J. Vis. Commun. Image R. 35 (2016) 15–24

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A DAISY descriptor based multi-view stereo method for large-scale scenes ☆

Bindang Xue *, Lei Cao, Donghai Han, Xiangzhi Bai, Fugen Zhou, Zhiguo Jiang

Image Processing Center, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 27 March 2015
Accepted 15 November 2015
Available online 2 December 2015

Keywords:
Computer vision
3D reconstruction
Multi-view stereo
Large-scale
DAISY descriptor
Photometric discrepancy function
Feature matching
Feature point growing

ABSTRACT

Normalized cross-correlation (NCC) has been widely used as the matching cost function in multi-view stereo methods. However, NCC is vulnerable in the occlusion area and edge region of large-scale scenes because of color distortion and illumination changes. To alleviate the above problems, we present an improved patch based multi-view stereo method by introducing a photometric discrepancy function based on DAISY descriptor. In the patch extraction stage, a new corresponding point matching method based on the DAISY descriptor is proposed and the epipolar constraint is used to filter mismatched points. In the patch optimization stage, a photometric discrepancy function based on DAISY descriptor is proposed to measure the photo-consistency among reconstructed patches to identify reliable patches. Finally, dense patches are obtained by expanding sparse patches with global visibility information and patch optimization. Experimental results show that the proposed algorithm obtains better reconstruction results in occlusion and edge regions of large-scale scenes.

1. Introduction

Multi-view stereo (MVS) reconstruction is to obtain 3D scene models from a collection of images. The existed MVS algorithms can be roughly classified into four categories [1]: the voxel based methods [4–9], the surface evolution based methods [10–12], the depth-map merging based methods [3,13–15,17,18], and the feature point growing based methods [19–23]. The voxel based methods, most of which are applicable for small and compact objects within a tight bounding box, calculate a cost function on a 3D volume at first and then extract a surface from this volume. The surface evolution based methods evolve a surface to minimize a cost function iteratively. These methods require a reliable initial guess which is difficult to obtain for large-scale scenes. The depth-map merging based methods consists of depth maps calculating and depth maps merging, which require sufficient images to increase the reconstruction accuracy and performance. The feature point growing based methods extract feature points from textured regions firstly, then reconstruct space points based on these features, and expand reconstructed space points to regions with less texture. Usually, NCC is used as the matching cost between two patches. However, NCC can be easily influenced by color distortion and illumination changes in the occlusion area and edge region of large-scale scenes.

Aiming to alleviate the above-mentioned problems, a DAISY descriptor based photometric discrepancy function is proposed to identify reliable patches. DAISY [16,17] descriptor is a gradient information based and local invariant descriptor, which is robust to affine transform and illumination changes. Especially, this descriptor is highly discriminative to those simple pixels opposite to corner points and therefore suited for wide-baseline dense matching [17]. The main idea of generating initial sparse patches can be described as a two-stage procedure as in patch-based multi-view stereo (PMVS) method [21]: the initial patch candidate extraction stage and the candidate patch optimization stage. In the former stage, the proposed method utilizes DAISY descriptor and the main orientation of the descriptor to conduct corresponding points matching, and filters mismatched points with epipolar constraint. In the following stage, the conjugate gradient method is used to optimize positions and orientations of the remaining patches along the view directions, and then to minimize photometric discrepancy function to obtain refined sparse patches. Finally, dense patches are obtained by growing the sparse patches with global visible information and patch optimization strategy.

2. Proposed method

The proposed method can be implemented in three steps: Step 1, obtaining reference images and neighbor images to narrow the scope of searching corresponding projected points; Step 2, patch initialization by the DAISY descriptor based photometric

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author.
E-mail address: xuebd@buaa.edu.cn (B. Xue).

http://dx.doi.org/10.1016/j.jvcir.2015.11.007
1047-3203/

discrepancy function, and this step consists of feature extraction, candidate patches generation based on geometrical optical triangulation [27], and initial sparse patches extraction by utilizing DAISY descriptor based photometric discrepancy function; Step 3, generating dense patches and filtering false patches by global visibility. In step 3, we assume that the scenes to be reconstructed are nearly Lambertian as in most algorithms. With this assumption, we can use depth continuity and normal consistency between reconstructed patches and their neighborhoods to expand the sparse patches. In the end, a dense collection of patches can be obtained and represented as the estimated surface close to large-scale scene. Fig. 1 shows the pipeline of the proposed method.

2.1. DAISY descriptor

DAISY [16,17] is an improved version of SIFT and GLOH [25,26] descriptor. It is implemented by replacing weighted responses of orientation histograms in SIFT with simple convolutional computation for different isotropic Gaussian kernels and gradient histograms.

Given an input image, H orientation maps are firstly calculated, i.e. conduct gradient computation along H sampling directions and $G_o(u,v)$ represents gradient scores at point (u,v) of orientation o.

$$G_o = \left(\frac{\partial I}{\partial o}\right)^{+} (a)^{+} = \max(a, 0) \tag{1}$$

Then $H \times Q$ convolutional orientation maps are obtained by convoluting orientation maps with different Gaussian kernels with an increasing standard deviation $\sum$. Here, Q represents convolution times, v indicates current index of convolution, and $G_o^{\sum_1} = G_{\sum_1} * G_0$

$$G_o^{\sum_v} = G_{\sum} * G_o^{\sum_{v-1}}, \quad \sum = \sqrt{\sum\nolimits_v^2 - \sum\nolimits_{v-1}^2} \tag{2}$$

To maintain DAISY descriptor's isotropy, T directions are selected evenly in the range of $[0°, 360°]$, and Q key points are evenly sampled in each direction. Then we totally get $(T \times Q + 1)$ key points. Convolutional orientation histograms of each key point are represented as corresponding feature vector $h_{\sum}(u,v)$, and the whole descriptor $D(u_o, v_o)$ is the combination of $h_{\sum}(u,v)$. Its dimension is $H \times (T \times Q + 1)$.

$$h_{\sum}(u,v) = \left[G_1^{\sum}(u,v), \ldots, G_H^{\sum}(u,v)\right]^T \tag{3}$$

$$\begin{aligned} D(u_o, v_o) = \Big[& h_{\sum_1}^T(u_o, v_o), \\ & h_{\sum_1}^T(l_1(u_o, v_o, R_1)), \ldots, h_{\sum_1}^T(l_T(u_o, v_o, R_1)), \\ & \vdots \\ & h_{\sum_Q}^T(l_1(u_o, v_o, R_Q)), \ldots, h_{\sum_Q}^T(l_T(u_o, v_o, R_Q))\Big] \end{aligned} \tag{4}$$

2.2. Photometric discrepancy function based on DAISY descriptor

Photometric discrepancy function describes photo-consistency between the projected image points on the reference image and all the visible images of the given patch. When the photometric discrepancy value is smaller than the given threshold, the corresponding projected image points are photo-consistent. This information will be used to determine whether the patch is photo-consistent or not, i.e. whether the patch is a reliable estimate of scene's surface or not.

2.2.1. Assign the main orientation for DAISY descriptor

DAISY descriptor does not contain orientation information. We introduce the main orientation of DAISY descriptor to help evaluate image projections of reconstructed patches in the corresponding visible images.

Given a patch p, the corresponding reference image I_1 and a neighbor image I_2, project the center point of p to I_1 and I_2 respectively, and denote the projected image points as cp_1 and cp_2. To calculate the main orientation of projected points, two space points along positive direction of the local coordinate axes x, y of p are projected to I_1 and I_2. The directions of projected local axes in I_1 and I_2 can be calculated respectively, and denote the direction vectors of projected local axes in I_1 as $\vec{d}_{x1} = [u_{x1}, v_{x1}]$ and $\vec{d}_{y1} = [u_{y1}, v_{y1}]$, the direction vectors of projected local axes in I_2 as $\vec{d}_{x2} = [u_{x2}, v_{x2}]$ and $\vec{d}_{y2} = [u_{y2}, v_{y2}]$.

The orientation o of the DAISY descriptor at the projected image point can be defined as the angle between the sum of the two direction vectors of projected local axes and the horizontal axis of the image as Eq. (5),

$$o = \begin{cases} \arccos(\langle \vec{d}_{x1} + \vec{d}_{y1}, \vec{u} \rangle / \|\vec{d}_{x1} + \vec{d}_{y1}\|), & (u_{x1} + u_{x2}) \geqslant 0 \\ \arccos(\langle \vec{d}_{x1} + \vec{d}_{y1}, \vec{u} \rangle / \|\vec{d}_{x1} + \vec{d}_{y1}\|) + \pi, & (u_{x1} + u_{x2}) < 0 \end{cases} \tag{5}$$

where $\vec{u}$ is a unit vector along the positive direction of the horizontal axis of the image.

2.2.2. Compute the photometric discrepancy function based on DAISY descriptor

After assigning the main orientation for DAISY descriptor, photometric discrepancy function can be calculated. Firstly, multi-channel DAISY descriptor of the projected points of patch p on image I_i is introduced as Eq. (6),

$$D^t(p, I_i) = D^t(cp_i, o_i, I_i), \quad t \in \{R, G, B\} \tag{6}$$

where cp_i is the projected point on image I_i, o_i is the main orientation of DAISY descriptor at cp_i.

In order to get a more robust descriptor and reduce the influence of illumination changes and occlusions, $D^t(p, I_1)$ is normalized as follows,

$$D_n^t = \frac{D^t - \bar{D}^t}{\sigma} \tag{7}$$

where $\bar{D}^t$ is the average of feature vectors and σ is the standard variance of feature vectors.

The photometric discrepancy function based on DAISY for patch p on image I_1 and I_2 is defined as Eq. (8),

$$C(p, I_1, I_2) = 1 - \sum_{t \in \{R,G,B\}} \left\| D_n^t(cp_1, o_1, I_1) - D_n^t(cp_2, o_2, I_2) \right\| \tag{8}$$

The DAISY based photometric discrepancy function for patch p on all its visible images is defined as Eq. (9).

$$T(p) = \frac{1}{|VN(p)|} \sum_{I \in VN(p)} C(p, R(p), I) \tag{9}$$

In this paper, we assume that the objects or scenes are Lambertian so that pixels in different images projected from the same space point on the scene surface are photo-consistent. In the presence of specular highlights or obstacles, the DAISY based photometric discrepancy function may not be able to accurately identify reliable patches. To handle this situation, we restricted the value of DAISY based photometric discrepancy function within a reasonable range in order to remove false matches, i.e., a pair of pixels is a good match if the corresponding discrepancy value of