January 15, 2019

Determining region-of-interest based on motion patterns

Ruiduo Yang

Determining region-of-interest based on motion patterns

ABSTRACT

An auto-focus camera determines and focuses on a region-of-interest (RoI), e.g., faces, plates of food, flowers, etc. There is a slight lag, of the order of milliseconds, between the determination of the RoI and the capture of an image. During the time period of the lag, the subject of the photograph can sometimes move out of the just-determined RoI which can cause an off-target focus. This disclosure describes techniques to predict the RoI for a frame based on the RoIs of immediately-previous frames, the camera state, machine-learned patterns of motion, etc. The techniques enable robust and accurate auto-focus on the subject during capture of the photograph.

KEYWORDS

automatic focus; region of interest; RoI; camera; machine learning; motion tracking

BACKGROUND

An auto-focus camera determines and focuses on a region-of-interest (RoI), e.g., faces, plates of food, flowers, etc. An auto-focus camera typically determines and focuses on a region-of-interest (RoI). Accurate determination of RoI improves photo quality and narrows down search regions for important objects. RoI can be determined, e.g., by the touch of a user on a touchscreen, by techniques of image analysis, etc.

To focus on a region-of-interest, the following steps are performed: a frame (f0) is captured with the camera and the RoI of the captured frame is determined; next, the determined RoI is provided to the focusing hardware. The focus for a subsequent frame (f1) is determined based on the received RoI.

There is a slight lag, of the order of milliseconds, between the determination of the RoI and the capture of an image. During the time period of the lag, the subject of the photograph can sometimes move out of the just-determined RoI which can cause an off-target focus. For example, when frame f0 is received and used to determine focus for capture of a subsequent frame, several milliseconds may have passed. As another example, when the focus is set based on the determined RoI during capture of the frame f1, the subject of the photograph may have moved out of the RoI due to natural movement within the scene, due to computational latency, delays of the focusing hardware, due to shaking of the hands of the user holding the camera, etc.

DESCRIPTION

This disclosure describes techniques to predict the RoI for a frame based on the RoIs of immediately-previous frames, the camera state, machine-learned patterns of motion, etc. The techniques enable robust and accurate auto-focus on the subject during capture of the photograph.

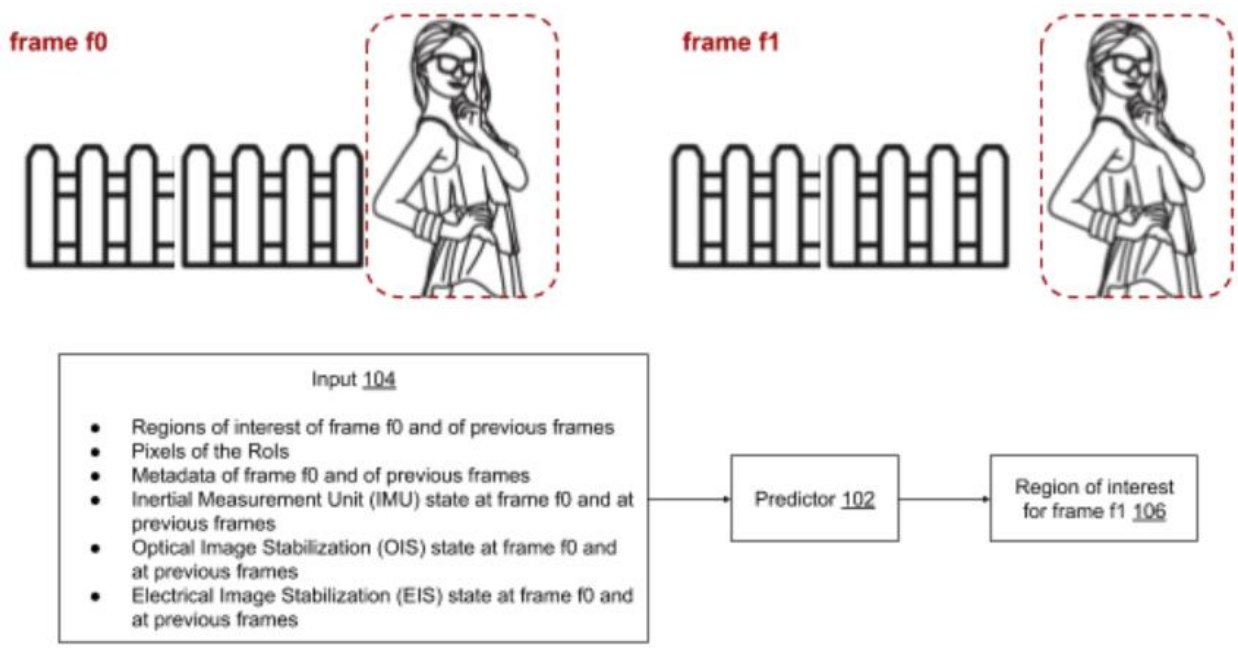

Fig. 1: Predicting the region of interest of a frame

As illustrated in Fig. 1, a predictor (102) accepts as input (104) factors such as:

- the regions of interest of frame f0, and of previous frames extending to some window-length W;
- the pixels of the regions of interest;
- metadata of the frame f0, and of previous frames within the window W;
- the states of the inertial measurement unit at frame f0 and at previous frames within the window W;
- the states of the optical image stabilizer at frame f0 and at previous frames within the window W;
- the states of the electrical image stabilizer at frame f0 and at previous frames within the window W, if electrical image stabilization happens before RoI generation; etc.

The predictor produces a predicted region of interest for frame f1 (106). In addition, the predictor accepts as feedback the true RoI for frame f1. The true RoI is used to perform regression analysis and reduce prediction error. The true RoI can be obtained, e.g., in a training phase. In this example, the predictor is trained using offline learning. If contrast-based, e.g., scanning, auto-focus is used, the RoI estimated for f1 is used. If direct distance-estimating technology, e.g., phase-based or dual-pixel based technology, is used, the distance change from f0 to f1 is estimated and added to the distance difference. The distance difference can be predicted from a machine learning model, or can be calculated based on the RoI size change, etc.

The predictor can be a machine learning model, e.g., a generative machine learning model, a regression learning model, a neural network, etc. Example types of neural networks that can be used include long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other models, e.g., support vector machines, random forests, boosted decision trees, etc., can also be used. Techniques such as Hidden Markov Models (HMM), conditional random field (CRF) can also be used.

Under online learning, RoI is predicted in a manner similar to the illustration of Fig. 1. In addition, other inputs, e.g., camera location and time as reported by GPS, etc., are used as permitted by the user, to predict RoI. Training advantageously occurs on the device that includes the camera, e.g., smartphone, tablet, personal computer, etc. Training can also occur within the cloud. The obtained training data (frames captured with the device camera) is used to fine-tune the model and is subsequently removed. Online learning can be used to tailor the learning model for individual users e.g., to a particular user’s hand-shake pattern, captured subjects, etc.

As shown in Fig. 1, the region of interest (in dotted-lines) is different in frame f0 from subsequent frame f1, as the subject has moved between the capture of the two frames. With the techniques described herein, it is ensured the subject remains in focus during capture.

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 the user 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 to predict the RoI for a frame based on the RoIs of immediately-previous frames, the camera state, machine-learned patterns of motion, etc. The techniques enable robust and accurate auto-focus on the subject during capture of the photograph.

REFERENCES

1. Guan, Haike and Norikatsu Ninami, "Object Tracking and Motion Prediction Technology for Digital Camera." *Ricoh technical report* 41(2015):19-30, available online at https://jp.ricoh.com/technology/techreport/41/pdf/RTR41a02.pdf