Neurocomputing 154 (2015) 325–336

Contents lists available at ScienceDirect

Neurocomputing

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

Depth map upsampling using compressive sensing based model

Longquan Dai, Haoxing Wang, Xiaopeng Zhang *

National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing, China

ARTICLE INFO

Article history:
Received 20 November 2013
Received in revised form
24 November 2014
Accepted 30 November 2014
Communicated by Zhouchen Lin
Available online 18 December 2014

Keywords:
Depth map
Compressive sensing
Upsampling

ABSTRACT

We propose a new method to enhance the lateral resolution of depth maps with registered high-resolution color images. Inspired by the theory of compressive sensing (CS), we formulate the upsampling task as a sparse signal recovery problem that solves an underdetermined system. With a reference color image, the low-resolution depth map is converted into suitable sampling data (measurements). The signal recovery problem, defined in a constrained optimization framework, can be efficiently solved by variable splitting and alternating minimization. Experimental results demonstrate the effectiveness of our CS-based method: it competes favorably with other state-of-the-art methods with large upsampling factors and noisy depth inputs.

1. Introduction

In recent years, a wide range of devices have been developed to measure the 3D information in the real world, such as laser scanners, structured-light systems, time-of-flight cameras and passive stereo systems. The depth maps (range images) captured with most active sensors usually suffer from relatively low resolution, limited precision and significant sensor noise. Therefore, effective depth map post-processing techniques are essential for practical applications such as scene reconstruction and 3D video production, especially for 3D face recognition [1] and 3D object recognition [2,3].

In this paper, we present a method to enhance the spatial resolution of a depth map with a registered high-resolution color image. Our method is based on two key assumptions: first, neighboring pixels with similar colors are likely to have similar depth values; second, just like most natural images, an ideal depth map without noise corruption has large smooth regions and relatively few discontinuities, and therefore can be approximated with a sparse representation in some transform domain such as multiscale wavelets. Although the first assumption has been extensively explored in recent depth post-processing work [4–9], relative less attention has been given to the second assumption [10,11].

Inspired by the theory of compressive sensing [12,13], we try to recover the upsampled depth map in a sparse signal reconstruction process. We first compute a set of measurement data from the low-resolution depth map. The measurement data near depth discontinuities are generated with a cellular automaton algorithm, and no filtering techniques are involved in the process. Then we reconstruct the depth signal in an optimization model, with constraints on measurements, smoothness and representation sparseness. An efficient numerical method is provided to solve the model with linear complexity in the number of the image pixels. Experimental results show that, by solving the problem in a CS-based framework, our algorithm can produce high quality depth results with relatively low resolution depth maps. And it shows stable performance under noisy conditions.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. Section 3 provides a brief introduction to the CS theory, whereas our CS-based upsampling model is presented in Section 4. After that, in Section 5, we describe how to generate the sampling data for the model, and we provide a numerical solution in Section 6. Section 7 reports the experimental results and discusses how to register a low resolution depth map and its companion high resolution color image as well as the influence of sampling pattern. At last, conclusions are given in Section 8.

2. Related work

As stated in Section 1, the idea of enhancing a depth map with a coupled color image is not new. Existing methods can be roughly classified as either filtering-based methods [5–8] or optimization-based methods [4,9].

Filtering-based methods employ color information with various edge-preserving filters [14,15]. Kopf et al. [5] use a joint bilateral filter to refine the upsampled depth results. Yang et al. [6] instead initialize a cost volume and iteratively smooth each cost slice with a bilateral filter. Sub-pixel accuracy is achieved with an interpolation

* Corresponding author. Tel./fax: +86 10 82544696.
E-mail addresses: lqdai@nlpr.ia.ac.cn (L. Dai), Haox.wang@nlpr.ia.ac.cn (H. Wang), xpzhang@nlpr.ia.ac.cn (X. Zhang).

http://dx.doi.org/10.1016/j.neucom.2014.11.060
0925-2312/

scheme. Huhle et al. [8] rely on nonlocal means filters (NLM) for depth denoising and upsampling. One advantage of filtering-based methods is that they can be easily parallelized on graphics hardware [7,8]. However, to find enough support for each pixel, large filtering kernels are often used, or the filters have to be performed iteratively, which might lead to over-smoothed depth results.

The methods which are more closely related to our algorithm are the optimization-based methods [4,9]. In [4], Diebel and Thrun construct a two-layer Markov Random Field model for depth map upsampling. The color information of neighboring pixels is encoded as edge weights of the graph. Recently, Park et al. [9] improve this model by including a multi-cue edge weighting scheme and an NLM energy term, which turns out to be very effective for preserving fine structures and depth discontinuities. To make the problem tractable, both methods use quadratic cost functions, which can be solved using standard numerical methods such as conjugate gradient. Our method differs from these methods in which we formulate the model with l_1 sparseness and total variation constraints, which shows the more robust behavior against noise and low sampling rates.

Recently, some researchers have explored sparse representations for depth map processing [10,11]. Tošić et al. [10] use sparse coding techniques [16] to learn a dictionary from Middlebury disparity data sets. This dictionary is exploited in a MRF model, which brings accuracy improvements for stereo depth estimation and range image denoising. Hawe et al. [11] propose a CS-based depth estimation method from sparse measurements. They show that, by taking only 5% of the disparity data as measurements, their method can recover the full disparity map with high accuracy, which is quite impressive. An essential point of their method is that the pixels lying at depth boundaries should be selected as sampling points, otherwise the reconstruction accuracy would be seriously affected. Unfortunately, such information is unavailable in our low-resolution depth inputs. We provide a novel method to generate measurements at these sampling positions with a registered reference color image, which proves crucial for the upsampling accuracy. Moreover, we employ a different CS model with better regularization ability.

3. CS theory and underdetermined linear system

CS theory finds an optimal solution $\mathbf{x}^n$ from the observed data $\mathbf{y} \in \mathbb{R}^m$ by reducing the problem to solving an underdetermined linear system. In mathematical terms, the observed data $\mathbf{y}^m$ is connected to the signal $\mathbf{x}^n$ of interest via

$$\Phi \mathbf{x} = \mathbf{y} \tag{1}$$

where $m < n$, x is the s-sparse vector which only has s nonzero components and the measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ models the linear measurement process. Traditional wisdom of linear algebra suggests that the number m of measurements must be at least as large as the signal length n. Indeed, if $m < n$, the classical linear algebra indicates that the underdetermined linear system Eq. (1) has infinite solutions. In other words, without additional information, it is impossible to recover $\mathbf{x}$ from $\mathbf{y}$ in the case $m < n$. However, with additional sparsity assumption, it is actually possible to reconstruct the sparse vector x from underdetermined measurements $\mathbf{y} = \Phi \mathbf{x}$ because many real-world signals are sparse. Even though they are acquired with seemingly too few measurements, exploiting sparsity enables us to solve the resulting underdetermined systems of linear equations. More importantly, there are many efficient algorithms for the reconstruction [17–19].

Specifically, CS theory reconstructs $\mathbf{x}$ as a solution of following combinatorial optimization problem

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \Phi \mathbf{x} = \mathbf{y} \tag{2}$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of a vector. However, the minimization problem is nonconvex and NP-hard. It thus is intractable for a modern computer. An alternative method is ℓ_1 minimization, which can be interpreted as the convex relaxation ℓ_0 minimization.

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \Phi \mathbf{x} = \mathbf{y} \tag{3}$$

One major shortcoming of above considerations is that they do not carry over to the complex setting such as the contaminated measurements $\mathbf{y}$. As a remedy, we can directly extend the ℓ_1 minimization (3) to a more general ℓ_1 minimization taking measurement error into account, namely,

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \le \epsilon \tag{4}$$

It is worth noting that the solution of Eq. (4) is strongly linked to the output of the ℓ_1 denoising, which consists in solving, for some parameter $\beta \ge 0$

$$\min_{\mathbf{x}} \beta \|\mathbf{x}\|_1 + \tfrac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \tag{5}$$

Expecting Eq. (3) can restore any $\mathbf{x}$ for any Φ is unreasonable. Instead, CS theory only proves that for any integer $n > 2$ s, there exists a measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m = 2$ s rows such that every s sparse vector $\mathbf{x} \in \mathbb{R}^n$ can be recovered from its measurement vector $\mathbf{y} \in \Phi \mathbf{x}$ as a solution of Eq. (3). However, finding out the measurement matrix Φ is a remarkably intriguing endeavor. To date, it is still an open problem to construct explicit matrices which are provably optimal in a compressive sensing setting. One novelty of the work is just providing a method to construct the measurement matrix Φ for depth upsampling.

The depth upsampling problem can be reduced to the problem of using the underdetermined linear system Eq. (1) to find an optimal solution $\mathbf{x}^n$ from few measurements $\mathbf{y}^m$ because depth upsampling aims to restore a high-resolution depth map from a low-resolution depth map and the values of the low-resolution depth map can be viewed as the samplers of the high-resolution depth map. Unfortunately, depth maps are not usually sparse in the canonical (pixel) basis. But they are often sparse after a suitable transformation, for instance, a wavelet transform or discrete cosine transform. This means that we can write $\mathbf{x} = \Psi \mathbf{z}$, where $\mathbf{z}^n$ is a sparse vector and $\Psi^{n \times n}$ is a unitary matrix representing the transform. Recalling $\mathbf{y} = \Phi \mathbf{x}$, depth upsampling finds a solution $\mathbf{x} = \Psi^T \mathbf{z}$ from the underdetermined linear system Eq. (6).

$$\Phi \Psi^T \mathbf{z} = \mathbf{y} \tag{6}$$

Obviously, the underdetermined linear system (6) is similar to the underdetermined linear system (1) and the similarity leads us to guess that CS theory also deals with this kind of underdetermined linear system. Indeed, studying Eq. (6) is the origin of CS theory. Without any doubt, CS theory can efficiently solve it [12,13,20] by using the optimization problem (3).

4. CS-based upsampling model

We build our upsampling model upon a fundamental fact that many signals can be represented or approximated with only a few coefficients in a suitable basis. Consider a high-resolution depth map $\mathbf{d} \in \mathbb{R}^n$ in column vector form, it can be linearly represented with an orthonormal basis $\Psi \in \mathbb{R}^{n \times n}$ and a set of coefficients $\mathbf{x} \in \mathbb{R}^n$: $\mathbf{d} = \Psi \mathbf{x}, \mathbf{x} = \Psi^T \mathbf{d}$. The map $\mathbf{d}$ is linearly measured m times ($m \ll n$), which leads to a set of measurements $\mathbf{y} \in \mathbb{R}^m$ with a measurement matrix $\Phi \in \mathbb{R}^{m \times n}$: $\mathbf{y} = \Phi \mathbf{d}$. The CS theory tries to recover depth map