J. Vis. Commun. Image R. xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Stereo matching cost computation based on nonsubsampled contourlet transform

Ka Zhang [a,b,c,*], Yehua Sheng [a,b,c], Haiyang Lv [a]

[a] Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, 1# WenYuan Road, Nanjing, Jiangsu 210023, PR China
[b] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, 1# WenYuan Road, Nanjing, Jiangsu 210023, PR China
[c] Key Laboratory of Police Geographic Information Technology, Ministry of Public Security, Nanjing Normal University, 1# WenYuan Road, Nanjing, Jiangsu 210023, PR China

ARTICLE INFO

Article history:
Received 12 April 2014
Accepted 22 September 2014
Available online xxxx

Keywords:
Stereo image matching
Nonsubsampled contourlet transform
Feature vector
Weighted matching cost
Matching accuracy

ABSTRACT

A new matching cost computation method based on nonsubsampled contourlet transform (NSCT) for stereo image matching is proposed in this paper. Firstly, stereo image is decomposed into high frequency sub-band images at different scales and along different directions by NSCT. Secondly, by utilizing coefficients in high frequency domain and grayscales in RGB color space, the computation model of weighted matching cost between two pixels is designed based on the gestalt laws. Lastly, two types of experiments are carried out with standard stereopairs in the Middlebury benchmark. One of the experiments is to confirm optimum values of NSCT scale and direction parameters, and the other is to compare proposed matching cost with nine known matching costs. Experimental results show that the optimum values of scale and direction parameters are respectively 2 and 3, and the matching accuracy of the proposed matching cost is twice higher than that of traditional NCC cost.

1. Introduction

One of the chief tasks in stereo image matching is computation of matching cost. The matching cost is defined as the degree of similarity between the reference window of matching pixel in the left image and the searching window of candidate conjugate pixel in the right image. The matching cost is the basis for finding conjugate pixels in two images taken at the same scene, and plays an important role in such fields as remote sensing image registration [1], 3D vision based on images [2], and digital photogrammetry [3]. The robustness of the matching cost has a direct impact on reliability of matching results.

Matching cost computation methods can be divided into two types. One is based on grayscales of pixels, and the other on post-processed features of grayscales. The gray-based matching cost is computed by comparing the similarity of grayscale distribution between two matching windows. Indexes of NCC (Normalized Cross Correlation) [4], ZNCC (Zero Mean Normalized Cross Correlation) [5], TAD (Truncated Absolute Difference) [6], SSD (Sum of Squared Difference) [7] and SAD (Sum of Absolute Difference) [8,9] are some common gray-based costs. These matching costs have been widely used and can acquire good matching results in image regions with open and smooth terrain. But they cannot work well in poor information regions of image such as textureless water bodies, and repetitive texture regions.

There are two steps in feature-based matching cost computation. The first step is to carry out some kinds of feature transform on stereo image for constructing feature vectors. In feature transform, scale invariant feature transform (SIFT), affine-invariant feature extraction, wavelet transform and census transform are often used [10–13]. The second step is to compute the matching cost with Euclidean distance or hamming distance between feature vectors [14].

However, heterogeneity between the central pixel and its neighboring pixels of the matching window is ignored in above matching costs. During computation, all pixels are equally treated, which leads to some shortcomings on their performance of adaptability and anti-interference. Therefore, weighted matching cost computation methods, which consider heterogeneity of pixels in the matching window, are studied by more and more researchers. Yoon and Kweon [15] presented a weighted matching cost computation model based on color similarity in the CIELab color space and geometric proximity in the image space. Richardt et al. [16] introduced a real-time stereo matching technique using the bilateral grid to improve computation efficiency of Yoon's method. Hosni et al. [17] designed a local weighted matching cost computation model based on geodesic support weights. They used

* Corresponding author at: Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, No. 1, WenYuan Road, Nanjing, China.
E-mail addresses: zhangka81@126.com (K. Zhang), shengyehua@njnu.edu.cn (Y. Sheng), lvhaiyang368@163.com (H. Lv).

http://dx.doi.org/10.1016/j.jvcir.2014.10.002
1047-3203/© 2014 Elsevier Inc. All rights reserved.

geodesic distance between two pixels of the support window in RGB color space to compute geodesic support weight for a pixel. Nalpantidis and Gasteratos [18,19] incorporated biologically and psychologically inspired features to design an adaptive weighted SAD computation framework in RGB color space, and proposed a luminosity-compensated dissimilarity measure (LCDM) computation method in HSL color space. Based on these costs, they designed a new adaptive support weight stereo correspondence algorithm for robotic applications in non-ideal lighting conditions. Zhou and Boulanger [20] computed relative gradient of each pixel in the matching window firstly, then designed Gaussian weighted SAD matching cost based on relative gradients to complete stereo matching. Yang [21] developed a non-local stereo matching cost aggregation method based on a tree structure. In this method, all image pixels were treated as nodes of the tree based on color similarity, and every node received support weights from all other nodes on the tree during cost aggregation.

In traditional computation model of weighted matching cost, structure features of pixels are less taken into consideration. Instead, only grayscale similarity and image space proximity are utilized to compute support weights. However, geometric distance proximity in image space cannot effectively represent heterogeneity characteristics of pixels, especially pixels in disparity discontinuity regions of the image. In addition, after comparing the accuracy of disparity maps generated by 360 different match cost measures, Neilson and Yang [22] believed that the choice of match cost measure used in an algorithm had a large impact on the accuracy of the disparity maps generated by this algorithm. Gong et al. [23] also obtained the result that properly designed cost aggregation approaches could significantly improve the quality of the disparity maps by evaluating the performances of six different cost aggregation approaches. They suggested that further study on new cost aggregation approaches would be a promising research direction. Therefore, a new weighted matching cost computation based on nonsubsampled contourlet transform (NSCT) is proposed for stereo image matching in this paper, and it consists of two main steps: structure feature extraction based on NSCT and weighted matching cost computation. The main purpose of the paper is to enhance reliability of matching cost and accuracy of stereo matching through utilization of NSCT high frequency structure features.

The organization of remainder paper is given as follows: Section 2 details computation principle of the proposed approach. Section 3 demonstrates detailed experimental data, design scheme and corresponding results. Finally, conclusions and potential future work will be presented in Section 4.

2. Proposed approach

2.1. General idea

The flow diagram of the proposed algorithm is shown in Fig. 1. Firstly, left and right images of one stereopair are respectively decomposed into high frequency subband images by nonsubsampled contourlet transform (NSCT) at different scales and along different directions. Secondly, the structure feature of each pixel is constructed with high frequency coefficients of NSCT subband images. Thirdly, by utilizing high frequency structure feature vectors and RGB color vectors of pixels, the NSCT-based weighted matching cost computation model between left reference window and right searching window is designed based on the gestalt laws. Lastly, by utilizing standard stereopairs of the famous Middlebury platform [7], the proposed weighted matching cost is tested by two experiments to show its feasibility and validity. One experiment is designed to analyze the influence on matching accuracy with the change of NSCT parameters, and confirm their optimum values with optimal matching accuracy. The other experiment is designed to compare the proposed matching cost with some existing costs.

2.2. Structure feature vector extraction based on NSCT

NSCT is a new multi-scale geometric analysis tool for images developed on the basis of contourlet transform [24]. NSCT are characterized by multi-scale, multi-direction expansion, powerful anisotropy, shift-invariant and good localization in space domain and frequency domain. Compared to wavelet transform, which has only horizontal, vertical and diagonal components in high frequency information, NSCT can decompose high frequency information into combination of multi-direction components, and it gives NSCT a good nonlinear approximation capability for approximating 2D piecewise smooth functions with linear singularity. After one image is decomposed by NSCT with J scales, it is transformed into following images: one low frequency approximate image and $\sum_{i=1}^{J} 2^{d_i}$ high frequency subband images (d_i is the number of direction parameter in nonsubsampled directional filter bank (NSDFB) at the ith scale). All these decomposed images have the same resolution with the original image. The structure diagram of NSCT is shown in Fig. 2.

Fig. 3 shows an example of NSCT decomposition results (values of transform scale and direction parameters in NSDFB at each scale are respectively 1 and 2) of one image (shown in Fig. 3(a), 384 columns $\times$ 288 rows). After NSCT, the image is decomposed into one low frequency image (shown in Fig. 3(b)) and four high frequency subband images along four directions (shown in Fig. 3(c)). In addition, to clearly display high frequency subband images, their high frequency coefficients are multiplied with 255.

Seen from Fig. 3, high frequency detailed structures of the original image along different directions are displayed on NSCT high frequency images. We can use these details to construct the structure feature vector for each pixel. After the image $f(x, y)$ is decomposed by NSCT (values of scale and direction parameters at each scale are respectively n and d_i, $i = 1, 2, \ldots, n$), high frequency directional subband images $G_j(x, y)$ ($j = 1, 2, \ldots, m$) can be obtained, and the total number m of high frequency images is $\sum_{i=1}^{n} 2^{d_i}$. Then for each pixel $p(x_k, y_k)$ in the image $f(x, y)$ (x_k and y_k are respectively

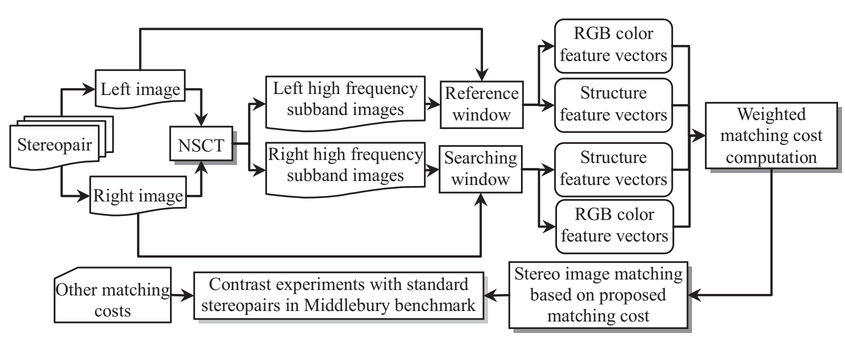

Fig. 1. Flow diagram of the proposed approach.

Please cite this article in press as: K. Zhang et al., Stereo matching cost computation based on nonsubsampled contourlet transform, J. Vis. Commun. (2014), http://dx.doi.org/10.1016/j.jvcir.2014.10.002