Pattern Recognition 47 (2014) 1926–1940

Contents lists available at ScienceDirect

Pattern Recognition

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

A combined post-filtering method to improve accuracy of variational optical flow estimation

Zhigang Tu, Nico van der Aa, Coert Van Gemeren, Remco C. Veltkamp *

Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

ARTICLE INFO

Article history:
Received 11 July 2013
Received in revised form
17 October 2013
Accepted 25 November 2013
Available online 11 December 2013

Keywords:
Optical flow
Combined post-filtering (CPF)
Multi-scale nonlinear 3D structure tensor
Hybrid GBF and Gaussian Filter smoothing (HGBGF)
Spatial-scale gradient signal-to-noise ratio (SNR)

ABSTRACT

We present a novel combined post-filtering (CPF) method to improve the accuracy of optical flow estimation. Its attractive advantages are that outliers reduction is attained while discontinuities are well preserved, and occlusions are partially handled. Major contributions are the following: First, the structure tensor (ST) based edge detection is introduced to extract flow edges. Moreover, we improve the detection performance by extending the traditional 2D spatial edge detector into spatial-scale 3D space, and also using a gradient bilateral filter (GBF) to replace the linear Gaussian filter to construct a multi-scale nonlinear ST. GBF is useful to preserve discontinuity but it is computationally expensive. A hybrid GBF and Gaussian filter (HGBGF) approach is proposed by means of a spatial-scale gradient signal-to-noise ratio (SNR) measure to solve the low efficiency issue. Additionally, a piecewise occlusion detection method is used to extract occlusions. Second, we apply a CPF method, which uses a weighted median filter (WMF), a bilateral filter (BF) and a fast median filter (MF), to post-smooth the detected edges and occlusions, and the other flat regions of the flow field, respectively. Benchmark tests on both synthetic and real sequences demonstrate the effectiveness of our method.

1. Introduction

Motion is an intrinsic characteristic of the world [1], providing essential information that can be used in a wide variety of image processing and visual tasks, such as 3D-reconstruction, segmentation, tracking and video compression. One of the most successful motion estimation approaches is the variational optical flow method [2,3], due to two inherent advantages, i.e. its comprehensive functional form and an efficient numerical optimization. The variational optical flow method was introduced by Horn and Schnuck (HS) [4]. It combines a local, gradient-based data matching term with a global smoothness term. The data term assumes that each pixel's brightness remains invariant during a short time. The smoothness term regularizes each pixel's flow by its neighbors' flow. It assumes that the flow vector varies smoothly almost everywhere over the flow field. In practice, however, these two basic constraints are seriously violated. Various extensions and improvements have been proposed during the past 30 years in order to overcome the drawbacks of the original HS model.

These variations can generally be classified according to the following three aspects: (1) Modification of the variational formulation, such as improvements of the data term to make the algorithm more robust under illumination changes [5], invariant under different types of motion [6], and more resistant to noise [7] and outliers [6,7], and large displacements[8]. Modifications of the regularizer's capability to handle motion discontinuity [9,10]. Selection of the optimal weighting parameter λ to obtain a better balance between the data term and the regularization term [10,11]. (2) Pre-processing of the input frames to reduce violations, such as noise suppression of the frames to remove, for example, high frequencies that might have a negative influence on the result. Commonly used filtering methods include Gaussian filter [4], PDE filter [12] and non-local filter [13]. Most of these methods do not only reduce noise, but also enhance important structures of the frames [14]. (3) Post-processing of the flow field to improve the accuracy, e.g. usage of available filters for smoothing, such as, Kalman filter [15], median filter (MF) [14,16], and bilateral filter (BF) [17,18].

Wedel et al. [16] successfully introduced a MF to remove the flow noise. However, his MF approach over-smoothes the edges. Sun et al. [14] proposed a modified WMF method to prevent this kind of over-smoothing, and saved the computational time by solely smoothing the detected motion boundaries with the Sobel edge detector. However, this method still has some drawbacks. First, the Sobel detector often performs poorly in extracting flow boundaries. Second, wrong flow components in the MF [19] window can cause serious errors. Surprisingly, although smoothing the flow field

* Corresponding author. Tel.: +31 30 253 4091.
E-mail addresses: Z.Tu@uu.nl (Z. Tu), N.vanderAa@noldus.nl (N. van der Aa), C.J.Vangemeren@uu.nl (C. Van Gemeren), R.C.Veltkamp@uu.nl (R.C. Veltkamp).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.026

boundaries is a reasonable way to improve accuracy and efficiency, few efforts have been devoted so far to analysis of the connection between the smoothing performance and the extraction of flow features (e.g. edges and occlusions). To the best of our knowledge, this work is the first systematical analysis of the importance of the above mentioned connection.

We present a novel 3D nonlinear ST based Harris edge detector to identify flow edges, and apply a piecewise occlusion detection approach to detect flow occlusions. The ST has first been proposed by Förstner and Gülch [20]. Since it represents the first order derivative information of an image, it can be used as a local geometry indicator to analyze the geometric structure of a scalar-valued data set (e.g. an image) or a vector-valued data set (e.g. the flow field). Compared to traditional derivative-based methods, the ST has two outstanding characteristics due to two Gaussian smoothing operations: (1) smoothing the data set yields robustness under noise by introducing an integration scale and (2) integrating local structure information (e.g. orientation) from a neighborhood makes ST able to distinguish features [21].

Since Gaussian smoothing is isotropic, it has some disadvantages: (1) detailed and weak features, such as some textures are smoothed out, (2) distinctive discontinuities such as edges are blurred and, and (3) points belonging to different regions, such as occluded points and non-occluded points, would be roughly composed. These disadvantages are caused by the fact that the Gaussian filter is fixed in both size and shape, and it cannot adapt to the local structures. Therefore, the Gaussian filter based linear ST cannot detect edges accurately. For instance, the identified edges are often wider than the real edges or discontinuous.

Different anisotropic filtering methods have been proposed to replace the linear Gaussian filter to construct a nonlinear ST, like anisotropic diffusion [21], BF [17,18,22] and mean shift filtering [23]. They can adapt to local structures, avoiding smoothness across discontinuities and preserving useful information. The BF, which extends the concept of Gaussian filter by adding a Gaussian weighting function that depends on the difference between pixel intensities, is most attractive [24] due to its inherent advantages: (1) it is non-iterative, which makes it overcomes the instability of the iterative method – since small errors in derivatives will be magnified after each iteration, (2) only two parameters are needed and these parameters have explicit geometrical and graphical meaning, therefore, they are easy to be constructed and implemented, and (3) as illustrated in [17], the BF can handle occlusion. In this work, the BF is used to replace the Gaussian filter to construct a nonlinear ST, and also it is used to replace the MF to smooth the occlusions of the flow field.

The BF assigns higher weights to pixels with smaller spatial and/or color distances computed with respect to the central pixel. In this way, smoothing is implemented adapt to local structures. To distinguish trivial structures from true corners, Zhang et al. [22] introduced a GBF which uses both spatial and gradient distances to smooth the 2D spatial ST. In this work, we introduce the GBF to construct a nonlinear 3D ST to detect edges.

A direct implementation of the BF is computationally expensive. It requires $O(\sigma_s^2)$ operations per pixel. Especially when the data set is large, it is too slow to be executed in real-time. Porikli [25] proposed a fast O(1) BF using Taylor polynomials to approximate the standard Gaussian BF. However, the Taylor polynomials provide only good approximations of the range Gaussian function just locally around the origin. Different from the O(1) BF method, in this work, we apply a fast spatial-scale gradient based SNR segmentation and a hybrid smoothing approach – HGBGF to treat the low efficiency of the BF.

For a vector-valued data set, the primary derivative errors are concentrated at discontinuities [26]. Because the Gaussian filter is good enough to smooth flat regions and the BF is better at smoothing discontinuities, combing the advantages of the two filters can preserve edges, tackle occlusions and also reduce time consumption. Therefore, we present a novel spatial-scale gradient SNR measure to extract discontinuities. Then, we apply the BF and the Gaussian filter to smooth the ST elements in separate regions, respectively.

Multi-scale is an intrinsic property of the signal structure in nature [27]. Liu et al. [28] gave a definition of edge scale and pointed out that there exists an optimal scale of the edge – the optimal scale is a parameter to indicate at which resolution(s) an edge is most salient for a human. We extend the traditional spatial ST based edge detector into spatial-scale space by adding scale information. Integrating the HGBGF into our spatial-scale 3D ST, a local pattern adaptive framework is constructed, resulting in better detection of flow field edges.

Using a suitable filter, such as the MF, to post-filter the intermediate flow field during incremental estimation and warping is an effective way to remove outliers and a key technique to recent performance gains [14,16]. However, the MF is not good at handling occlusions. In contrast, the BF has been successfully applied to treat occlusion [17,18]. In this paper, we combine the advantages of the WMF [14] and the BF[16], and propose a Combined Post-Filtering (CPF) method to smooth the classified flow field regions.

The paper is organized as follows: Section 2 describes the proposed "Classic+CPF" optical flow algorithm. In Section 3, a nonlinear 3D spatial-scale Harris edge detector to detect flow edges, and a piecewise occlusion detection approach to extract occlusions are introduced. A CPF method for post-filtering different regions of the flow field with different filters is proposed in Section 4. In Section 5, experiments and evaluations are conducted to the proposed algorithm. The paper is concluded in Section 6, which includes possibilities for future development.

2. "Classic+CPF" optical flow algorithm

Based on the brightness constancy assumption (data term), and combined with a global smoothness constraint (smoothness term), Horn and Schunck [4] proposed the variational optical flow method for motion estimation between two successive frames I_1, I_2:

$$E(u,v) = \int_\Omega \underbrace{(I_2(x+u,y+v,t+dt)-I_1(x,y,t))^2}_{\text{data term}}\, d\Omega + \lambda \int_\Omega \underbrace{(|\nabla u|^2+|\nabla v|^2)}_{\text{smoothness term}}\, d\Omega \tag{1}$$

where $(u,v)=(dx/dt, dy/dt)$ is the displacement vector field. It is a 2D projection of the real 3D motion in the world.

One state-of-the-art variational method is the TV-L1 non-local algorithm – "Classic+NL" [14], which incorporates the WMF during optimization to smooth the flow field. Due to the WMF, the accuracy is significantly improved. However, the WMF is poor to handle occlusions (see Section 4.1). To overcome this problem, we classify the optical flow field into three parts – edge regions, occlusions and flat regions. As illustrated in Section 1, we combine the advantages of WMF and BF, and use a CPF method to smooth flow edges, occlusions as well as flat regions with three different filters. Based on the baseline algorithm of "Classic+NL" [14], a "Classic+CPF" algorithm is proposed:

$$\begin{aligned} E(u,v,\overline{u},\overline{v}) = \sum_{i,j}\{&\rho_D(I_1(i,j)-I_2(i+u,j+v))+\lambda_1(\rho_S(|u_x|)+\rho_S(|u_y|) \\ &+\rho_S(|v_x|)+\rho_S(|v_y|))\}+\lambda_2(||u-\overline{u}||^2+||v-\overline{v}||^2) \\ &+\underbrace{\sum_{i_E,j_E}\sum_{i'_E,j'_E \in N_{i_E,j_E}} w_{i_E,j_E,i'_E,j'_{EE}}(|\overline{u}_{i,j}-\overline{u}_{i',j'}|+|\overline{v}_{i,j}-\overline{v}_{i',j'}|)_{|i_E,j_E,i'_E,j'_E}}_{\text{edge regions}\rightarrow\text{weighted median Filter}} \end{aligned}$$