Pattern Recognition Letters 38 (2014) 70–77

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A robust cost function for stereo matching of road scenes ☆

Alina Miron [a,b,*], Samia Ainouz [a], Alexandrina Rogozan [a], Abdelaziz Bensrhair [a]

[a] INSA Rouen/LITIS laboratory – EA4108, 76801 Saint-Etienne du Rouvray, France
[b] Babes-Bolyai University, Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 2 June 2013
Available online 22 November 2013

Keywords:
Stereo vision
Census Transform
Cross Comparison Census
Graph cuts
Matching cost comparison

ABSTRACT

In this paper different matching cost functions used for stereo matching are evaluated in the context of intelligent vehicles applications. Classical costs are considered, like: sum of squared differences, normalised cross correlation or Census Transform that were already evaluated in previous studies, together with some recent functions that try to enhance the discriminative power of Census Transform (CT). These are evaluated with two different stereo matching algorithms: a global method based on graph cuts and a fast local one based on cross aggregation regions. Furthermore we propose a new cost function that combines the CT and alternatively a variant of CT called Cross-Comparison Census (CCC), with the mean sum of relative pixel intensity differences (DIFFCensus). Among all the tested cost functions, under the same constraints, the proposed DIFFCensus produces the lower error rate on the KITTI road scenes dataset[1] with both global and local stereo matching algorithms.

1. Introduction

Stereo matching has been an intensely studied topic in research due to its crucial applications that vary from 3D reconstruction to image-based rendering or object hypothesis generation.

Our field of application is intelligent vehicles, in particular the detection of road obstacles like pedestrians. The objective is to reduce the hypothesis space using the information provided by the disparity map. Classic techniques like sliding window produce an extensive search space while ground subtraction based techniques can not be applied to dynamic scenes. Robust disparity map is consequently essential in order to have pertinent hypothesis over the location of pedestrians.

Most of the stereo matching algorithms rely on four important steps: Cost computation; Cost aggregation; Disparity computation/optimisation and Disparity refinement (Scharstein and Szeliski, 2002). Each step is important for the quality of the disparity map, with the cost computation step being crucial as it stands at the basis of the stereo matching algorithms. A given cost function can be minimised using different methods within the step of *disparity computation/optimisation*. There exists many techniques for energy minimisation that vary from local methods that find the minimum of the cost function using a winner takes it all strategy like in Zhang et al. (2009) and Mei et al. (2011), to global techniques like graph cuts (Kolmogorov and Zabih, 2001), dynamic programming (Bleyer and Gelautz, 2008), or belief propagation (Felzenszwalb and Huttenlocher, 2006; Klaus et al., 2006). Szeliski et al. (2008) and Kolmogorov and Rother (2006) compared different optimisation algorithms based on energy functions and showed that the lowest energy is produced by the graph cuts.

Choosing a cost function has to take into account the radiometric distortions, since in real traffic situations these are very pronounced. Some of the causes are sun flares, reflections or just camera sensor differences. In this context, our contribution is twofold.

- First, we compare different cost functions in order to be able to choose the most adapted one for our field of application. For this, we combine different cost functions with two stereo matching methods: a global technique based on Graph cuts (GC) (Kolmogorov and Zabih, 2001) and a local stereo matching algorithm based on cross zones aggregation with local voting (Zhang et al., 2009).
- Secondly, we propose a new cost function, based on a combination between census bitstring and the mean sum of relative differences, that is robust to radiometric distortions.

2. Related works

Choosing the right cost function is paramount for having a good disparity map. As presented in Hirschmuller and Scharstein (2009), the costs can be divided into parametric functions, where the cost

☆ This paper has been recommended for acceptance by Andrea Torsello.

* Corresponding author at: INSA Rouen/LITIS laboratory – EA4108, 76801 Saint-Etienne du Rouvray, France.

E-mail addresses: alina.miron@insa-rouen.fr (A. Miron), samia.ainouz@insa-rouen.fr (S. Ainouz), alexandrina.rogozan@insa-rouen.fr (A. Rogozan), abdelaziz.bensrhair@insa-rouen.fr (A. Bensrhair).

[1] http://www.cvlibs.net/datasets/kitti

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.11.009

incorporates the magnitude of pixel intensity, and non-parametric ones. Common parametric costs include those based on absolute differences and square differences, along with the window-based approaches: Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD) (Birchfield and Tomasi, 1998), Normalised Cross-Correlation (NCC), Zero-mean based costs (like ZSAD, ZSSD and ZNCC), or costs computed on the first (gradient) or second (Laplacian of Gaussian) image derivatives. Non-parametric costs include the popular Census and Rank methods (Zabih and Woodfill, 1994).

There exists several studies where comparison of cost functions is performed, the most extended ones being made by Hirschmuller and Scharstein (2007) and Hirschmuller and Scharstein (2009). In comparison with the study from 2007, where six cost functions where tested, the one from 2009 compared fifteen different stereo matching costs in relation with images affected by radiometric differences. These costs are compared using three different stereo matching algorithms: one based on global energy optimisation (Graph cuts), one using semi-global matching (Hirschmuller, 2008) and a local window-based algorithm. They conclude that the cost based on CT gives the best overall performance. In comparison with (Hirschmuller and Scharstein, 2009) that use both simulated and real radiometric changes in a laboratory environment, in this paper the experiments are performed on real road images from the KITTI dataset (Geiger et al., 2012) which presents significant radiometric differences. Besides the cost functions that provided the best results in Hirschmuller and Scharstein (2009), we also test some recent functions based on CT that gave good results on the Middlebury dataset.[2] Moreover we propose a new cost function $C_{DiffCensus}$ that remains robust to radiometric changes. These costs will be presented in the following two sections.

3. State of the art matching costs

In this section we define each matching cost function used in the experiments. Along with our new proposed function, eight different cost functions will be compared: Squared Intensity Differences C_{SD}, Zero-mean Normalised Cross-Correlation C_{ZNCC} (Hirschmüller et al., 2002; Hirschmuller and Scharstein, 2009), Census Tranform C_{CT} (Zabih and Woodfill, 1994), Cross Comparison Census C_{CCC} (Miron et al., 2012), a function combining the Sum of Absolute Differences with gradient C_{klaus} (Klaus et al., 2006), a function combining Absolute Differences with Census Transform $C_{ADCensus}$ (Mei et al., 2011) and one that combines Absolute Differences computed both on visible and gradient images with Census Tranform computed on gradient (C_{cstent}) (Stentoumis et al.).

The functions C_{SD}, C_{ZNCC} and C_{CT} were already compared by Hirschmuller and Scharstein (2009) on the Middlebury dataset composed of images with simulated or real radiometric distortions. We have chosen these functions as reference.

In the following, the functions presented are grouped into costs based on differences of intensities and costs based on CT.

3.1. Intensity differences based costs

$\boldsymbol{C_{AD}}, \boldsymbol{C_{SD}}$ & $\boldsymbol{C_{SAD}}$. One of the most popular cost matching function are the *absolute intensity differences (AD)* (see Eq. (1)) and *squared intensity differences (SD)* (see Eq. (2)).

Let p be a pixel in the left image with coordinates (x, y) and d the disparity value for which the cost of p is computed. Let $I_l(x, y)_i$ be the intensity value of pixel p in the left image on colour channel i, while $I_r(x, y-d)_i$ is the intensity value of the pixel given by coordinates $(x, y-d)$ in the right image. We consider n the number of colour channels ($n = 1$ for grey scale images and $n = 3$ for colour images).

$$C_{AD}(x,y,d) = \frac{1}{n}\sum_{i=\overline{1,n}} |I_l(x,y)_i - I_r(x,y-d)_i|, \tag{1}$$

$$C_{SD}(x,y,d) = \frac{1}{n}\sum_{i=\overline{1,n}} (I_l(x,y)_i - I_r(x,y-d)_i)^2. \tag{2}$$

If we consider $N(x, y)$ to be the neighbourhood of the pixel with coordinates (x, y), then the cost SAD on this neighbourhood is defined like in Eq. (3).

$$C_{SAD}(x,y,d) = \sum_{(a,b)\in N(x,y)} C_{AD}(a,b,d) \tag{3}$$

$\boldsymbol{C_{ZNCC}}$. Zero-mean normalised cross correlation (Eq. (4)) is a parametric window based matching function that provided one of the best results in the study performed by Szeliski et al. (2008) in presence of radiometric distortions.

$$C_{ZNCC}(x,y,d) = 1 - ZNCC(x,y,d) \tag{4}$$

where

$$ZNCC(x,y,d) = \frac{\sum_{(a,b)\in N_{(x,y)}} ZV(I_l,a,b)ZV(I_r,a,b-d)}{\sqrt{\sum_{(a,b)\in N_{(x,y)}} (ZV(I_l,a,b))^2 \sum_{(a,b)\in N_{(x,y)}} (ZV(I_r,a,b-d))^2}} \tag{5}$$

and

$$ZV(I,x,y) = I(x,y) - \bar{I}_{N(x,y)}(x,y), \tag{6}$$

where $\bar{I}_{N(x,y)}$ is the mean value computed in the neighbourhood $N(x, y)$.

$\boldsymbol{C_{klaus}}$. There exists several variations based on the costs previously described.[3] One of the top three algorithms on the Middlebury dataset (Klaus et al., 2006) proposes the combination of C_{SAD} with a gradient based measure C_{GRAD} (Eq. (7)). Both costs are computed in a neighbourhood $N(x, y)$ of 3×3 pixels and are weighted by w, which is computed by a grid search.

$$C_{klaus}(x,y,d) = (1-w) * C_{SAD}(x,y,d) + w * C_{GRAD}(x,y,d), \tag{7}$$

where

$$\begin{aligned} C_{GRAD}(x,y,d) = &\sum_{(a,b)\in N(x,y)} |\Delta_x I_l(a,b) - \Delta_x I_r(i,j-d)| \\ &+ \sum_{(a,b)\in N(x,y)} |\Delta_y I_l(a,b) - \Delta_y I_r(i,j-d)|, \end{aligned} \tag{8}$$

where Δ_x and Δ_y are the horizontal and vertical gradients of the image.

3.2. CT based cost functions

$\boldsymbol{C_{CT}}$. As demonstrated in Hirschmuller and Scharstein (2009), the Census Transform (CT) (Zabih and Woodfill, 1994) is one of the most robust cost function to radiometric changes. CT will basically replace all the intensity of pixels with a bitstring obtained by comparing the intensity of each pixel with the intensities of pixels in its vicinity. The *CT* cost is given by the Hamming distance (D_H) between two bit strings (Eq. (9)).

$$C_{CT}(x,y,d) = D_H(CT(x,y), CT(x,y-d)), \tag{9}$$

where *CT* is the bit string build like in Eq. (10).

[2] http://vision.middlebury.edu/stereo/

[3] If the authors did not name the proposed cost functions we are going to use the first name on the article to name the cost