Neurocomputing 149 (2015) 1396–1406

Contents lists available at ScienceDirect

Neurocomputing

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

Depth map super-resolution using stereo-vision-assisted model

Yuxiang Yang [a,*], Mingyu Gao [a,*], Jing Zhang [b], Zhengjun Zha [c], Zengfu Wang [b,c]

[a] Department of Electronics Information, Hangzhou Dianzi University, China
[b] Department of Automation, University of Science and Technology of China, China
[c] Institute of Intelligent Machines, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 10 March 2014
Received in revised form
26 August 2014
Accepted 30 August 2014
Communicated by L. Shao
Available online 6 September 2014

Keywords:
Depth map
Super-resolution
Stereo vision
Non-local prior
Local prior

ABSTRACT

In this paper, we propose a novel Stereo-Vision-Assisted (SVA) model for depth map super-resolution. Given a low-resolution depth map as input, we recover a high-resolution depth map using the registered high-resolution color stereo image pair. First, based on the mutual benefits between raw depth map and features of high resolution color image, we model the relationship with two constraint terms of local and non-local priors and sufficiently explore their complementary nature. Moreover, by considering reliable disparity pixels calculated from stereo matching algorithm, we formulate a stereo disparity regularization term to further reinforce the preservation of fine depth detail. Hence, our SVA objective function includes a non-local prior constraint, a local prior constraint and a stereo disparity prior constraint. In addition, we employ an efficient algorithm to optimize the objective function. Experimental results demonstrate that our approach can obtain superior performance in comparison with other methods.

1. Introduction

The ability to capture depth information of static real world objects has reached increased importance in many fields of application, such as three-dimensional reconstruction, manufacturing and prototyping, but also in the design of virtual worlds for movies and games. There exist a variety of depth measuring technologies to acquire depth information about our world. In general, they can be categorized into two major classes: passive range sensing methods and active range sensing methods.

Among passive range sensing approaches, stereo matching [1–6] is probably the most well-known and most widely used method. Classical two-frame stereo matching computes a stereo disparity from a pair of images under given camera configuration. The key objective of stereo matching methods is to reduce the matching ambiguities introduced by low-texture regions, and can be generally classified into two categories: local methods [1–4] and global methods [5,6]. A stereo algorithm is called a global method if there is a global objective function to be optimized. Otherwise, it is called a local method. Recently, the local stereo matching has recently experienced a renaissance with the up-come of adaptive support-weight schemes [2–4]. These local methods can deliver results close to the quality of global approaches. But unfortunately, neither local nor global approaches could solve the fundamental problems in stereo such as occlusion and low-texture regions. Hence, satisfactory depth results cannot be achieved by stereo matching approaches. Thus, depth maps from stereo matching for real scenes are often quite fragile.

Compare to passive range sensing approaches, time-of-flight sensors [7,8] and microsoft kinect [9] use active techniques to obtain near real-time scene depth. Time-of-flight sensors measure time delay between transmission of a light pulse and detection of the reflected signal on an entire frame once by using extremely faster shutter. But in the current generation, these time-of-flight sensors are limited in terms of resolution, swiss ranger SR-4000 [7] can only obtain depth maps with resolution 176×144. And PMD technologies camcube 2.0 [8] can get depth maps with resolution 204×204. In the past three years, the microsoft kinect [9,10] has obtained a high degree of concern in academic and games industry. The kinect sensor calculates the scene depth based on structured light technique and can provide depth maps with resolution up to 640×480. However, compared to time-of-flight sensors, the maximum measuring distance of kinect is about 3.5 m, which is about 10 m for swiss ranger SR-4000 and PMD technologies camcube 2.0. The depth measurement error of kinect greatly increases when the distance between the scene and the sensor exceeds 3.5 m, which limits the applications of kinect. Recently, Microsoft just launched the second generation of kinect. But the maximum measuring distance of kinect version 2 is only 4.5 m. Thus, it is meaningful to study how to overcome the low-resolution defect of those time-of-flight sensors [7,8].

In recent, many depth map super-resolution methods have been proposed which mainly focus on using a registered high-quality color image to provide information to enhance the raw depth map

* Corresponding authors.
E-mail addresses: yyx@hdu.edu.cn (Y. Yang), mackgao@hdu.edu.cn (M. Gao).

http://dx.doi.org/10.1016/j.neucom.2014.08.056
0925-2312/

of time-of-flight sensor. These approaches assume that there exist a joint occurrence between depth discontinuities and image color edges. In [11,12], Markov–Random–Field was taken as the form of this relationship. In [13–18], image filtering techniques such as joint bilateral filtering [15] or guided image filter [18] are presented to solve this depth super-resolution problem, where filter kernel is calculated by utilizing the color information. Although these methods are often computationally efficient, they will frequently result in copying of intensity texture into 3D geometry. Recently, Park et al. [19] used structural aware filter [20] and a local-weighting scheme to form an adaptive weighting filter. Specifically, they combine the initial depth inputs to calculate the filtering weights, which can avoid copying of intensity texture into 3D geometry. However, as accompany with low-quality initial depth maps, the super-resolution results will be affected by jagged effect, which reduces the quality of the reconstructed depth images (Fig. 1).

Thus, conventional depth map super-resolution methods still cannot well reconstruct high-quality depth maps. Motivated by above observations, we investigate the benefits of stereo matching and propose a novel SVA model for solving depth map reconstruction problem. The proposed SVA system is described in Fig. 2, the left color image is registered with the raw depth map and the right color image is registered with the left color image under given camera configuration. Specifically, based on the mutual benefits between raw depth map and features of high resolution color image, we model the relationship with two constraint terms of local and non-local priors and sufficiently explore their complementary nature. Then, by utilizing the reliable disparity pixels of stereo matching, we form a stereo disparity regularization term to further improve the quality of reconstructed depth maps. The experiments show that our approach can achieve high-quality depth maps in terms of both spatial resolution and depth precision.

The rest of this paper is organized as follows: Section 2 introduces the related work. In Section 3, we illustrate the proposed optimization

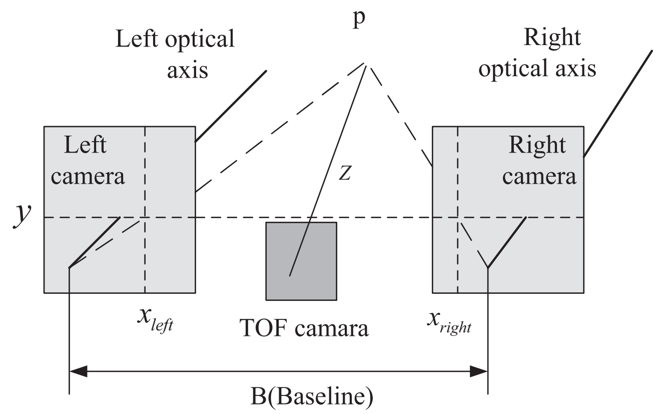

Fig. 2. The proposed Stereo-Vision-Assisted (SVA) system.

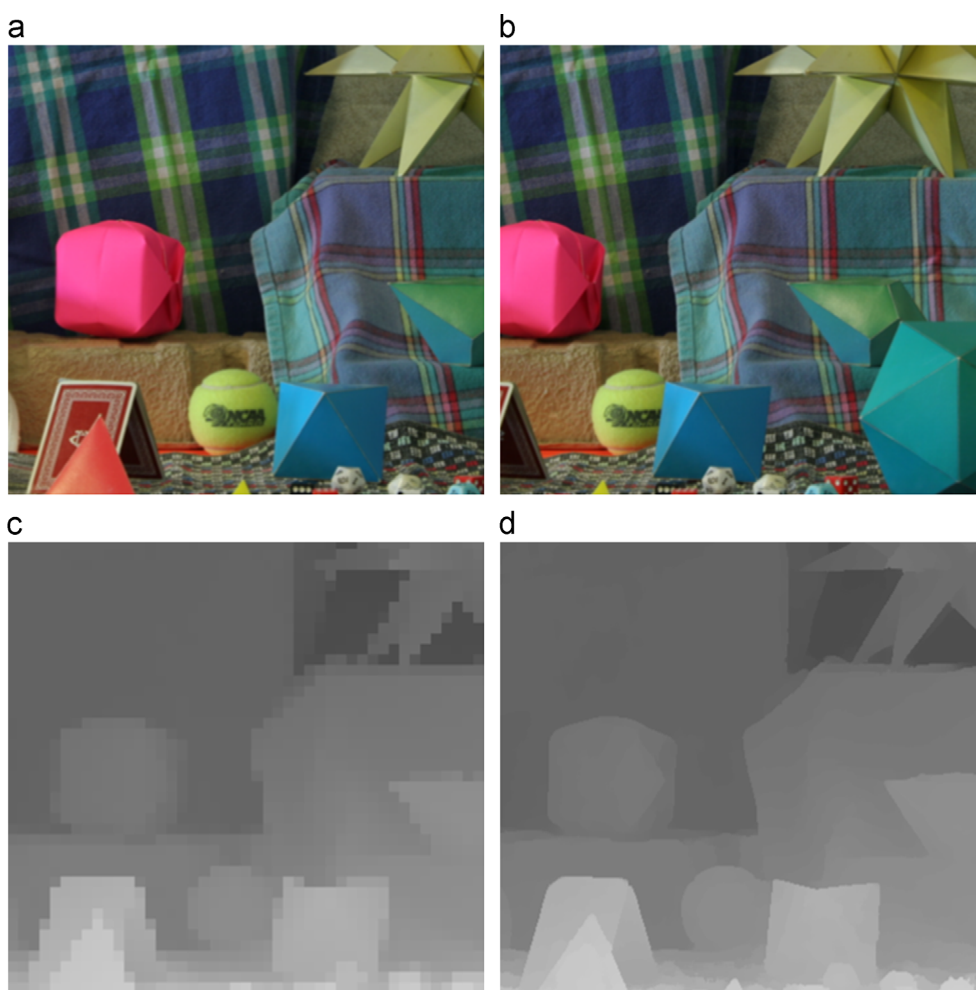

Fig. 1. Super-resolution results of Moebius (original resolution 695 × 555). (a) High-resolution left color image, (b) high-resolution right color image, (c) low-resolution depth map (down-sampling factor is 8), and (d) our reconstructed depth map based on the proposed SVA model. Note that our super-resolution result has sharp edges and rich details.