ARTICLE IN PRESS

Neurocomputing ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Neurocomputing

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

Filling Kinect depth holes via position-guided matrix completion

Zhongyuan Wang [a,*], Xiaowei Song [b], ShiZheng Wang [c], Jing Xiao [a], Rui Zhong [a], Ruimin Hu [a]

[a] NERCMS, School of Computer, Wuhan University, Wuhan 430072, China
[b] School of Electric and Information Engineer, Zhongyuan University of Technology, Zhengzhou 450007, China
[c] School of Electrical & Electronic Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form
11 May 2015
Accepted 14 May 2015

Keywords:
Kinect
Depth map
Hole filling
Matrix completion

ABSTRACT

Due to occlusion and measurement errors, there commonly exist holes, inaccurate depth values and noises in depth maps acquired by low-cost Kinect devices. The artifacts seriously affect the practical applicability of depth maps and thus filling holes is a critical pre-processing task for 3D applications. Without the assistance of the accompanied color image, the representative bilateral filtering and inpainting methods hardly provide satisfactory recovery results. Since the depth map containing holes can be naturally regarded as a corrupted low-rank matrix of missing entries, this paper addresses hole filling problem from the perspective of low rank matrix completion. Our method identifies the positions of invalid pixels in hole regions, and then incorporates the known entries into the formulation which considers the low-rank constraint on results and the sparse constraint on residuals. Owning to the well-established principle component pursuit theory, our method substantially boosts the Kinect depth recovery performance in terms of accuracy and reliability.

1. Introduction

In many visual scenarios, depth information plays an important role, such as depth based virtual viewpoint rendering, 3D modeling and gesture recognition. These stereo vision related applications pose great challenges for the acquisition, coding and synthesis of depth information [1,2]. Microsoft's Kinect provides a great convenience for real-time and active acquisition to the scene depth information. However, the structured light measurement method used by Kinect is susceptible to the impact of occlusion, transparent objects and rich texture regions, leading to holes or wrong depth values at object edges and flat regions in captured depth maps. Because such artifacts seriously affect the practical applicability of depth maps, filling holes is a critical pre-processing step for Kinect based applications.

Various methods have been developed to recover lost depth information and eliminate noise simultaneously for Kinect depth maps, which roughly fall into two categories: filtering and inpainting. Buades et al. [3] used a non-local filtering scheme to enhance depth maps, whereas it significantly blurs the sharp edges. Based on a joint histogram of a high-resolution color video and its corresponding low-quality depth video, Min et al. [4] instead proposed a weighted mode filtering method to prevent the depth boundaries from being blurry. As a widely used edge-preserving filtering approach, Fu et al. [5] incorporated bilateral filter into temporal motion compensation to recover the missing depth pixels. However, since depth maps are inherently discontinuous in temporal domain, a certain degree of edge dilation is present and a number of small holes remain. Camplani et al. [6] further proposed to apply iteratively a joint-bilateral filter to Kinect depths, in which the filter weights are specified with respect to three different factors: visual data, depth information and a temporal-consistency map. Unfortunately, this method fails to account for the overall contribution of all the pixels around the hole centre and often yields unsatisfactory results when large holes exist. Yang et al. [7] employed a frequency-counting based non-linear filter to improve the accuracy of the high resolution depth map, which is whereas primarily severed to remove sensor noise in the captured low-resolution depth map rather than fill in holes. Hu et al. [8] advocated a color image guided locality regularized representation to reconstruct the missing depth pixels, which analytically computes the filter weights by a ridge regression model in the accompanied color image. Due to the closed-loop approach for solving weights, this method can generally obtain more optimal reconstruction weights than above mentioned bilateral filtering methods and thus produce impressive results.

In contrast to filtering algorithms which more or less result in poor results near depth discontinuities, inpainting techniques seem more appealing. A popular inpainting algorithm is the Fast Marching Method (FMM) by Alexander Telea [9], but it does poorly when applied to depth maps as it is designed for generic hole filling in colored images. Liu et al. [10] proposed an extended FMM

* Corresponding author.
E-mail address: wzy_hope@163.com (Z. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.05.146

Please cite this article as: Z. Wang, et al., Filling Kinect depth holes via position-guided matrix completion, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.05.146

approach with an aligned color image guiding depth inpainting. In an energy minimization based approach [11], an additional TV regularization term is introduced to produce smooth depth maps with sharp boundaries preserved. Relatively, the techniques adequately accounting for depth discontinuities between objects can faithfully restore depth maps by Kinect.

Previous methods generally took advantages of the corresponding color image to enforce their filtering or inpainting performance. Without the assistance of the accompanied color image, they hardly provide satisfactory results in terms of depth-edge preservation. Since high-quality color images are not always available in some scenarios, for instance, in a poor illumination environment, and meanwhile the noisy color image can often mislead depth map enhancement because adjacent objects may have exceedingly close colors [12], it is meaningful to develop approaches dedicated to the restoration of depth maps independent on color images.

Depth maps have both characteristics of sudden changes at edge regions and large flat areas inside objects. As shown in [13], the regular patterns within the object, especially some symmetry structures, can be seen as constituting a low-rank matrix, while the presence of occlusion, corruption and other errors can be seen as a sparse matrix. Inspired by recent progress in low rank matrix recovery research [14–16], we propose a new hole filling method for Kinect depth maps via low rank matrix completion. This method recovers missing depth values by decomposing the Kinect depth map into a low rank matrix and a sparse error matrix, which represent a "desirable" depth map and an error data destroying its low-rank structure, respectively. Moreover, as the spatial locations of holes in Kinect depth maps can be explicitly identified, in other words, the missing entries in constituted incomplete depth matrix are already available, their coordinates can be used to facilitate the optimization on low-rank matrix completion. Experimental results show that our method can reliably remove artifacts, smooth depth maps in homogeneous regions and improve the accuracy near object boundaries.

The remainder of this paper is organized as follows. Section 2 particularly presents the depth filling method via position-guided low rank matrix completion. Section 3 shows the experimental results. Finally, conclusions are given in Section 4.

2. Proposed method

This section mainly concerns the proposed representation for position-guided low rank matrix completion and its optimization as well.

Low rank matrix completion problem considers how to complete a low rank matrix from an incomplete observation with both missing and corrupted entries. To be more specific, let matrix $\boldsymbol{P} \in \mathrm{R}^{m\times n}$ be the observed Kinect depth map, unknown matrix $\boldsymbol{Q} \in \mathrm{R}^{m\times n}$ denote the objective counterpart with the missing entries to be assigned, $\boldsymbol{Q}$ is expected to be reconstructed by filling in holes in $\boldsymbol{P}$ with the help of matrix complete theory. This procedure depends on two constraints: small differences between $\boldsymbol{P}$ and $\boldsymbol{Q}$; low rank of $\boldsymbol{Q}$. The first constraint is measured with the Frobenius norm $\|\cdot\|_F$ of a matrix, while the second is related to nuclear norm $\|\cdot\|_*$. The nuclear norm $\|\cdot\|_*$ of a matrix $\boldsymbol{X}$ is defined as the summation of its singular values, i.e., $\|\boldsymbol{X}\|_* = \sum_i \sigma_i(\boldsymbol{X})$, where $\sigma_i(\boldsymbol{X})$ denotes the i^{th} largest singular value. In the typical matrix completion problem [14], $\boldsymbol{Q}$ is recovered from its incomplete observation $\boldsymbol{P}$ by solving the following minimization:

$$\min_{\boldsymbol{Q}} \|\boldsymbol{Q} - \boldsymbol{P}\|_F^2 + \mu\|\boldsymbol{Q}\|_*, \tag{1}$$

where $\mu > 0$ is a weighting parameter. When the observation matrix has only a very few number of missing values, Eq. (1) can be used to address a variety of image processing questions. For example, Lu et al. [17] cast depth completion and denoising as an incomplete matrix factorization problem. However, due to large errors or invalid depth values, the occluded depth map tends to have outliers in the known entries so that the direct application of Eq. (1) to depth completion will damage fine depth details.

In order to overcome the dilemma of typical matrix completion algorithms, robust principle component pursuit (PCP) [16] was proposed to recover a low rank matrix with sparse outliers on the observed entries. Particularly, it introduces another unknown sparse matrix $\boldsymbol{E}$ to describe the outliers and recovers both $\boldsymbol{Q}$ and $\boldsymbol{E}$ by solving

$$\min_{\boldsymbol{Q},\boldsymbol{E}} \|\boldsymbol{Q}\|_* + \lambda\|\boldsymbol{E}\|_1 \text{ s.t. } \boldsymbol{P} = \boldsymbol{Q} + \boldsymbol{E}, \tag{2}$$

where ℓ_1 norm denoted by $\|\cdot\|_1$ accumulates the absolute values of all the entries in a matrix, reflecting the sparsity constraint. $\lambda > 0$ is an appropriately chosen regularization parameter, balancing the tradeoff between the rank of matrix $\boldsymbol{Q}$ and the sparseness of error matrix $\boldsymbol{E}$. As suggested in [16], the penalty parameter λ can be fixed as $\lambda = \frac{1}{\sqrt{\max(m,n)}}$. Alternatively, λ will be empirically tuned to give the best results by slightly adjusting it around the theoretical candidate.

PCP assumes the locations of the missing entries in the observed depth to be unavailable. However, if the locations of some of the corrupted entries are known a *priori*, we can leverage that information to facilitate solving the matrix completion problem. In reality, the missing pixels in the Kinect depth map have distinctive characteristics, which can thus be used to identify their spatial positions. To illustrate this issue, we show a depth map and its 3-dimensional surface in Fig. 1. The dark black hole areas in the depth map in Fig. 1(a) are associated with the missing pixels, and the depth levels are depicted in different colors in Fig. 1(b). As obviously shown in the 3-dimensional plot, the depth magnitudes in holes are exactly zeros or close to zeros while others are much larger than zeros. Therefore, depth levels imply their spatial coordinates in depth map so that we can exploit such information to pursue the optimal solution of PCP problem.

Mathematically, we have a problem of recovering a low rank matrix with known exact indexes of the missing entries. We denote by Ω the locations of available entries in the observed matrix $\boldsymbol{P}$ defined in Eq. (2). The linear subspace Ω of $m \times n$ matrices just corresponds to the complementary to the locations of missing entries. Let π_Ω represent the orthogonal projection operator in the subspace Ω,

$$\pi_\Omega(\boldsymbol{X})(i,j) = \begin{cases} \boldsymbol{X}(i,j), & \text{if}(i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

Since the depth pixels in holes are completely unavailable, to be filled in, it is meaningless to pursue the minimal deviation from their latent values. Therefore, we incorporate known entries of subspace Ω into the formulation Eq. (2) and have the following position-guided form:

$$\min_{\boldsymbol{Q},\boldsymbol{E}} \|\boldsymbol{Q}\|_* + \lambda\|\boldsymbol{E}\|_1 \text{ s.t. } \pi_\Omega(\boldsymbol{P}) = \pi_\Omega(\boldsymbol{Q} + \boldsymbol{E}). \tag{4}$$

The above problem is just a variant of Eq. (2) imposed on the spatial constraint of position prior information. Thus, it can be seen an extension of existing results for the low-rank matrix completion problem [14,16], which wishes to recover a low-rank matrix from large but sparse errors. Relative to the naïve application of Eq. (2) to the hole-filling problem in depth maps, our representation explicitly exploits the full use of the known coordinate positions of missing entries, leading to more ideal recovery and efficiency computation.

Instead of direct optimization on Eq. (4), we solve its unconstrained Lagrangian version: