Signal Processing: *Image Communication* 29 (2014) 191–202

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

High dynamic range video reconstruction from a stereo camera setup

Michel Bätz [a,*], Thomas Richter [a], Jens-Uwe Garbas [b], Anton Papst [b], Jürgen Seiler [a], André Kaup [a]

[a] *Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander University Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany*
[b] *Fraunhofer Institute for Integrated Circuits, Electronic Imaging, Am Wolfsmantel 33, 91058 Erlangen, Germany*

ARTICLE INFO

Available online 6 September 2013

Keywords:
High dynamic range video
Stereo camera setup
Stereo matching

ABSTRACT

To overcome the dynamic range limitations in images taken with regular consumer cameras, several methods exist for creating high dynamic range (HDR) content. Current low-budget solutions apply a temporal exposure bracketing which is not applicable for dynamic scenes or HDR video. In this article, a framework is presented that utilizes two cameras to realize a spatial exposure bracketing, for which the different exposures are distributed among the cameras. Such a setup allows for HDR images of dynamic scenes and HDR video due to its frame by frame operating principle, but faces challenges in the stereo matching and HDR generation steps. Therefore, the modules in this framework are selected to alleviate these challenges and to properly handle under- and oversaturated regions. In comparison to existing work, the camera response calculation is shifted to an offline process and a masking with a saturation map before the actual HDR generation is proposed. The first aspect enables the use of more complex camera setups with different sensors and provides robust camera responses. The second one makes sure that only necessary pixel values are used from the additional camera view, and thus, reduces errors in the final HDR image. The resulting HDR images are compared with the quality metric HDR-VDP-2 and numerical results are given for the first time. For the Middlebury test images, an average gain of 52 points on a 0-100 mean opinion score is achieved in comparison to temporal exposure bracketing with camera motion. Finally, HDR video results are provided.

1. Introduction

The high dynamic range of a real scene cannot be fully captured by regular CCD or CMOS camera sensors. The resulting images are therefore called low dynamic range (LDR) images. However, since the human visual system (HVS) is sensitive to a far wider dynamic range compared to LDR images, techniques for creating HDR content are useful to capture more information of a real scene, and thus, improve image quality. Several methods exist to generate HDR images which can be categorized into three classes according to [1]: Direct acquisition using particular HDR cameras [2], HDR rendering for synthetically created images, and exposure bracketing. The first class has two major drawbacks. The special hardware is very expensive and it cannot capture a dynamic range as large as when using an exposure bracketing technique. The second class

* Corresponding author. Tel.: +49 9131 85 28904; fax: +49 9131 85 28849.
E-mail addresses: baetz@lnt.de (M. Bätz), richter@lnt.de (T. Richter), jens.garbas@iis.fraunhofer.de (J.-U. Garbas), anton.papst@iis.fraunhofer.de (A. Papst), seiler@lnt.de (J. Seiler), kaup@lnt.de (A. Kaup).

0923-5965/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2013.08.016

is only relevant for computer-generated content, and hence, is not applicable for improving images of a real scene. The third class of HDR acquisition is a low-budget solution. In recent years, many consumer cameras started to offer exposure bracketing as a method to create HDR images of static scenes. For exposure bracketing, a series of LDR pictures is taken with different exposure times one after the other in order to estimate the camera response, convert the pixel values into radiance values and then merge them together to generate the HDR output. Two well-known approaches are the algorithm by Debevec and Malik [3] and the method by Robertson et al. [4]. The different exposures can also be realized by using different apertures or neutral density (ND) filters instead of exposure times.

However, this approach is not applicable for capturing dynamic scenes or even HDR video due to the temporal fashion of the exposure bracketing since the scene is changing over time. An example for a static scene without and with camera motion is illustrated in Fig. 1. It can be seen that temporal exposure bracketing completely fails when there is motion. There are approaches which compensate for the motion [5], but they have the drawback not being able to easily apply ND-filters, and hence, yielding artifacts from different motion blur or depth of field.

The goal of this work is therefore to allow for HDR image acquisition of dynamic scenes and HDR video. The proposed framework utilizes a stereo camera setup to realize a spatial exposure bracketing where the differently exposed images are distributed among the two views. Due to this, all necessary information to create an HDR image is available at the same time instant. This stereo setup leads to two challenges. On the one hand, HDR image reconstruction requires perfect alignment of the views after image warping. Therefore, a robust disparity estimation is needed. On the other hand, typical stereo matching algorithms rely on the brightness constancy assumption which does not hold for differently exposed views. Each stage of the proposed framework is built to properly handle the just mentioned challenges and also under-/oversaturated regions. Instead of applying a stereo matching approach, it would also be possible to utilize time-of-flight cameras in order to obtain the disparity maps. However, they can only generate low resolution disparity maps, increase the complexity of the setup, and are still considerably expensive. For testing purposes, datasets from the Middlebury Stereo Vision Page [6] and two self-recorded datasets are used. Besides visual quality evaluation, a suitable quality metric is applied with respect to a single-view HDR reference image.

The key novel contributions are the new one-stage framework compared to the two-stage frameworks in the literature [7,8] and the introduction of a saturation map to reduce errors in the final HDR image. Beyond this, an objective quality assessment, which is directly computed on the HDR images, is done for the first time.

The remainder of this article is structured as follows. The upcoming section presents previous work, whereas Section 3 covers the layout and the details of the framework and its modules. Experimental results are then given in Section 4. Finally, Section 5 concludes this contribution.

2. Previous work

This contribution is mainly inspired by the work in [7,8] where stereo matching is done in two stages and the camera responses are computed from the matches found in the first stage. These camera responses are then used for converting the input images to radiance space and a second stereo matching for disparity refinement is performed. In this work, however, unlike the existing algorithms, the camera responses are not estimated from the stereo images. This step is done only once in an offline step using temporal exposure bracketing on a single-view setup. For that reason, the disparity maps can be computed directly from the radiance space images (RSI) and the need for a second stereo matching stage is removed. In addition, the proposed offline step allows for camera setups with different sensors since each camera can be calibrated separately. Besides the novel one-stage framework, this work also presents the concept of applying a so-called saturation map before the actual HDR generation. In addition to that, objective quality evaluation results are provided for the first time. Furthermore, many ideas for the disparity estimation are taken from the currently

Fig. 1. Temporal exposure bracketing without (top) and with (bottom) camera motion.