Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2504–2520
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A statistical approach to sparse multi-scale phase-based stereo

I. Ulusoy[a,*], E.R. Hancock[b]

[a]Computer Vision and Intelligent Systems Research Laboratory, Electrical and Electronics Engineering Department, Middle East Technical University, 06531 Ankara, Turkey
[b]Department of Computer Science, University of York, York Y01 5DD, UK

Received 14 December 2005; received in revised form 22 March 2006; accepted 23 October 2006

Abstract

In this study, a multi-scale phase based sparse disparity algorithm and a probabilistic model for matching uncertain phase are proposed. The features used are oriented edges extracted using steerable filters. Feature correspondences are estimated using phase-similarity at multiple scale using a magnitude weighting scheme. In order to achieve sub-pixel accuracy in disparity, we use a fine tuning procedure which employs the phase difference between corresponding feature points. We also derive a probabilistic model, where phase uncertainty is trained using data from a single image pair. The model is used to provide stable matches. The disparity algorithm and the probabilistic phase uncertainty model are verified on various stereo image pairs.

Keywords: Stereo; Probabilistic stereo; Multi-scale; Phase; Steerable filter; Orientation

1. Introduction and motivation

In sparse stereo [1–5], distinctive image features are extracted and corresponding pairs are matched using a feature-based similarity criterion. The advantage of these methods is that they can produce very accurate results. The disadvantage is that the methods fail in structureless or textureless image regions, and the resulting pattern of correspondences can become rather sparse.

To be useful for successful stereo matching, local features must be stable or robust under typical image deformations, such as scale changes, noise, brightness variations and rotation. Moreover, once detected the features and their locations should provide salient information that can be used to establish unambiguous correspondence information. There are a number of ways in which this can be achieved. For instance, a brute force approach is to perform correlation matching of the normalized brightness values in an image neighborhood. A more subtle approach is to capture the local brightness structure using differential operators [6]. However, one of the most elegant methods is to use local phase information, and to locate correspondences using phase congruence.

There are several examples of the phase-based approach in the literature. For instance, Jenkin and Jepson [7], Jepson and Fleet [8,9] and Sanger [10] developed methods based on the phase behavior of the output of band-pass Gabor filters. Jepson and Fleet [9] provide a justification for phase-based techniques based on an analysis of the stability of the band-pass phase behavior under typical distortions that exist between the left and right stereo views. Carneiro and Jepson [11] have shown that the phase information provided by steerable filters is often locally stable with respect to scale changes, noise and mutual brightness variations. In a more recent study, they take this work one step further and show that it is also possible to achieve stability under rotation by selecting the steerable filter [12]. They also conclude that although phase-based local features perform better in terms of mutual illumination changes, in the case of 2D rotation and sub-pixel translation, for scale and large shear changes, the robustness of phase-based features to scale changes must be improved. This is done by using a denser sampling in the scale-space, and this in turn provides stable multi-scale phase-based features [11,12].

* Corresponding author. Tel.: +90 312 2104558; fax: +90 312 2101261.
E-mail address: ilkay@metu.edu.tr (I. Ulusoy).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.10.019

Experience shows that disparity estimates from local phase-differences are reliable near edges, but yield poor results at intermediate locations. In Ref. [13] a probabilistic lattice structure is proposed to fill unreliable regions that result after phase-based disparity estimation. The method uses a simple smoothness constraint motivated by Markov random fields.

There are many other probabilistic algorithms that can be used to estimate the disparity map between stereo images. These methods share the feature of maximizing the conditional probability of the observed disparities given the stereo images. The main problem in developing such methods is how to calculate the conditional probability given only the stereo image set. In Ref. [14] a Gibbs distribution is used to develop a Markov Random Field model for the distribution of feature points. In this paper it is proved that it is possible to estimate the disparity of a position, if all of the joint probability distributions between neighborhood disparities are known in advance. Myers et al., have used graph edit distance to find stereo correspondence in wide-baseline uncalibrated stereo images [15]. Olson [16,17] has done work on the stereo matching problem in the context of matching edge images for robot localization. Using a probabilistic framework, his method considers the distance from each pixel in the template to the closest matching pixel in the image. The joint probability density function is approximated as a product of individual probability distribution functions, assuming that the distance measurements are independent.

In this study we propose a sparse disparity algorithm where multi-scale phase similarity is used as a matching criterion. Our approach is as follows: We commence from feature points detected using a steerable filtering method [18], as done in Ref. [19] (Section 2.1). Instead of corners, which are very sparse, we use oriented edges as the features to be matched. In this way the points at depth and intensity discontinuities are selected as features independent of their scale and orientation. In contrast, most stereo algorithms fail at discontinuities [20]. In this study we deal with points at object edges and within textured regions which correspond to depth and intensity discontinuities. These features provide reliable information for the computation of multi-scale phase. The phases at different scales are calculated and a phase vector is formed for each feature point. Then, correspondences are estimated using the phase-similarity at multiple scales together with a magnitude weighting scheme (Section 2.2). In this way we avoid the singular points encountered in the method of Jepson and Fleet [8,9]. After calculating disparity from the positional difference between corresponding points, fine-tuning in disparity is performed using the phase difference between corresponding feature points (Section 2.3). In this way we achieve a sub-pixel accuracy in disparity. We also develop an alternative method of arriving at feature point correspondences where we use a probabilistic model (Section 3). We train our model from data in a single image pair, and use the model to provide stable matches in other image pairs. We model the probability distribution of phase differences for pairs of corresponding points using a mixture of von Mises distributions at each scale. In this way the probability of a phase difference can be used as a measure of the confidence of correspondence (Section 3.1). We base our decision concerning correspondences on the product of probability distributions over the different scales used in the analysis (Section 3.2). This provides a higher degree of discrimination than the method discussed in Section 2. Finally Section 4 presents our results and Section 5 provides some conclusions and discusses directions for future research.

2. Multi-scale phase based disparity algorithm

In this section we describe our filter-based approach to disparity estimation. We commence by describing the steerable filters used in our study. Next we explain how the filter outputs can be used to establish correspondences. Finally, we explain how the correspondences can be used to estimate disparity with sub-pixel accuracy.

2.1. Feature extraction using steerable filters

We follow Ludtke et al. [21] and construct our feature detection model using the idea of orientation selective cells and hyper column structure in the visual cortex. The feature points used in our study are detected using steerable filters as in Ref. [19]. The analytic filter used as the template filter is constructed from the filters described by Freeman and Adelson [18]. If $\mathbf{x}=(x, y)$ is the pixel location in the image, $I(\mathbf{x})$ then the template filter is given by

$$h(\mathbf{x}) = g(\mathbf{x}) + jq(\mathbf{x}), \tag{1}$$

where $g(x, y)$ is chosen to be the 4th derivative of a Gaussian with a variance $\sigma = 1/\sqrt{2}$ normalized to unit energy (Eq. (2)) and $q(x, y)$ is chosen to be a steerable approximation to the Hilbert Transform of $g(x, y)$ (Eq. (3)),

$$g(x, y) = (0.934 - 3.738x^2 + 1.246x^4)\mathrm{e}^{-(x^2+y^2)} \tag{2}$$

and

$$q(x, y) = (2.858x - 2.982x^3 + 0.3975x^5)\mathrm{e}^{-(x^2+y^2)}. \tag{3}$$

For an arbitrary orientation θ, $g_\theta(\mathbf{x})$ can be synthesized using five basis filters of orientation 0°, 36°, 72°, 108° and 144° as in Eq. (4). Also, $q_\theta(\mathbf{x})$ can be synthesized using six basis filters of orientation 0°, 30°, 60°, 90°, 120° and 150° as in Eq. (5). $\mathbf{R}_{\theta_i}$ is the rotation matrix in both the cases:

$$\begin{aligned} g(\mathbf{R}_\theta\mathbf{x}) = \sum_{i=1}^{5} \frac{1}{5}\,(1 + 2\cos(2(\theta-\theta_i)) \\ + 2\cos(4(\theta-\theta_i)))g(\mathbf{R}_{\theta_i}\mathbf{x}), \end{aligned} \tag{4}$$