Pattern Recognition 48 (2015) 3310–3323

Contents lists available at ScienceDirect

Pattern Recognition

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

A focus fusion framework with anisotropic depth map smoothing

Madina Boshtayeva *, David Hafner, Joachim Weickert

Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Campus E1.7, Saarland University, 66041 Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 3 February 2014
Received in revised form
23 September 2014
Accepted 6 October 2014
Available online 18 October 2014

Keywords:
Focus fusion
In-focus measure
Depth map
Anisotropic smoothness

ABSTRACT

Focus fusion is the task of combining a set of images focused at different depths into a single image that is entirely in-focus. The crucial point of all focus fusion methods is the decision about the in-focus areas. To this end, we present a general framework for focus fusion that introduces a modern regularisation strategy on these per-pixel decisions. We assume that neighbouring pixels in the fused image belong to similar depth layers. Following this assumption, we smooth the depth map with a sophisticated anisotropic diffusion process combined with a robust data fidelity term. The experiments with synthetic and real-world data demonstrate that our new model yields a better quality than several existing focus fusion methods. Moreover, our methodology is general and can be applied to improve many fusion approaches.

1. Introduction

In applications such as macro-photography or optical microscopy, the limited depth of field of standard cameras poses a severe problem: it is not possible to capture an image that is totally in focus. A common remedy is to acquire a set of images while varying the position of the focal plane. In this way, the image stack contains all required information to produce a single image that is sharp everywhere. The task of combining these images of the focus stack into an all-in-focus composite is called *focus fusion*.

1.1. Related work

We categorise focus fusion techniques into two main groups: the methods in the first group work on multiscale decompositions of the images. In the first step, they apply a multiscale transformation of the complete image stack. Next all images are combined in the transform domain by selecting the coefficients that have the highest probability of belonging to in-focus areas. Finally the composed multiresolution representation is transformed back to the spatial domain. The result is the all-in-focus image. In this class, the pioneering work is the Gaussian and Laplacian pyramid-based method by Ogden et al. [1]. Later Burt and Kolczynski [2] generalised it to alternative pyramid representations, and Petrovic and Xydeas [3] proposed a multiresolution gradient map representation. Also wavelet-based methods belong to this class of algorithms. Here a first approach with application to focus fusion was published by Li et al. [4]. Modifications and extensions have, for instance, been proposed by Forster et al. [5] or Lewis et al. [6]. In [7], Zhang and Blum present a generic framework for multiscale image fusion and compare different approaches. All these multiresolution-based techniques share the same constitutional drawback: performing the fusion in the transform domain may change the intensity values and create artificial colours. This produces undesirable artefacts in the fused result.

To overcome this drawback, the algorithms of the second group work in the image domain. Here the basic idea is first to select the regions from all frames that are in focus, and then to combine them to one composite. Recently, many methods for focus fusion have been reported in the literature which employ machine learning techniques to build a sharp image: Wu et al. [8] propose a method using a hidden Markov model, Wan et al. [9] employ principal component analysis for the focus fusion task, and Wang et al. [10] use pulse coupled neural networks to obtain a sharp composite. All of these models work well in the image domain. However, in general, operating in the image domain can cause unpleasant visible seams that appear when simply arranging the identified in-focus areas in a mosaic-like fashion. To tackle these artefacts, Pop et al.

Table 1
Overview of applied in-focus measures to compute the initial depth map d^{init}.

Measure	Formula
Gradient magnitude	$m_1 = \lvert \nabla f_\sigma \rvert$
Norm of the Laplacian	$m_2 = \lvert \Delta f_\sigma \rvert$
Frobenius norm of the Hessian	$m_3 = \lVert \mathcal{H} f_\sigma \rVert_F$
Trace of the structure tensor	$m_4 = \operatorname{tr}(\boldsymbol{J}_\rho(\nabla f_\sigma))$
Determinant of the structure tensor	$m_5 = \det(\boldsymbol{J}_\rho(\nabla f_\sigma))$
Variance	$m_6 = \frac{1}{\lvert \mathcal{N}(\boldsymbol{x}) \rvert} \int_{\mathcal{N}(\boldsymbol{x})} \left(f_\sigma(\tilde{\boldsymbol{x}}) - \mu(\boldsymbol{x}) \right)^2 \, d\tilde{\boldsymbol{x}}$

* Corresponding author. Tel.: +49 681 302 57353; fax: +49 681 302 57342.
E-mail addresses: boshtayeva@mia.uni-saarland.de (M. Boshtayeva), hafner@mia.uni-saarland.de (D. Hafner), weickert@mia.uni-saarland.de (J. Weickert).

http://dx.doi.org/10.1016/j.patcog.2014.10.008
0031-3203/

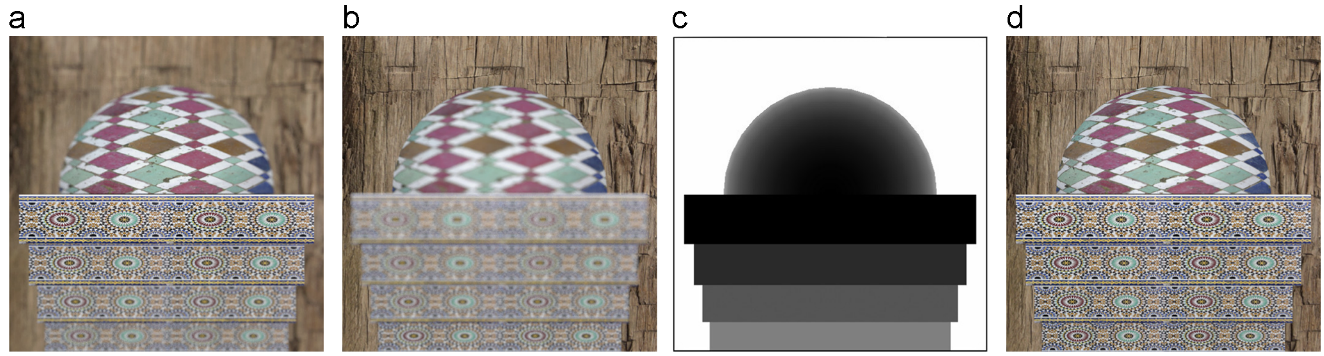

Fig. 1. Synthetic data set. (From left to right) (a) Frame 1 with the shortest focal distance. (b) Frame 13 with the largest focal distance. (c) Ground truth depth map: brighter grey tones describe larger depth values. (d) ground truth image (all-in-focus).

[11] as well as Wang et al. [12] explicitly model a smoothness constraint of the resulting composite image by means of a partial differential equation (PDE). Unfortunately this may also cause smoothing of important image structures such that the resulting images appear blurred and not sharp everywhere.

Hence, researchers came up with the idea of not applying the smoothness constraint on the resulting image itself, but on the per-pixel decision of the in-focus areas: in [13–18] the authors determine an initial decision map by means of a specific sharpness criterion. Subsequently they segment these maps into regions that belong to the same input frames. These segments are then used to fuse the input images to an all-in-focus composite, or even to recover an underlying 3-D surface. Agarwala et al. [13] use graph-cut optimisation to segment different in-focus areas and fuse the input images in the gradient domain. Šroubek et al. [14] propose a level-set segmentation on the decision map solving a suitable PDE. Similarly, in the method of Li and Yang [15] the images are segmented with the normalized cut method; this method is further extended in the work of Liu et al. [19].

There are many other approaches that offer improvements to focus fusion techniques and algorithms: Muhammad and Choi [20] derive the optimal sampling to obtain a reasonable 3-D shape. In [16], Shim and Choi introduce a novel iterative algorithm to reconstruct the 3-D shape. Mahmood et al. [17] propose a combination of different focus measures for constructing the optimal decision map through genetic programming. Mahmood and Choi [18] employ 3-D anisotropic diffusion to enhance the input images, and in turn, to obtain an accurate decision map. Staying with the idea of operating on decision maps, Bae et al. [21] apply bilateral filtering to this decision map in a related context. However, they do not consider the focus fusion task, but perform a defocus magnification given a single image. While most of the related research aims for piecewise constant solutions, we aim to achieve more realistic piecewise smooth decision maps to the focus fusion problem.

1.2. Contributions

In our work, we follow the idea of processing an initial decision map. However, instead of a segmentation-based approach, we introduce a modern regularisation technique which aims to smooth the initial decision map. Moreover, since each image is sharp at a particular depth value, we interpret the decision maps as depth maps. Consequently, we aim for *piecewise smooth* solutions as opposed to *piecewise constant* ones that are obtained by segmentation-based methods. In this way, we are even able to adequately handle pixels that are never captured totally in-focus since they lie between two focal planes. The explicit modelling of smooth transitions in depth not only provides more accurate depth maps, but also counteracts unpleasant seams in the final image.

In our approach, we formulate a similarity to a precomputed depth map or even to a composite of multiple depth maps by a robust data term and combine it with a modern adaptive regularisation technique: our *joint image- and depth-driven* diffusion is guided by the structures of the evolving all-in-focus image, while the amount of smoothing is determined by the depth map gradients.

Compared to our conference publication [22] these are the following new contributions in the present paper:

(i) In [22], we applied the gradient magnitude as indicator of sharp image regions. However, our method is very general and not limited to this specific choice: it creates a high quality depth map using one or multiple depth maps that can been precomputed with various sharpness measures. Our experiments demonstrate this by means of six measures.

(ii) In the conference paper, we computed the solution of our model via gradient descent, i.e. as the steady state ($t \to \infty$) of a parabolic PDE. Here we present an alternative elliptic formulation and solve the resulting system of equations with a modern well-parallelisable algorithm implemented on GPU. In this way, we reduce the runtime of our approach significantly.

(iii) Last but not least, we conduct a thorough evaluation of our method on synthetic and various real-world focus stacks. We show the performance of the proposed nonlinear anisotropic diffusion in comparison to the linear isotropic one. We demonstrate the flexibility and general applicability of our technique, and compare the results with several focus fusion methods from the literature.

1.3. Organisation

Our paper is organised as follows: in Section 2, we introduce our diffusion-based approach and explain its algorithmic realisation in full detail. Section 3 illustrates the performance of our method on synthetic as well as on real-world experiments. This includes a comparison to several focus fusion approaches. Finally, we conclude the paper with a summary and an outlook in Section 4.

2. Our focus fusion framework

Let $f(\boldsymbol{x}, z)$ be a 3D volume where $\boldsymbol{x} := (x, y)^\top$ denotes the location within a rectangular image domain $\Omega \subset \mathbb{R}^2$ and $z \in \mathbb{R}$ the depth. We interpret the K input images $f(\boldsymbol{x}, z_k)$ with $k = 1, \ldots, K$ as equidistant slices of this volume.[1] Our goal is to find a depth map $d(\boldsymbol{x})$ that selects for each location $\boldsymbol{x}$ the frame that is in focus. To this end, our focus fusion framework consists of three main parts: in the first step, we

[1] The required number of depth samples is scene dependent, see Muhammad and Choi [20].