Information Sciences 376 (2017) 190–201

Contents lists available at ScienceDirect

Information Sciences

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

Motion-free exposure fusion based on inter-consistency and intra-consistency

Wei Zhang[a], Shengnan Hu[a,*], Kan Liu[a,b], Jian Yao[c]

[a] School of Control Science and Engineering, Shandong University, China
[b] Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, China
[c] School of Remote Sensing and Information Engineering, Wuhan University, China

ARTICLE INFO

Article history:
Received 19 April 2016
Revised 1 October 2016
Accepted 9 October 2016
Available online 15 October 2016

MSC:
00-01
99-00

Keywords:
Deghosting
Exposure fusion
Inter-consistency
Intra-consistency
Gradient

ABSTRACT

Exposure fusion often suffers from ghost artifacts, which are caused by the movement of objects when a dynamic scene is captured. In this paper, two types of consistency concepts are introduced for enforcing the guidance of a reference image for motion detection and ghost removal. Specifically, the inter-consistency, which represents the similarities of pixel intensities among different exposures, is weakened by the use of different exposure settings. Histogram matching is employed to recover the inter-consistency. Following this, pixel differences are mostly the result of changes in content caused by object movements, so motion can easily be detected. To further restrain the weights of outliers in fusion, motion detection is performed at a super-pixel level, to ensure that pixels with similar intensities and structures share similar fusion weights. This is referred to as intra-consistency. Experiments in various dynamic scenes demonstrate that the proposed algorithm can determine the motion more effectively than existing methods, and produce high quality fusion results that are free of ghost artifacts.

1. Introduction

Despite recent advances in imaging technology, common digital cameras often fail to capture the range of irradiance that is visible to the human eye in natural scenes. To address this limitation, much effort has been given to acquiring images representing the full dynamic range of the real world, referred to as high dynamic range (HDR) images. Specifically, because each exposure can be designed to capture a certain dynamic range, HDR imaging can be achieved by capturing a stack of differently exposed images of the same scene and then merging them. However, because the scenes we encounter in practice are often dynamic and contain moving objects, a simple composition of pictures may introduce ghosting or blurring artifacts. Therefore, the problem of removing ghosting artifacts in sequential HDR imaging has become an important topic of research.

Common approaches for generating HDR images of static scenes generally consist of two steps. First, the camera response function (CRF) [2,7] is estimated, and the radiance maps of the multiple exposures are combined to create a latent image that stores pixel values covering the entire tonal range of the real scene. Second, tone mapping is applied to re-map the obtained latent HDR image to an image of low dynamic range (LDR), in order to make it displayable on commonly used LDR devices [3,4,16,28]. Some other methods [19,21,26] skip the typical HDR process, and create a tone-mapped-like HDR image

* Corresponding author. Fax: +86 21 6564 2219.
E-mail address: sarahhu0109@gmail.com (S. Hu).

http://dx.doi.org/10.1016/j.ins.2016.10.020
0020-0255/

directly using image fusion [15,33]. Therefore, methods of this type are more efficient, and do not require tone mapping. However, because all of the above-mentioned approaches are only intended for completely static scenes, any movement during the exposure sequence will introduce ghosting artifacts into the resulting images.

1.1. Related work

There exists previous work addressing the removal of HDR ghosting artifacts in dynamic scenes. In general, previous methods can be categorized into two types, as follows.

The first type assumes that the images are mostly static, and only small parts of the scene contain motion. They align the images to each other by utilizing the input image sequence to determine whether a given pixel is static or corrupted by a moving object, and then combine the pixels without movement to obtain an artifact-free HDR image. For instance, Kang et al. [13] computed the optical flow between successive frames, and then warped pixels to detect which of them exhibit motion. Reinhard et al. [27] proposed a method based on radiance variation values, while Jacobs et al. [12] proposed a method based on entropy maps. Grosch [6] performed thresholding on the error map estimated from the input stack to detect motion. In addition, Khan et al. [14] and Pedone and Heikkilä [25] used the kernel density estimators to compute the probability that a pixel belongs to a moving object. One common problem of these algorithms is that they cannot handle scenes containing extensive motion or in which some portions change frequently, because they assume that there is a predominant pattern in the captured sequence for each location of the target scene. This can be referred as the 'majority assumption,' which implies that moving objects only appear in small parts of the scene, or appear in a small percentage of the exposures at a single location.

To account for this limitation, methods of the second type attempt to align the images in a stack to a selected reference image, before merging them into an HDR image. To avoid ghosting artifacts, additional details extracted from the other exposures will only be included if they are consistent with the scene defined by the reference image. Methods of this type can be referred as reference-based fusion methods. For example, some deghosting methods [1,8,10,11,29] have been proposed that find inconsistencies between the reference exposure and others by using patch matching. Specifically, HaCohen et al. [8] employed a new coarse-to-fine scheme, in which nearest-neighbor fields are computed based on the generalized PatchMatch technique [1] to determine the correspondences between two images. Hu et al. [10] employed the algorithm proposed in [8] to locate dense correspondences between the reference image and the others in the given image stack, and corrected wrong correspondences using local homographies. Sen et al. [29] developed a new patch-based optimization method that jointly solves for both the HDR image and the aligned images simultaneously. Hu et al. [11] employed a PatchMatch based algorithm to tackle ghosts in saturated sections of pictures. In addition, Zimmer et al. [38] proposed an energy-based optic flow approach to detect ghosts. Zhang and Cham [37] proposed a reference-guided method, using changes in gradient direction as a cue for motion detection. It has been proved that reference-based methods do not require the 'majority assumption,' and can handle extensive or frequent movements in a stack effectively. Nevertheless, because all such methods reconstruct the HDR image by maximizing the similarity with a selected reference image, the performance strongly relies on the exposure quality of the reference. If some parts of the reference image are saturated, or colored in black (under-exposed) or white (over-exposed), they can hardly provide accurate guidance for motion detection and exposure fusion. Occasionally, it is not even possible to select a 'good' reference for deghosting from the exposure stack.

1.2. Proposed approach

In this paper, we present a new reference-based method for producing tone-mapped-like HDR images that are free of ghosts, by using exposure fusion in dynamic scenes. To tackle the limitations of reference-based techniques, we propose consistency concepts of two types to enforce the guidance of the reference image for motion detection and ghost removal. One is the inter-consistency, which describes the similarities of pixel intensities among different exposures. However, owing to the different exposure settings, pixel intensities of different exposures vary widely. Hence, the inter-consistency is rather weak, and direct motion detection via the cue of intensity variations between the reference exposure and others is unreliable. To recover the inter-consistency, histogram matching is introduced, in order that the reference and target images share a similar degree of exposure. Following this, motion can easily be detected, because remaining discrepancies in the inter-consistency are mostly the result of content changes caused by object movements. The second concept is intra-consistency, which describes the mutual relationships of neighboring pixels in each exposure. Such consistency is imposed to eliminate outliers and ensure that adjacent pixels have similar fusion weights if they share similar intensities and structures. To achieve this goal, we group the pixels of each exposure into super-pixels, and then perform motion detection at a super-pixel level. This also reduces the complexity of the fusion procedure. Various experimental results demonstrate that the two types of proposed consistency considerations can help to determine motion more effectively than previous methods in exposure fusion.

In fact, consistency has already been considered in some studies [31]. For example, Pece and Kautz [24] employed median threshold bitmaps, which impose consistency relations between exposures to detect inconsistent pixels. Min et al. [22] imposed consistency by assuming that the gray levels at a particular pixel location must exhibit a non-decreasing property when the exposure values of an image change. Sidibe et al. [30] and Gallo et al. [5] both rely on the order relation between pixel values in images with different exposures to determine inconsistencies. Zhang and Cham [35,36] proposed the use of