October 2021

Automatic Zoom Based on Image Saliency

Lin Chen

Wei Hong

David Massoud

Automatic Zoom Based on Image Saliency

Abstract:

This publication describes techniques for image-capture devices (e.g., smartphones, laptops, tablets) to automatically zoom in on object(s) of interest in a photographic scene. In aspects, through utilization of a machine-learned technique (e.g., a Visual Saliency Model), an Automatic Zoom Manager can detect a visual saliency region within an image, generate a bounding box enclosing the visual saliency region, and calculate an appropriate zoom ratio. The Automatic Zoom Manager can then implement the calculated zoom ratio to automatically achieve a suitable zoom for the photographic scene.

Keywords:

Object of interest, object detection, camera, image-capture device, camera application, image sensor, electronic imaging system, RGB photograph, adaptive zoom, aspect ratio, field of view, machine-learned model, convolution neural network (CNN), saliency model, saliency region, image saliency, pixel data, smartphone

Background:

To capture an image using an image-capture device, including digital cameras, smartphones, laptops, and so on, a user must power on the device and initiate an image sensor (e.g., camera) boot-up sequence. A user may initiate the boot-up sequence by selecting an application or simply turning on the device. The boot-up sequence typically involves an iterative optical and software-settings-adjustment process (e.g., automatic focus, automatic exposure, automatic white balance). After the booting sequence is complete, the image-capture device can then capture a quality image. Seldom, however, does the captured image contain an optimally zoomed presentation of the photographic scene. For example, a user may desire that the captured image contain a zoomed-in representation of the photographic scene, focusing on an object of interest.

Current techniques to adjust the zoom of an image sensor involve the users providing manual input (e.g., adjusting a knob) or performing a gesture (e.g., a pinch gesture, double-tap the screen). Oftentimes, manually adjusting the zoom requires time and effort by the user. Ideally, an image-capture device would automatically achieve a suitable zoom for a photographic scene.

Description:

This publication describes techniques for image-capture devices to automatically zoom to the object(s) of interest in a photographic scene. In aspects, through utilization of a machine-learned technique, an Automatic Zoom Manager detects a visual saliency region within an image, generates a bounding box enclosing the visual saliency region, and calculates an appropriate zoom ratio. By so doing, the Automatic Zoom Manager can implement the calculated zoom ratio to automatically achieve a suitable zoom for the photographic scene.

While the example image-capture device described in this publication is a smartphone, other types of image-capture devices, including tablets, laptops, and digital cameras, may also support the techniques described herein. An image-capture device may include one or more processors having logic for executing instructions, at least one built-in or peripheral image sensor (e.g., a camera), and an input/output device for displaying a user interface (e.g., a display panel).

The image-capture device further includes a computer-readable medium (CRM). The CRM may include any suitable memory or storage device like random-access memory (RAM), static RAM (SRAM), dynamic RAM (DRAM), non-volatile RAM (NVRAM), read-only memory (ROM), or flash memory. The image-capture device stores device data (e.g., user data, multimedia data, applications, and/or an operating system of the device) on the CRM. The device data may include executable instructions of an Automatic Zoom Manager. The Automatic Zoom Manager may be part of an operating system executing on the image-capture device, or may be a separate component executing within an application environment (e.g., a camera application) or a "framework" provided by the operating system.

The Automatic Zoom Manager implements a machine-learned technique ("Visual Saliency Model"). The Visual Saliency Model may be implemented as one or more of a support vector machine (SVM), a recurrent neural network (RNN), a convolutional neural network (CNN), a dense neural network (DNN), one or more heuristics, other machine-learning techniques, a combination thereof, and so forth. The Visual Saliency Model is iteratively trained, off-device, by exposure to training scenes, sequences, and/or events. For example, training may involve exposing the Visual Saliency Model to images (e.g., digital photographs), including user-drawn bounding boxes containing a visual saliency region (e.g., a region wherein one or more objects of particular interest to a user may reside). Exposure to images including user-drawn bounding boxes trains the Visual Saliency Model to identify visual saliency regions within images. As a result of the training, the Visual Saliency Model can generate a visual saliency heatmap for a given image and produce a bounding box enclosing the region with the greatest probability of visual saliency. In this way, the Visual Saliency Model can predict visual saliency regions within images. After sufficient training, model compression using distillation can be implemented on the Visual Saliency Model, enabling the selection of an optimal model architecture based on model latency and power consumption. The Visual Saliency Model can then be deployed to the CRM of the image-capture device as an independent module or implemented into the Automatic Zoom Manager.

The image-capture device performs operations under the direction of the Automatic Zoom Manager. Either automatically or in response to a received triggering signal, including, for example, a user performed gesture (e.g., tapping, pressing) enacted on the input/output device, the Automatic Zoom Manager implements several steps that calibrate the image-capture device. In a first step, the Automatic Zoom Manager receives one or more captured images from the image sensor.

In a second step, the Automatic Zoom Manager utilizes the Visual Saliency Model to generate a visual saliency heatmap using the one or more captured images. Figure 1A and 1B, below, illustrate a captured image and a visual saliency heatmap, respectively.

Figure 1A

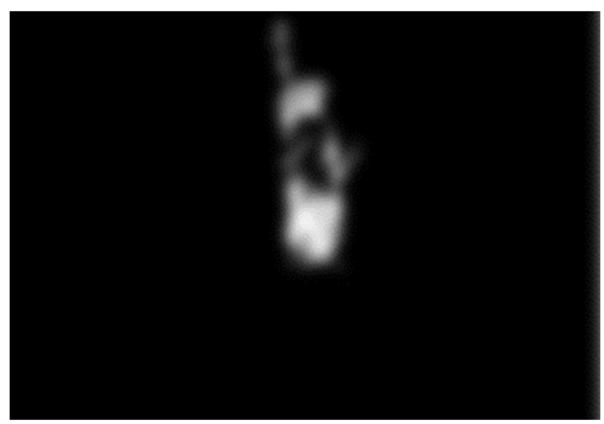

Figure 1B

Figure 1A illustrates a photographic scene captured by an image sensor. The Automatic Zoom Manager utilizes the Visual Saliency Model to generate a visual saliency heatmap of the captured image, as illustrated in Figure 1B. One or more processors calculate the visual saliency heatmap in the background operations of the device. The image-capture device does not display the visual saliency heatmap to the user. As illustrated, the visual saliency heatmap depicts the magnitude of the visual saliency probability on a scale from black to white, where white indicates a high probability of saliency and black indicates a low probability of saliency.

In a third step, the Visual Saliency Model produces a bounding box enclosing the region with the greatest probability of visual saliency. Figure 2 below illustrates the visual saliency heatmap from Figure 1B, along with a bounding box.

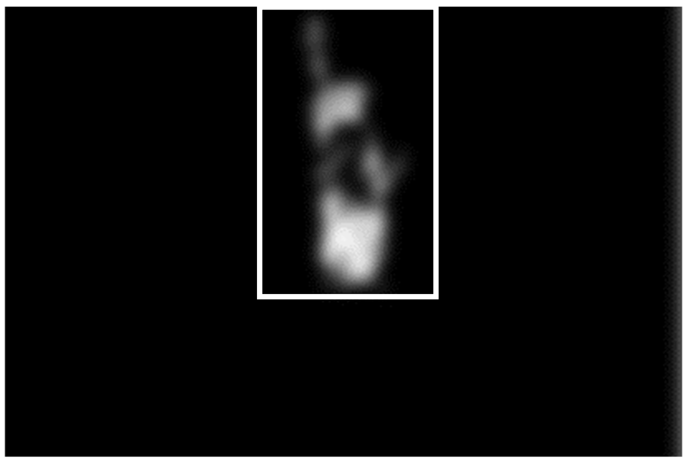

Figure 2

As illustrated in Figure 2, the visual saliency heatmap includes a bounding box enclosing the region within the image containing the greatest probability of visual saliency. In the event that there are multiple objects of interest in a photographic scene, causing the Visual Saliency Model to identify multiple saliency regions within a captured image, the Visual Saliency Model can be trained to produce a bounding box enclosing the saliency region nearest the center of the captured image. This trained technique assumes that a user is interested in the most centralized object in the image. Alternatively, the Visual Saliency Model can be trained to produce a bounding box enclosing all the objects of interest in a captured image.

In a fourth step, using Equation 1, the Automatic Zoom Manager calculates a targeted zoom ratio based on the bounding box dimensions:

$$zmRatio = max(boundingBoxWidth/imageWidth,\ boundingBoxHeight/imageHeight) \quad (1)$$

Equation 1 enables the Automatic Zoom Manager to calculate a zoom ratio (zmRatio) based on the bounding box width (boundingBoxWidth) and the image width (imageWidth), as well as the bounding box height (boundingBoxHeight) and the image height (imageHeight). In a final step, the Automatic Zoom Manager utilizes the zoom ratio value to adjust the zoom settings of the image-capture device.

Adjusting the zoom settings may involve the Automatic Zoom Manager directing one or more processors to adjust the arrangement of the optical lenses of an image sensor (i.e., optical zoom). In another aspect, the Automatic Zoom Manager utilizes a different image sensor to implement the calculated zoom ratio. In yet another aspect, the Automatic Zoom Manager digitally edits and enhances the image. For example, the Automatic Zoom Manager may crop and scale up the image, as well as add pixels (i.e., digital zoom). A combination of these aspects (e.g., hybrid zoom) may be utilized to achieve a suitable zoom, as well.

Figure 3 below illustrates a zoomed-in image of Figure 1A based on the bounding box as illustrated in Figure 2.

Figure 3

As illustrated in Figure 3, the Visual Saliency Model identified the object of interest within the image illustrated in Figure 1A, and the Automatic Zoom Manager implemented the calculated zoom ratio. Figure 3 illustrates an automatically zoomed-in image of Figure 1A.

In an aspect, the Visual Saliency Model identifies visual saliency regions and creates bounding boxes only when the center of the visual saliency region is within a predetermined distance of the center of the image. When this condition is met, the Automatic Zoom Manager zooms in on the center of an image. In another aspect, the Automatic Zoom Manager enables the user to zoom into arbitrary regions, where the zoom center is the center of the saliency region, not the center of the image.

In addition to the above descriptions, when an image-capture device displays a captured image on the input/output device, the Automatic Zoom Manager may display a bounding box enclosing the region within the image containing the greatest probability of visual saliency. In this way, the user can visualize the suggested automatic zoom before the device implements the zoom ratio. Figure 4 illustrates an example in which a captured image includes a bounding box enclosing the visual saliency region.

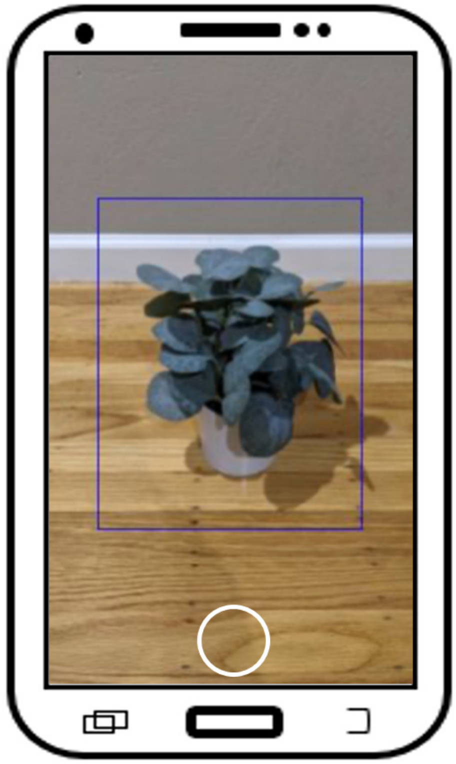

Figure 4

As illustrated, in Figure 4, the image-capture device is a smartphone. The smartphone displays a captured image to a user on the input/output device. Operating in the background of the device, the Automatic Zoom Manager utilizes the Visual Saliency Model to generate a visual saliency heatmap, identify the visual saliency region, and produce a bounding box enclosing the visual saliency region. As illustrated in Figure 4, the Automatic Zoom Manager presents the bounding box in the capture image enclosing the object of interest. The bounding box is presented to the user as a suggested automatic zoom.

Responsive to a user gesture, for example, including selecting the bounding box or tapping the input-output device, the Automatic Zoom Manager implements the suggested automatic zoom. Figure 5 illustrates the smartphone and captured image of Figure 4 with the suggested automatic zoom implemented.

Figure 5

As illustrated in Figure 5, the smartphone displays a zoomed-in version of the captured image from Figure 4. In this manner, the Automatic Zoom Manager, through the utilization of the visual saliency model, can detect a visual saliency region within an image, generate a bounding box enclosing the visual saliency region, calculate an appropriate zoom ratio, and implement the calculated zoom ratio to automatically achieve a suitable zoom for the photographic scene.

References:

[1] Patent Publication: US20190132520A1. Generating Image Previews based on Capture Information. Priority Date: November 2, 2017.

[2] Patent Publication: US20110267499A1. Method, Apparatus and System for Performing a Zoom Operation. Priority Date: April 30, 2010.

[3] Patent Publication: US20160117798A1. Image Zooming. Priority Date: October 27, 2014.

[4] Patent Publication: US20180077210A1. Method and Apparatus for Controlled Observation Point and Orientation Selection Audiovisual Content. Priority Date: September 9, 2016.