Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 106 (2007) 85–96
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Bayesian stereo matching

Li Cheng [a,b,*], Terry Caelli [b]

[a] Department of Computing Science, University of Alberta, Edmonton, Alta., Canada T6G 2E8
[b] National ICT Australia, Research School of Information Science and Engineering, Australian National University, Canberra ACT 2601, Australia

Received 17 November 2004; accepted 8 September 2005
Available online 12 February 2007
Communicated by Arthur E.C. Pece

Abstract

A Bayesian framework is proposed for stereo vision where solutions to both the model parameters and the disparity map are posed in terms of predictions of latent variables, given the observed stereo images. A mixed sampling and deterministic strategy is adopted to balance between effectiveness and efficiency: the parameters are estimated via Markov Chain Monte Carlo sampling techniques and the Maximum A Posteriori (MAP) disparity map is inferred by a deterministic approximation algorithm. Additionally, a new method is provided to evaluate the partition function of the associated Markov random field model. Encouraging results are obtained on a standard set of stereo images as well as on synthetic forest images.

Keywords: Generative model; Stereo vision; Monte Carlo sampling; Bayesian analysis; Markov random field

1. Introduction

The goal of stereo matching is to infer the optimal disparity map for a given pair of images. Unfortunately, hand-crafting of model parameters is often necessary to ensure satisfactory results for specific image pairs [1]. A remedy is to adopt the Bayesian paradigm which naturally solves this problem of automatic parameter tuning, by treating both the unknown disparity map and the related parameters as random variables. The problem is then to infer the optimal distributions of the random variables. The merit of this scheme has been demonstrated in the related area of medical image processing by Higdon et al. [2]. However, the Bayesian approach is typically computationally demanding due to the use of sampling algorithms to explore the space of plausible distributions.

We propose the use of a generative Bayesian framework for stereo matching, which addresses the inference of disparity map and the estimation of parameters under a unified scheme. Further, efficient Markov Chain Monte Carlo (MCMC) methods [3] are proposed for parameter estimation, and a deterministic approximation algorithm, loopy belief propagation (LBP)[1] [6], is adopted to infer the disparity map.

Recently, a number of optimization methods have been used to solve the stereo problem. These include using simulated annealing [7], dynamic programming [8] and LBP [9] to infer the optimal disparity map. However, unlike the proposed method, existing models are not fully Bayesian, and their solution techniques are substantially different. The novel contributions of this work are threefold. First, stereo matching is explicitly addressed as a generative process, as illustrated in Fig. 1. Second, a Bayesian framework that naturally unifies the tasks of inferring the disparity

* Corresponding author. Present address: National ICT Australia, Research School of Information Science and Engineering, Australian National University, Canberra ACT 2601, Australia. Fax: +61 2 6230 7499.

E-mail addresses: Li.Cheng@anu.edu.au (L. Cheng), Terry.Caelli@anu.edu.au (T. Caelli).

[1] Two other papers of this issue [5] also use the LBP algorithm.

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2005.09.009

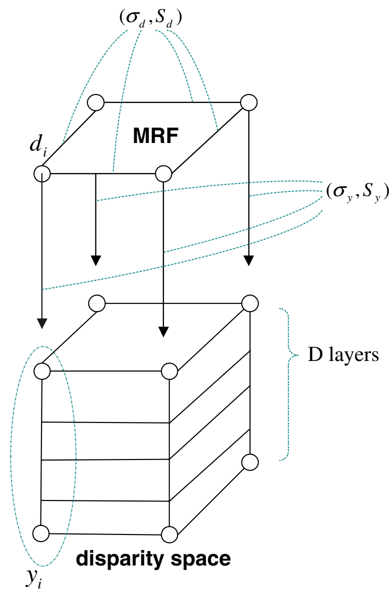

Fig. 1. A 2×2 2D lattice example that illustrates the proposed generative model for stereo matching. On the bottom, the 3D disparity space y is compiled by measuring the pixelwise dissimilarities of the left and right images with respect to shifts along the epipolar line. On the top, the disparity map d is modelled as a Markov random field. For a node i, given the latent disparity d_i, the pre-compiled observation y_i is independent of the rest of the disparity space y.

map and estimating the model parameters, is proposed. Third, a new method, based on the path sampling approach [10], is derived to evaluate the partition function of the underlying Markov random field (MRF). In particular, the proposed evaluation method is shown to bear theoretical advantages over both the coding and the pseudo-likelihood method [11]. Moreover, it greatly reduces the computational load when integrated into the MCMC samplers, and empirical experiments demonstrate the convergence behaviors of the proposed mixing strategy.

The Bayesian model is presented in Section 2, followed by a mixed updating strategy in Section 3. Details regarding the coding and the pseudo-likelihood methods are shown in Appendix C.1, and details of the proposed partition function evaluation method are presented in Appendix C.2. Finally, experiments are conducted in Section 4, with an empirical analysis of convergence behavior of the proposed approach addressed in Section 5.

2. The Generative Model

We assume a dense binocular stereo setting (e.g. [1]), where two views (left and right images, rectified to satisfy the epipolar constraint) of the same scene are presented. With the left image being the reference view, the task is to infer the disparity of each pixel, and to automatically estimate the model parameters for the image pair. This model, however, could be easily extended to more general scenarios.

Let $i = 1, \ldots, n$ index a 2D lattice of image pixels. Let $y = \{y_i\}$ denote a 3D disparity space with each y_i a vector of length D, where D is the range of possible disparity values. Essentially, y stores sufficient statistics about the input images, with each layer (see Fig. 1) storing the pixelwise dissimilarities of the two images, after shifting the left image horizontally a certain number of pixels. Therefore, y is referred to as the "*observed*" disparity space. The disparity map $d = \{d_i \in \{1, \ldots, D\}\}$ is modelled as a Markov random field (MRF) [12]. The proposed model consists of two components: the sensor model and the prior model. For the sensor model, $p(y|\, d, \sigma_y, s_y)$ captures the statistical dependencies of the observation y on the latent disparity MRF d, while the prior model $p(d|\sigma_d, s_d)$ addresses the neighboring dependencies within the disparity map. For convenience, denote the model parameters as $\theta = \{\sigma_y, s_y, \sigma_d, s_d\}$, with (σ_d, s_d) parameters of the prior model and (σ_y, s_y) parameters of the sensor model.

Because of the uncertainty of θ for different image pairs (see Fig. 1), Bayesian theory [13] treats θ as unknown and assigns a prior distribution for θ. By establishing the likelihood $p(y|\, d, \theta)$, the priors $p(d \mid \theta)$ and $p(\theta)$, the joint posterior is defined as

$$p(d, \theta|y) \propto p(y|d, \theta)p(d|\theta)p(\theta). \tag{1}$$

Our task is then twofold. First, we want to infer the MAP disparity map d^*:

$$d^* = \arg\max_d p(d|\theta^*, y), \tag{2}$$

where θ^* denotes the optimal parameter estimate. Second, we have to estimate the model parameters θ by its expectation

$$\theta^* = \int_\theta \theta p(\theta|d, y) \mathrm{d}\theta. \tag{3}$$

2.1. The Sensor and the Prior Models

Given the random variable $x \in \mathbb{R}^n$ and parameters (σ, s), we consider a class of density functions [2]

$$p(x|\sigma, s) = \frac{1}{\sigma^n z(s)} \exp\left\{-\frac{1}{s} u(x|\sigma, s)\right\}. \tag{4}$$

Here $u(x|\sigma, s) = \sum_i \rho(x_i/\sigma, s)$ is the energy function, and $z(s)$ is the normalization constant to ensure $p(x|\sigma,s)$ a valid density distribution. $\rho(\cdot, \cdot)$ is the potential function, with scale parameter $\sigma \in (0, \infty)$ and shape parameter $s \in (0, 2]$. We further decompose $x = \{x_i\}_{i=1}^n$ to represent a random field that could be either the MRF d or the disparity space y.

One reason for choosing this type of function is that the potential function, $\rho(\cdot, \cdot)$, unifies many existing function forms, both convex and non-convex, into one general representation [2]. In particular, it includes the generalized Gaussian distribution, when the potential function admits the following form,

$$\rho\left(\frac{x}{\sigma}, s\right) = |\frac{x}{\sigma}|^s, \tag{5}$$

when $s = 2$ we have the Gaussian distribution.

In Fig. 2, the two panels in each row show the effect of varying the shape parameter s, and the two panels in each