October 09, 2019

INTELLIGENT MEDIA CREATION UTILIZING PASSIVE SONG DETECTION

Kody Ferguson

INTELLIGENT MEDIA CREATION UTILIZING PASSIVE SONG DETECTION

ABSTRACT

A system is described that enables a computing system (e.g., mobile phone, camera, tablet computer, etc.) to associate images with a sound that has been detected when an image was captured. The computing system may passively capture an audio sample when a user clicks a photo and compare the audio sample with a song database (e.g., as detected by one or more input device). If a match is found, information, such as the title of the song and the name of the musician, may be retrieved and correlated to the photo. A user may play a song associated with a photo by clicking on the title of the song within an information panel. Alternatively, with instructions from the user, the computing system may create a playlist of songs associated with photos. The computing system may also pair people recognized in the photo with the song associated with the photo to identify the songs that are played most often when the recognized people are together. In addition, with the user's consent, the computing system may generate a combined playlist of songs associated with photos of a given event when a user shares photos with other users. The computing system may also utilize the songs associated with the photos to create an intelligent movie, collage, home album, or animation.

DESCRIPTION

Image capture is a common feature for computing devices. Existing image capture device could include location information, time information, and image settings in a photo without any intervention, which creates a summary of a moment and helps to bring a user back to that moment. However, a moment is much more than a still shot. Fortunately, music has a way of invoking memories and helps to transport a person back to a moment seemingly forgotten. At such, it would

be desirable for a computing system to be able to associate an image with a sound that has been detected when the image was captured.

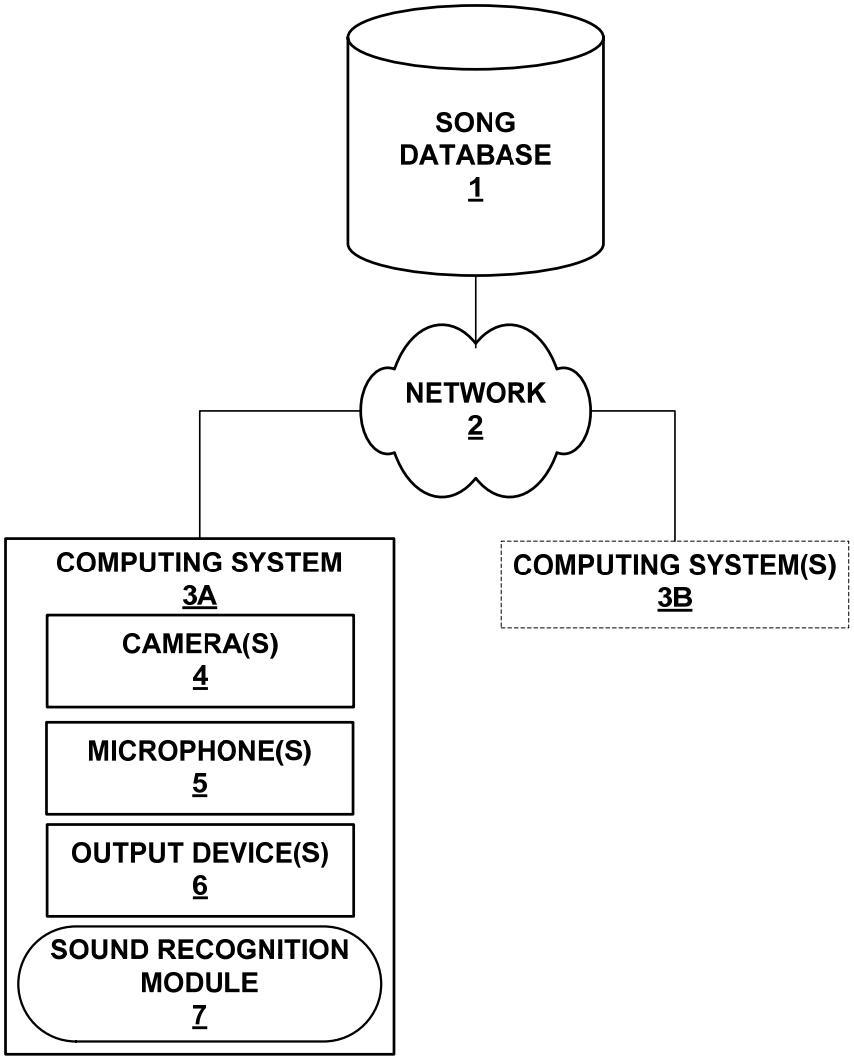

FIG. 1

A computing system, such as shown in the example of Figure 1 above, may contain a sound recognition module to associate with an image a sound that has been detected when the image was captured. The computing system may include, or otherwise be included in, a mobile device (e.g., smartphone), a tablet computer, a laptop computer, a camera, or any other type of suitable media capture device.

Computing system 3A may associate images with a sound (e.g., a song, melody, ringtone, or other types of sound) that have been detected when an image was captured at an event (e.g., wedding, concert, graduation ceremony, or the like). Computing system 3A may associate the recognized sound's information with the image and display to a user through a user interface. For example, the sound recognition model may passively recognize the song played at the user's wedding when a wedding picture was captured, and computing system 3A may associate the title of the song with the picture and display the title of the song in an information panel to refresh the user's memory of the wedding.

Computing system 3A of Figure 1 includes one or more cameras, microphones, output devices, and a sound recognize module to associate an image with a sound that has been detected when the image was captured. In the example of Figure 1, computing system 3A may access song database 1 via network 2. Network 2 may represent a combination of any one or more public or private communication networks, for instance, television broadcast networks, cable or satellite networks, cellular networks, Wi-Fi networks, broadband networks, and/or any other type of network for transmitting data (e.g., telecommunications and/or media data) between various computing devices, systems, and other communications and media equipment.

As a user takes a phone, microphone 5 may start to record an audio sample for a predetermined period. Sound recognition module 7 may then process and compare the audio sample with song database 1 via network 2. For example, microphone 5 may passively collect an audio sample of the background music when a selfie is taken and sound recognition module 7 may then compare the collected audio sample with song database 1 to determine the title and the artist of the background music. Sound recognition module 7 may be a software application. In such examples, sound recognition module 7 may be a native application or a web-based application.

Native applications may be provided by a first-party or a third-party developer and may be pre-installed or downloaded from an application market.

If a match is found, computing system 3A may retrieve information such as the title of the song and the name of the artist from song database 1 and may correlate the retrieved information to the photo. Computing system 3A may then display the retrieved information to the user within an information panel. As shown in example Figure 2, the information panel may display time, title of a song, artist name, camera setting, information associated with people recognized in the photo, location information or any other relevant information.

A user may play a song associated with a photo by clicking on the title of the song within the information panel associated with the photo. Once the user clicks on the song title, computing system 3A may access the song from song database 1 via network 2 and output device(s) 6 may play the song while the image is displayed. Examples of output devices include, but are not limited to, speakers, portable music players, or any other type of suitable output devices. In some examples, an additional user interface overlay may be added to the corner of the photo to show the user there is a song associated with the photo and provide an option to play the song directly from the photo. A download icon may also be included near the song title, which provides the user with a convenient way to download the audio related to the recognized song by clicking on the download icon.

Based upon user request, computing system 3A may create a playlist of songs associated with photos by selecting multiple photos. For example, suppose photos A, B, and C are associated with songs A, B, and C, respectively. Once a user selects photos A, B, and C, computing system 3A may suggest the user create a playlist consisting of songs A, B, and C. The user may also, at

any time, add or remove a song from the playlist. For example, the user may add song D after creating a playlist consisting of songs A, B, and C.

Computing system 3A may connect with computing system 3B via network 2, which allows the user to share photos and songs with other computing systems. For example, a user may share a playlist with a friend via network 2.

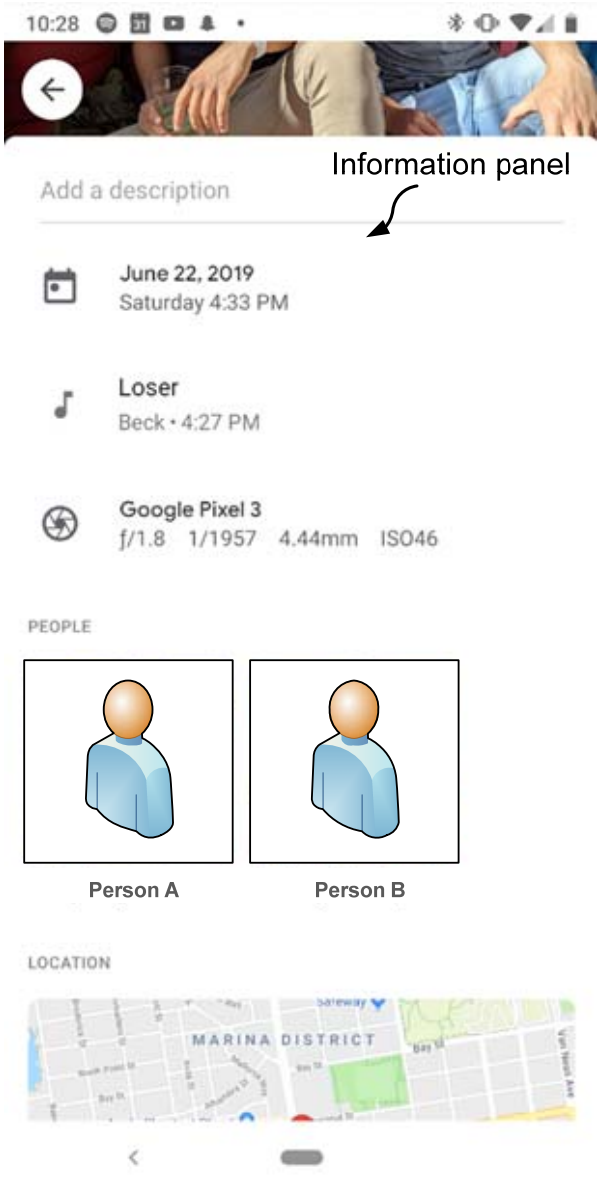

FIG. 2

Computing system 3A may also pair people recognized in the photo with the song associated with the photo to identify the songs that are played most often when the recognized people are together. As shown in Figure 2 above, computing system 3A may associate names of people recognized in the photo (identified as Person A and Person B in Figure 2), icons of people recognized in the photo, and songs that are played most often when the recognized people are together (identified as Loser by musician Beck in Figure 2) with the photo and display these information to the user within the information panel.

Based upon user request, computing system 3A may generate a combined playlist of songs associated with photos of a given event when a user shares photos with other users. Computing system 3A may use formation associated with the photos, such as location information, time information, people identified in the photos, or any other relevant information to generate the combined playlists of songs. For example, with the user's consent, the user may share photos with a friend. If two photos are identified as being taken at the same event, such as photos with the same location information and with same people identified, computing system 3A may generate a combined playlist of songs associated with the two photos. In this way, computing system 3A may supplement missing information for one picture by using information associated with other pictures took the same time of the same event.

Computing system 3A may also utilize the sound associated with the photos to create an intelligent movie, collage, home album, or animation. Computing system 3A may select relevant pictures of an event from a user's picture album based on keywords provided by the user. Computing system 3A may order the selected pictures based on the timestamp associated with the picture. Computing system 3A may then paly songs associated with the image while the image is displayed to create an intelligent movie, collage, home album, or animation.

It is noted that the techniques of this disclosure may be combined with any other suitable technique or combination of techniques. As one example, the techniques of this disclosure may be combined with the techniques described in US Patent Application Publication 2017/0272995A1. As another example, the techniques of this disclosure may be combined with the techniques described by Mark D. Wood "Matching Songs to Events in Image Collections," available at https://franz.com/agraph/cresources/white_papers/Kodak_Song-Images.pdf.