Pattern Recognition Letters 56 (2015) 30–37

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Improving the accuracy and low-light performance of contrast-based autofocus using supervised machine learning ☆

Rudi Chen, Peter van Beek*

University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada

A R T I C L E I N F O

Article history:
Received 21 May 2014
Available online 3 February 2015

Keywords:
Autofocusing
Digital cameras
Supervised machine learning
Decision trees

A B S T R A C T

The passive autofocus mechanism is an essential feature of modern digital cameras and needs to be highly accurate to obtain quality images. In this paper, we address the problem of finding a lens position where the image is in focus. We show that supervised machine learning techniques can be used to construct heuristics for a hill-climbing approach for finding such positions that out-performs previously proposed approaches in accuracy and robustly handles scenes with multiple objects at different focus distances and low-light situations. We gather a suite of 32 benchmarks representative of common photography situations and label them in an automated manner. A decision tree learning algorithm is used to induce heuristics from the data and the heuristics are then integrated into a control algorithm. Our experimental evaluation shows improved accuracy over previous work from 91.5% to 98.5% in regular settings and from 70.3% to 94.0% in low-light.

1. Introduction

The passive autofocus mechanism is an essential feature of modern digital cameras. To maximize the likelihood of obtaining high quality images, the mechanism must be quick and accurate. The two main forms of passive autofocus are phase-detection and contrast-detection. Phase-detection is faster and more efficient at tracking subject movement, but typically requires hardware found on higher-end cameras. Contrast-detection can be more accurate and uses image processing techniques to obtain a measure of the sharpness of an image that can be used on a wide range of devices, including point-and-shoot cameras, mobile phones and DSLRs. This paper focuses on contrast-detection.

We present a novel algorithm for finding an in-focus lens position through a local search, in contrast to sweeping through the entire range of lens positions. We build a control algorithm supported by supervised machine learning that is used to train a classifier to transition between the states of the algorithm. In supervised learning, classifiers are built using training examples (instances) consisting of a vector of feature values and labeled with the correct answer. We obtain training and test data using an offline simulation on a suite of 32 benchmarks with each instance labeled in an automated manner. From the gathered data, a decision tree learning algorithm [15] was used to induce multiple heuristics. In a decision tree, the internal nodes of the tree are labeled with features, the edges to the children of a node are labeled with the possible values of the feature, and the leaves of the tree are labeled with a classification. To classify a new example, one starts at the root and repeatedly tests the feature at a node and follows the appropriate branch until a leaf is reached. The label of the leaf is the predicted classification of the new instance.

The final result is compared with previous work by performing an extensive evaluation over a range of real-life photography situations. Our approach is shown to be more accurate, including in low-light scenarios and situations where the focus measure is not unimodal.

2. Background

In this section, we review the necessary background in auto-focus, focus measures, and focus search algorithms.

2.1. Focus measures

Contrast-detection autofocus (AF) makes use of a focus measure that maps an image to a value that represents the degree of focus of the image. Many focus measures have been proposed and evaluated in the literature see, for example, [8,13]. In our work, we make use of two focus measures: (i) the squared gradient focus measure [17] and (ii) a focus measure based on the convolution of the image with the derivatives of the Gaussian [5,7]. Let $f(x, y)$ be the luminance or grayscale at pixel (x, y) in an image of size $M \times N$. The value $\phi(p)$ of the squared gradient focus measure for an image acquired when the

☆ This paper has been recommended for acceptance by Egon L. van den Broek.

* Corresponding author: Tel.: +1 519 888 4567, x35344; fax: +1 519 885 1208.
E-mail address: vanbeek@cs.uwaterloo.ca (P. van Beek).

http://dx.doi.org/10.1016/j.patrec.2015.01.010
0167-8655/

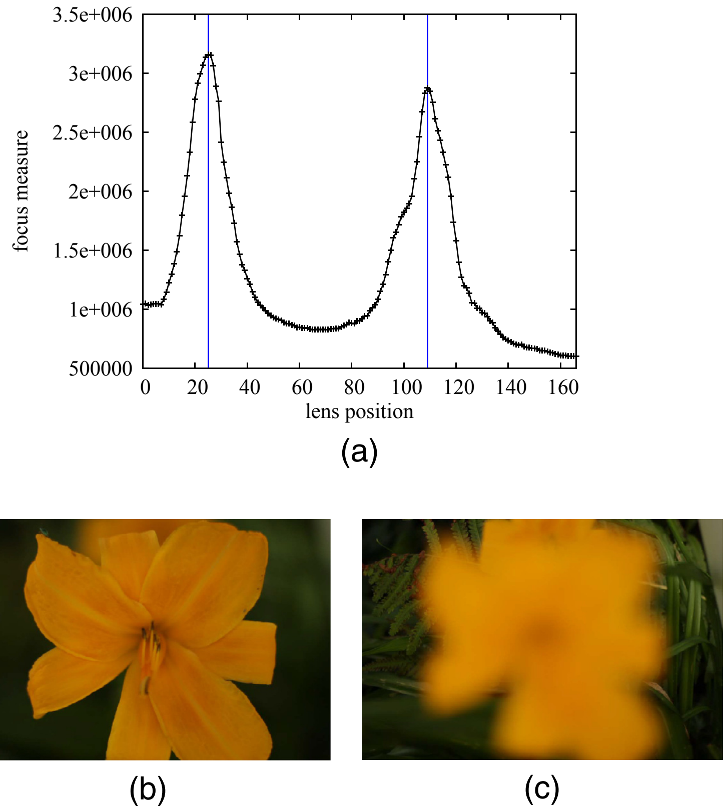

Fig. 1. (a) Focus measures of images at each of the 167 lens positions (Canon 50 mm lens) for an example scene using the squared gradient focus measure. The two (blue) vertical bars refer to the two images that have objects that are in maximal focus: (b) flower in focus, and (c) fern and grasses in focus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lens is at position p is then given by,

$$\phi(p) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-2} (f(x, y+1) - f(x, y))^2.$$

The value $\phi(p)$ of the focus measure based on convolution with the first derivatives of the Gaussian is given by,

$$\phi(p) = \sum_{x} \sum_{y} [(f * G_x^\sigma)(x, y)]^2 + [(f * G_y^\sigma)(x, y)]^2,$$

where f is the image, G_x^σ and G_y^σ are the first-order derivatives of the Gaussian in the vertical and horizontal direction, σ is the scale of the filter (higher values cut off more of the high frequencies of the image), and * is the 2D convolution operator. Both of these focus measures have been shown to be highly effective across many types of images [13] and the Gaussian focus measure has excellent noise reduction properties for low-light situations [5].

Following Kehtarnavaz and Oh [11], we assume that the region of interest (ROI) is the entire image. In practice, a user can either (i) specify the ROI by moving a rectangle over the desired part of the image when the camera is in live previous move, or (ii) have the camera automatically determine the object or region of interest to bring into focus, say by using face or object recognition [16]. Our proposals are easily adapted to the case where the ROI is an arbitrary sub-area of an image. Fig. 1 shows the focus measures acquired at all possible lens positions (Canon 50 mm lens) in a sample scene.

2.2. *Focus search algorithms*

A contrast-based AF algorithm searches for the lens position with the sharpest image as reported by the focus measure by iteratively moving the lens. Lenses are moved in discrete steps using a step motor. The images are streamed from the sensor at video frame rates (e.g., 24 frames per second on many Canon cameras) and also shown on the camera's live preview mode.

At each iteration, step motors can be moved in a single step or larger steps. In our case, the largest step is equivalent to eight small steps. Each step, small or large, is followed by a latency of hundreds of milliseconds. As well, step motors can suffer from backlash when the lens movement changes direction, making any absolute measure of lens position unreliable [11,14]. In our work, we assume that the only information available to the camera in regards to lens position is whether the lens has reached the first or last lens position. An efficient AF algorithm will therefore take as large steps as possible and be able to handle losses in accuracy due to backlash.

Given a set of lens positions $\{a, a+1, \ldots, b\}$, an autofocus algorithm can solve one of the several search problems. The first is to find the lens position that corresponds to the *maximum* or *highest* peak. This often involves finding *all* peaks. Another problem is to find a *nearby* peak. During real usage, the lens could be resting at any lens position prior to autofocus activation. In fact, the lens is likely to be near a peak already when images are taken consecutively. In that case, it is preferable to start the search from the current position rather than using an approach that requires moving the lens back to the first lens position. For lenses with a large number of positions, such as those used with DSLRs, moving the lens back to the first lens position would also incur a significant visual artifact in the live preview where the image goes significantly out of focus before coming back to focus again. This would not be desirable for the user. In this paper, we address the problem of finding a nearby peak without moving the lens back to the first lens position.