Computer Vision and Image Understanding 125 (2014) 223–236

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Stereo reconstruction using high-order likelihoods ☆

Ho Yub Jung [a], Haesol Park [a], In Kyu Park [b], Kyoung Mu Lee [a,*], Sang Uk Lee [a]

[a] Department of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, Republic of Korea
[b] Department of Information and Communication Engineering, Inha University, Incheon 402-751, Republic of Korea

ARTICLE INFO

Article history:
Received 23 September 2013
Accepted 21 April 2014
Available online 2 May 2014

Keywords:
High-order likelihood
Markov random field
Stereo vision
Global matching framework
Graph cuts
High-order cliques
Occlusion handling

ABSTRACT

Under the popular Markov random field (MRF) model, low-level vision problems are usually formulated by prior and likelihood models. In recent years, the priors have been formulated from high-order cliques and have demonstrated their robustness in many problems. However, the likelihoods have remained zeroth-order clique potentials. This zeroth-order clique assumption causes inaccurate solution and gives rise to undesirable fattening effect especially when window-based matching costs are employed. In this paper, we investigate high-order likelihood modeling for the stereo matching problem which advocates the dissimilarity measure between the whole reference image and the warped non-reference image. If the dissimilarity measure is evaluated between filtered stereo images, the matching cost can be modeled as high-order clique potentials. When linear filters and nonparametric census filter are used, it is shown that the high-order clique potentials can be reduced to pairwise energy functions. Consequently, a global optimization is possible by employing efficient graph cuts algorithm. Experimental results show that the proposed high-order likelihood models produce significantly better results than the conventional zeroth-order models qualitatively as well as quantitatively.

1. Introduction

Graphical models based on Markov random field (MRF) have become dominant in low-level computer vision problems such as image denoising, segmentation, and stereo vision. The ill-posed nature of these problems necessitates a strong presence of regularizing priors. MRF provides an effective representation of local dependency between the target variables. Efficient global optimization algorithms such as α-expansion, partial quadratic pseudo-Boolean optimization (QPBO) fusion, and belief propagation allow convincing estimates of the minimum energy states [3,7,20,26,30] of such models. The MRF framework has proven effective for incorporating smoothness priors into pixel labeling problems.

Recently, considerable progress has been made in modeling priors. The first-order priors are the most popular because these can be implemented directly in pairwise MRFs. In addition, second-order and high-order priors are introduced for the stereo and segmentation problems [14,35]. The success of the high-order models for priors has encouraged the development of high-order energy minimization algorithms, which also have become a persistent topic in computer vision [11,17,19]. The high-order models yield more accurate results than their low-order counterparts, though with an increase of computational time. To solve this problem on a conventional optimization framework, the larger cliques are first reduced to pairwise cliques before the optimization procedure.

However, high-order likelihoods have not been considered yet and the independent distribution has often been assumed so far. If we narrow our perspective to the stereo vision problem, the matching cost (likelihood) is modeled as a sum of single-variable functions, while the surface smoothness cost (prior) is modeled as the sum of two-variable or three-variable functions. Although various window-based and segment-based matching costs have been proposed, the matching costs have been nominally encased in the zeroth-order potentials until recently. Under the conventional zeroth-order clique assumption, the left and right images are filtered before matching. Note that due to this simple assumption, the conventional methods suffer form inaccuracy in matching and have noticeable undesired artifacts such as fattening effect around the disparity discontinuity. While it is desirable to filter the non-reference image after it is warped to the reference image plane, it causes high-order cliques because a pixel value in the filtered image becomes a function of many disparities inside the supporting window.

☆ This paper has been recommended for acceptance by Nicu Sebe.

* Corresponding author. Fax: +82 2 878 1452.
E-mail addresses: hoyubjung25@yahoo.com (H.Y. Jung), sankata0@snu.ac.kr (H. Park), pik@inha.ac.kr (I.K. Park), kyoungmu@snu.ac.kr (K.M. Lee), sanguk@ipl.snu.ac.kr (S.U. Lee).

http://dx.doi.org/10.1016/j.cviu.2014.04.009
1077-3142/

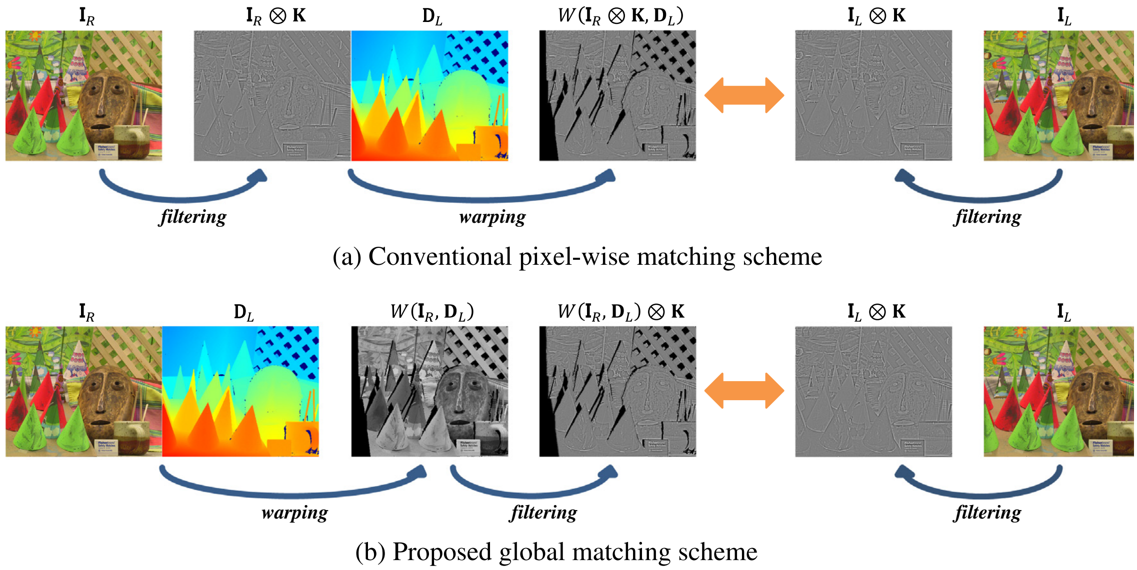

Fig. 1. Matching scheme for the zeroth-order (conventional) and high-order (proposed) potentials. (a) The filtered images are typically matched pixel-wise. The left and right stereo images are filtered before the match. For each pixel position x, the difference in filter responses becomes the matching cost, and the likelihood is a product of the independent distributions at each pixel position. (b) In the proposed global likelihood approach, the left image is first synthesized using the right image and a disparity map. Then, the filter is applied to the original left image and the synthesized left image. The single likelihood distribution is defined by an exponential in the measure of dissimilarity between these two images. The depth map in this figure is the ground truth which is used for visualizing the warping process. Note that the proposed method incorporates occlusion modeling (Section 3.5).

In this paper, we investigate the global matching framework and the subsequent high-order likelihood modeling for the stereo reconstruction problem. This paper differs from the previous works in proposing the elevation of the existing window-based matching costs to high order. We demonstrate how the global matching framework provides a natural interpretation for the high-order likelihoods. Furthermore, the global optimization is made possible through the pairwise clique reduction. Many of the previously introduced matching costs, such as those calculated by mean subtraction and census filters, are modeled as high-order likelihoods and optimized using graph cuts. The tests of various matching costs show that the elevation of likelihood to high order significantly eliminates the fattening effect that is inevitable when using the window-based zeroth-order likelihoods. Although the proposed method has high computational complexity due to the large neighborhood system and non-submodular edges, we show that the use of parallel implementation of the optimization algorithm on graphics processing unit (GPU) is an effective way to compensate the penalty. The early version of this paper was published in [12]. In this paper, it is expanded significantly to accommodate all high-order linear filters with L2 distances, in addition to the Census filter discussed in [12].

The rest of this paper is organized as follows. Section 2 introduces previous works related to the proposed approach. Section 3 describes the high-order likelihood from the global matching framework. In the same section, the pairwise clique reductions are provided for high-order matching costs obtained with linear and census filters. The fusion move allows asymmetrical occlusion handling by combining multiple proposal states. In Section 4, the energy minimization technique and its parallel implementation on GPU are described. Section 5 presents the experimental results, in which different high-order likelihoods are compared for various regularizing values. Section 6 gives concluding remarks.

2. Related works

Binocular stereo vision represents the visual matching problem in its simplest form. Despite the straightforwardness of the problem, the stereo matching is difficult even for images taken in well-controlled environments. Numerous matching costs and their implementation have been proposed. We explain how these can benefit from the elevation to high order.

Window-based methods involve a basic premise that nearby pixels have similar depths. More elaborate segment-based methods assume that pixels with similar color belong to the same 3D planar surface. Note that those two assumptions are also shared by prior models. In segment-based or window-based matching, nearby pixels ought to have the same depth surface, and the depth discontinuities are expected to align with color discontinuities. Thus, matching costs are aggregated for each segment by assuming the same surface. There are various dissimilarity measures, ranging from the classic sum of absolute differences to cross correlation and recent examples based on the intensity histogram [38]. However, the critical concerns of segment-based matching reside in designing the shape and size of the segment or smoothness priors, rather than the matching cost of the segment-based matching. With few exceptions, the shape and size of the segment are based on the color and distance similarity. By segmenting the stereo images, the smoothness prior is applied as hard surface constraints on the set of similar pixels. Recently, a soft segmentation has been introduced using adaptive support weights [8,36]. The soft segmentation is also based on the same assumptions as the matching costs based on hard segments, and greater support weight is given to the pixels that are closer in terms of color and position.

Alternatively, the stereo images can be preprocessed before calculating the matching costs. Instead of a color consistency assumption, the matching features of interest are transformed by filtering the stereo images before the matching cost calculation. The filtered images are subsequently matched in a pixel-wise or window-wise manner. Marr and Poggio first introduced phase and filter banks that resemble edges for stereo images [21]. In real time stereo systems, a Laplacian of Gaussian (LoG) filter is used to eliminate an intensity offset after noise smoothing [18]. Other linear filters include bilateral subtraction filters and mean subtraction filters [2,32]. Moreover, nonparametric rank and census filters output a binary map that evaluates the relative pixel intensities in a