Computers & Graphics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

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

Automated outdoor depth-map generation and alignment☆

Q1 Martin Čadík[a,*], Daniel Sýkora[b], Sungkil Lee[c]

Q2 [a] CPhoto FIT, Faculty of Information Technology, Brno University of Technology, Czechia
[b] Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia
[c] College of Software, Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Article history:
Received 18 January 2018
Revised 23 March 2018
Accepted 1 May 2018
Available online xxx

Keywords:
Image enhancement
Synthetic depth
3D terrain
Free-form warping
Image registration
Synthetic camera

ABSTRACT

Image enhancement tasks can highly benefit from depth information, but the direct estimation of outdoor depth maps is difficult due to vast object distances. This paper presents a fully automatic framework for model-based generation of outdoor depth maps and its applications to image enhancements. We leverage 3D terrain models and camera pose estimation techniques to render approximate depth maps without resorting to manual alignment. Potential local misalignments, resulting from insufficient model details and rough registrations, are eliminated with our novel free-form warping. We first align synthetic depth edges with photo edges using the as-rigid-as-possible image registration and further refine the shape of the edges using the tight trimap-based alpha matting. The resulting synthetic depth maps are accurate, calibrated in the absolute distance. We demonstrate their benefit in image enhancement techniques including reblurring, depth-of-field simulation, haze removal, and guided texture synthesis.

1. Introduction

A limited configuration in taking photographs does not always lead to the highest quality, and often motivates enhancement of photographs. Computational photography has addressed such limitations, which introduces additional flexibility on focus, exposure, and depth [1–3]. Among them, depth information, on which we focus here, can greatly facilitate diverse image manipulations, such as refocusing, dehazing, texture synthesis, and image editing [4–7].

Outdoor photographs (e.g., natural landscapes) represent by far the biggest group in many media services [8], but their direct depth estimation poses a challenge. They are usually monocular, which has a low chance to work with typical structure-from-motion. Within-image features, such as airlights or textures [9,10], help, but are not always available. Range sensors [11] are applicable only to limited distance ranges (only up to tens of meters).

One better alternative can be an indirect estimation from a 3D terrain model, which renders the reference depth map as previously suggested by Kopf et al. [12]. The terrain model is already available for the whole planet and recent photographs are usually geo-tagged (e.g., the global positioning system; GPS), which can serve as a strong external cue. Further, this approach can be distinguished for its higher resolution and accuracy; the direct estimation may yield a coarser resolution or wrong outcomes.

However, the 3D terrain model may be insufficient in its resolution and details (e.g., textures and objects). Also, a precise registration between real and virtual views is challenging, where the alignment errors result in visible artifacts in edited images. Manual registration starting at a rough initial guess can help [13], but is laborious and inappropriate for massive batch processing.

In this work, we present a *fully automatic* framework for depth-map generation and alignment for an outdoor photograph. A virtual camera is first localized with the geo-tagging information of a photo and recent camera pose estimation techniques. Then, the 3D terrain model is rendered at the virtual camera to produce an initial approximate depth map. Local inaccuracies, resulting from the rough registration and insufficient details of the model, are subsequently reduced using our novel *automatic* free-form warping. We first align discontinuities in the synthetic depths with photo edges using the as-rigid-as-possible image registration. The shape of the edges is further refined using the tight trimap-based alpha matting. The resulting depth map, synthesized from the geo-referenced terrain model, is *absolute* (calibrated in meters). We show its benefit in image enhancement tasks, including refocusing, defocus manipulation (see Fig. 1), dehazing, and texture synthesis.

☆ This article was recommended for publication by Dr M Kim.

* Corresponding author.
E-mail addresses: cadik@fit.vutbr.cz, cadikm@centrum.cz (M. Čadík), sungkil@skku.edu (S. Lee).

https://doi.org/10.1016/j.cag.2018.05.001
0097-8493/

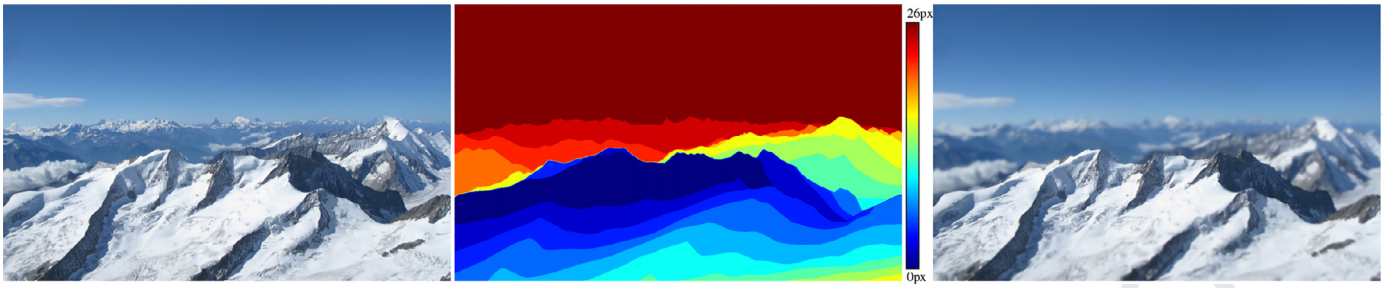

Fig. 1. Transforming an outdoor photograph into a model-like look. An automatically generated synthetic depth map is used to calculate plausible blur kernel size map (middle) to simulate shallow depth-of-field (right) in landscape images (left), where such an effect cannot be achieved using standard optics for physical reasons. Virtual camera: full-frame, f-number=1.0, focal length=1200 mm, focus distance=5 km.

2. Related work

We briefly review previous work on depth map reconstruction and its major applications including defocus manipulation and dehazing.

2.1. Depth map reconstruction

Robust depth map reconstruction is an ongoing subject of interest. A typical approach is to rely on stereo image pairs [18,19] or multiple/multiview images [20–22]. More recently, short-distance range-sensing devices, such as Kinect [11], improved the availability of depth maps in indoor environments [23].

Multiple images are not easily available in practice, and single-image processing has also been intensively studied. Semi-automatic user interaction often helps [24–26], and further modification to hardware or light patterns proved its utility, such as light fields [27–31], coded aperture [32] or structured light method [33].

Depth maps can further be generated semi-automatically using sparse samples (seeds) provided by the user [34–37]. In these approaches, anisotropic diffusion is used to propagate depth information from seeds to the rest of the image. The key assumption here is that the gradient of the resulting depth map should roughly correspond to the color gradient.

The previous methods are not applicable to ours which is targeting on *single outdoor* photographs. The stereo vision techniques require multiple images, while the range sensors work only for a small range of distances. Computational photography requires special hardware or modification to the aperture. For the diffusion-based techniques, real photos often violate their main assumption about compatible depth and color gradients, and also, the positions of depth seeds require to be accurate. Otherwise, the diffusion will propagate small misalignments to a notably larger area and lead to notable artifacts.

Kopf et al. showed the combination of geo-tagging and 3D models can be used for fairly accurate geo-registration and many applications including dehazing and relighting [12]. The geo-tagging already allows us to select an effective subset for structure from motion [38], but when combining with the available 3D models, we can directly render a depth map. The combination even enables to assign geo-locations and labels onto pixels [39], point clouds [38,40], or annotate photos [41]. Nevertheless, the depth map is not pixel-perfect and requires fine alignment; we address this issue in the present work.

While the majority of previous approaches estimate *relative* depth maps, our solution can generate *absolute* depth map from the geo-referenced digital terrain model. This is highly beneficial in many image enhancement applications; for instance, there are more chances in estimating kernels for defocus blur or other effects.

Similarly to our goals, the method proposed by Kopf et al. [12] allows to synthesize absolute depth maps. However, it requires a *user-assisted* interaction for registration, which motivates for our novel depth free-form warping step (Section 3).

2.2. Defocus manipulation

The defocus blur, caused by shallow depth of field (DOF), is pronounced in indoor photos or films, but hardly exhibited in outdoor photography (even with large lenses). Its capture is inherently restricted to a particular configuration (e.g., focus and *f*-number). Thus, its post-reproduction for novel configurations (with computational photography) drew attention for refocusing [32,33] or defocus magnification [42].

Another mainstream is the postprocessing of a usual single-view image, which comprises deblur and refocusing. Typically, the deblur involves blind deconvolution using known priors to estimate kernels [24,43]. However, a precise solution requires to consider geometric visibilities similarly to the distributed ray tracing [44]. Our solution and its absolute depth information can facilitate the estimation of per-pixel local blur kernels to some extents, enabling non-blind deconvolution. Refocusing can also benefit from ours, which can use a precise rendering technique [45].

2.3. Dehazing

Outdoor photographs are often hazed by atmospheric scattering that can be characterized by medium *transmission maps*, depending heavily on depths. Most previous work has focused on depth estimation and radiance recovery, including Markov random fields (MRF) [46], independent component analysis [47], dark channel prior [48], factorial MRF [49–52], and machine-learning approaches using random forests [53,54], and convolutional neural networks [55]. In our case, geo-referenced absolute depth maps are used for dehazing, which can be more precise than the previous ones. We obtain such maps automatically without the previous manual image-to-model registration [12].

3. Automatic depth-map generation and alignment

In this section, we describe our fully-automatic approach to depth-map generation from a single color image as well as a technique used for the final depth map refinement (see Figs. 2 and 3 for summary).

Retrieval of 3D terrain model. A Google-Earth-like digital terrain models are currently available for the whole planet. They are acquired from satellites and/or planes and published in form of geo-referred digital elevation maps (DEMs) even for less accessible regions. Such models are sufficient for our purposes (i.e., *outdoor photographs*), however in general case, a 3D model might not be

Please cite this article as: M. Čadík et al., Automated outdoor depth-map generation and alignment, Computers & Graphics (2018), https://doi.org/10.1016/j.cag.2018.05.001