recommendations. For example, humans in the recommendation photos can be detected using facial or body recognition techniques, if consent has been obtained to perform such recognition. In this example, photos would be classified as having humans vs. not having humans.

Creating and presenting recommendations from photos

Recommendations for taking a photo are created from the set of high-quality images, and are presented to a user on the device based on various conditions that may occur. For example, the user can open a recommendation app, request a recommendation feature, travel to a particular location, etc. Various types of recommendations are detailed below.

Location recommendations

A location recommendation provides a suggestion of a precise location at which to position a camera to capture a desired photo, e.g., of a designated subject (such as a location-based feature) or at a designated area. The quality of a photo depicting a location-based feature may be dependent on the precise location at which the photo is taken, which influences the view angle of the photo. For example, taking a photo of the Empire State Building on the sidewalk in front of the building is not usually desirable, since the camera is located too close to the building, the building is very tall, and the amount of sidewalk space is limited. As the time/date affects the sunlight, as one of all the features considered, such aspects are taken into account when determining the location recommendation. Details of time/date consideration are provided in later paragraphs.

Recommendation photos are selected from the set of high-quality photos to determine the location recommendation. The selected photos show a location-based feature that matches
the desired feature of the user. The selected photos may have captured locations that match a geographic location that is designated by the user or is the current location of the user’s device (if the device is at a desired location to take a photo). For the above example, location-based photos are selected as recommendation photos that depict the Empire State Building, have capture locations corresponding to that building, or are labelled “Empire State Building.”

Recommended camera locations can be presented as graphical locations on a displayed map and/or as names (if a name is available). For example, in the Empire State Building example, a recommended location can be an icon positioned at nearby Madison Square Park or at a tall building that has a view of the Empire State Building, such as Rockefeller Center. The selected recommendation photos may also be displayed, e.g., as a thumbnail next to the location, to provide the user with an example photo captured from that location.

The capture locations of the recommendation photos may be determined from location data, e.g., location tags in metadata of these photos, input manually or determined automatically (e.g., using GPS or other sensor) or using object recognition. In some cases, relative angles to a location-based feature in a photo (e.g., to a facade of a recognized landmark) are used to derive the capture location of the photo. Some cameras may tag photos with metadata that includes a direction the camera was pointing at the time of capture, which can be used to determine the capture location of an associated photo.

Many photos of the same location-based feature are commonly taken by different users at about the same camera location, e.g., to get a particular view of a popular monument or landmark. If a recommendation photo has a capture location that matches the capture locations of a large number of other photos of various users, the confidence level of the location
recommendation can be increased. Capture locations that are statistical outliers can be discarded as being in error due to GPS drift or other inaccuracies.

The user can request multiple different locations and display recommendation photo(s) taken at those different locations. Alternatively, the user can simply request top rated photos of the desired subject matter, and then choose the location based on the user’s preferred photos. Different locations may provide users with estimates of different views of a location-based feature, and help the user make a decision about where to go.

**Framing recommendations**

The recommendation system also provides a framing recommendation to a user. If the user goes to a recommended location as described above, or simply wants to take a photo of objects of interest, the user still must point the camera to frame the desired subjects in a way that produces a high-quality photo. A framing recommendation indicates a position and/or orientation of a camera to obtain a high-quality photo of the desired subject.

To promote flexibility, the framing recommendation can present multiple recommendation photos, selected from the set of high-quality photos, that were taken at the recommended location or of the specific objects, at different view angles or camera settings, to give the user ideas for photo composition.

If no location recommendation has been made (or in the case of recommendation based on a combination of location-based and non-location-based feature), the framing recommendation presents recommendation photos selected from the set of non-location-based high-quality photos. The recommendation photos show the same types of objects desired to be in the new photo. For example, the user can point the camera to a desired scene including objects, these objects can be recognized and matched to photos in the set of high-quality photos,
and the matched photos are displayed as recommendation photos. The user can also input text descriptions of types of objects to be included in recommendation photos. These text descriptions can be matched to the tags of the high-quality photos.

The user can also request more hands-on guidelines that directly indicate a camera position for a new photo. The framing recommendation includes guidance output that directs the user to look at a viewfinder image displayed by the camera and move the camera around until the framing of objects in the camera view is similar to the framing of a selected recommendation photo. As soon as the framing of the camera view partially overlaps with the framing of the recommendation photo, a target frame is displayed in the camera view that represents the framing of the recommendation photo. The user is directed to move the camera left, right, up, or down; move towards or away from the subject; and/or adjust zoom settings etc. until the camera view matches the target frame, e.g., by displaying directions within the camera view.

![Fig. 1: Guidance output of framing recommendation](image)

Fig. 1 shows one example of guidance output displayed in a camera view. The current camera view (100) shows a particular portion of the physical scene area (102) based on the
current position and orientation of the camera. A target frame (104) is displayed to guide the user to move the camera to match the position and size of the camera view with the position and size of the target frame. Here, the user can pan the camera to the right and zoom in to match the camera view to the target frame. Instructions (106) displayed in the camera view tell the user how to match the camera view with the target frame.

These features can use an online machine learning system that actively processes an image in the camera view as it is changing, by comparing that image to the recommendation photo. An exact match of camera view and photo is unlikely due to the recommendation image being captured under different circumstances and/or edited. An online system may be computationally expensive if serving data from a server system to a client device. Offline systems (or edge computing) can alternatively be used, where guidance output can be based on camera direction of recommendation photos determined from direction tags and/or derivation of direction from content of photos.

*Pose recommendation*

Subjects of a photo, such as persons, can be in a variety of positions and poses with respect to other features in a photo, which can affect the quality of the photo. The recommendation system can provide suggestions for poses of subjects, if requested by the user. One option is to present recommendation photos as examples having subjects and features (e.g., objects) similar to features of the new photo, e.g., features currently shown in the camera view. The user can view the recommendation photos and adjust poses of subjects to be similar for the new photo. For example, the recommendation photos may show poses relative to objects, such as poses on benches or steps.
In another option, a user selects a recommendation photo to use as a guide. The type of subject is recognized and a target shape matching that type of subject is displayed in the camera view, positioned with respect to the objects in the scene to match the selected recommendation photo. The subject moves or is moved within the scene to match the position and configuration of the target shape in the camera view. The target shape can be a cartoon equivalent of the user’s avatar.

![Figure 2: Camera view image showing pose recommendation](image)

Fig. 2: Camera view image showing pose recommendation

Fig. 2 shows an example of the system displaying an image (200) in the camera view of the camera (e.g., after matching camera view to target frame as in Fig. 1). A “carved out” target shape (202) is displayed in the camera view that shows a recommended pose for a person based on the recommendation photo. The subject person can move into the viewfinder frame and adjust their body to fit the target shape, allowing a photo to then be taken of the person in the suggested pose. For example, the user holding the camera can instruct the subject person to conform to the recommended pose. In another example, the camera can be set on automatic (e.g., when the photographer herself is a subject) and guide the subject to the recommended
pose with audio directions or visual directions in the case of a selfie/front-view camera, automatically capturing a photo when the subject (approximately) fits the recommended pose.

Additional recommendations

A recommendation can also include suggestions to take a photo at a designated time and/or under designated weather conditions. Scenery at a particular location may appear very different at different times of day, during different seasons, and/or under different weather conditions. If the user selects a recommendation photo as a guide, the recommendation system provides information on the time of year, time of the day, and/or the weather condition at the time of capture of the recommendation photo. This helps users plan ahead of time, for example, what time of the year to visit to capture the cherry blossoms in Washington DC. Time information can be obtained from photo metadata, from a photo app that was used to take the photo, from photo descriptions, or from similar photos having timestamps. Weather information can be obtained from online local weather forecast services and the weather history based on the timestamps of the photos in the database, or a machine learning model can identify the weather present in the photo based on its content.

In another option, if the user is at a recommended location, the current time and/or weather conditions at that location can be used to filter high-quality photos that have similar time and/or weather conditions. These filtered photos may be used as recommendation photos for the user in the current environmental setting.

Advanced users may be provided with other detailed information about a recommendation photo. For example, photo mode (horizontal, vertical, portrait, panoramic), photo aspect ratio (square, 2:3, etc.), zoom setting, aperture, and/or shutter speed characteristics of the recommendation photo can be presented. A user can adjust settings of the camera to
match the recommendation photo settings and learn which settings are appropriate to produce high quality photos. These types of information may be included in photo metadata (e.g., in EXIF tags) or identified based on attributes of the photo. For example, aspect ratio is based on dimensions of the photo, and portrait mode and aperture setting may be identified based on a blurred background.

After the recommendation system has been available for devices, with user permission, crowdsourced user data can be used in recommendations. For example, users can rate photos, help correct information on location, weather, time etc., This data can be used to update recommendations, thus making the recommendations more accurate.

Examples of use

- In one example, a user searches for photo recommendations at a particular location ahead of time. The user accesses a recommendation app or option on their device, which displays a geographic map of the user’s current location. The user inputs criteria, e.g., selecting an option, “inspire me,” selecting an option “near me,” or typing a query such as “Pisa Tower.” The app finds and shows recommendation photos that are the most popular or most highly-ranked photos that match the user’s criteria. The user can select a photo to view it at a larger size, select to display the next recommendation photo, or select to display the location at which a selected photo was captured. If the user sees a desired recommendation photo, the user can bookmark it, which stores a reference to the photo and its location and other metadata.

- In another example, the user is near a particular location, e.g., Pisa Tower, and opens the recommendation app on their device to see where to take photos. The app displays a map of the user’s current location, and previously bookmarked locations are highlighted in the map.
with icons of bookmarked recommendation photos at that location. The user selects a location, and information is displayed in the map for the location (e.g., name, address, etc.) along with associated bookmarked photos. The map directs the user to the location at which to take a photo. At this point, the user can select “Take a Photo” (e.g., the user has reviewed the recommendation photos and will take their own photo), or can select “Help Me Take a Photo Like This One” for a particular recommendation photo. The user points the camera of the device at the Pisa Tower, a target frame of the recommendation photo is displayed in the camera view, and the user adjusts the position of the camera until the framing of the camera view matches the target frame. The user puts the camera on automatic mode, moves into the scene and poses in the photo as desired. Or, the user selects “Pose Like the Model” to cause a target shape to be displayed in the camera view. The user moves into the camera view and conforms to the target shape, which triggers the camera to capture a photo of the scene including the user.

- In another example, the user arrives at a park that has a bench which the user wants to include in a photo. The user opens the recommendation app and selects the “inspire me” and “near me” options while pointing the camera at the bench. Recommendation photos (e.g., non-location based photos) of objects similar to the bench (e.g., with a person posing near the bench) are displayed. The remainder of the guiding process is similar as described above.

Machine learning models that are used to implement some of the described techniques using user data are trained and implemented only with user permission to access user data that serves as input to the models. Users are provided with options to indicate permission or denial of permission for access to various data, e.g., images, image metadata, video, and other content,
contextual factors such as time, location, application in use, etc. In implementing the described techniques, use is made only of user-permitted data, and certain techniques (e.g., ML models) are not implemented, if users deny permission. Model training is performed based on generalized data that is not attributable to individual users, and/or performed only locally on the user device with user data, e.g., using a federated learning approach.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein 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 the user is sent content or communications from a server. In addition, certain data is 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 is treated so that no personally identifiable information can be determined for the user, or a user’s geographic location is 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 has 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 to provide recommendations to users for taking high-quality photos. Recommendation photos relevant to a user’s desired photo subjects are provided as guiding examples, and these photos can also be used to guide the capture location, framing, and/or camera settings during the photo taking process. Described techniques allow a user flexibility in the amount of guidance received and thus the degree of creativity the user wishes to apply when taking photos.