Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Combination of spatio-temporal and transform domain for sparse occlusion estimation by optical flow

Pengguang Chen [a,*], Xingming Zhang [a], Pong C. Yuen [b], Aihua Mao [a]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[b] Department of Computer Science, Hong Kong Baptist University, Hong Kong, China

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form
6 May 2016
Accepted 11 June 2016
Communicated by Xiaoqin Zhang

MSC:
00-01
99-00

Keywords:
Occlusion estimation
Optical flow
Sparse representation
Dictionary learning

ABSTRACT

Lack of information in occluded regions leads to ambiguity inherent, which is a big challenge for motion estimation. Recently, the sparse model has been widely used since the essential content of the motion field could be effectively preserved with sparse representation. The methods exploiting sparsity acquire representations either directly in the spatio-temporal domain or indirectly in the transform domain. Usually, the sparse model with sparsifying transform is based on patches and thus is more robust agints noise, while the sparse model without sparsifying transform can directly work for an overall image treatment. Aiming at tackling the motion ambiguity efficiently, this paper employs a distinct sparse representation model into a variational framework for estimating occlusion with optical flow. In order to deal with dictionary learning which is computationally expensive and requires a preprocess for extending the sparsifying transform model for arbitrary image sizes, we present a new unified framework to directly generate an overall dictionary via the sparse model without sparsifying transform, and then optimize for small size dictionaries over corresponding patches with the overall dictionary. Our framework is based on the Stein–Weiss analysis function acting as a novel regulariser and a sparsifying transform function respectively in variational and sparsity models. Experiments show that the proposed method outperforms the existing estimation methods of jointing occlusion and optical flow.

1. Introduction

Occlusion is a crucial factor in digital image processing and provides important information in computer vision for perception, segmentation and recognition [1,2]. In this paper, we define the occluded region as the disappearance region in one of a pair of images. Occlusion estimation is closely related to optical flow, as the occluded region is only visible in one image while optical flow refers to the dense field of visual correspondence between two images. Optical flow is one of the most popular methods for occlusion estimation. However, occlusion estimation with optical flow is a significantly challenging problem, because the lack of information in occluded regions leads to the ambiguity inherent for the motion estimation [3–5]. This paper proposes an approach to occlusion estimation which tackles the motion ambiguity by modeling the occlusion with sparse representations in the spatio-temporal domain and transform domain in the optical flow framework.

In the optical flow community, most algorithms [6–9] did not take occlusions into account, which usually result in discontinuities in the motion field [10]. Therefore, jointing occlusion detection and optical flow [5,10,11,12,13,14] has been prevalent for better performance. The existing estimation algorithms of jointing detecting occlusion and optical flow can be roughly divided into two categories: implicit and explicit modeling of occlusion. The implicit modeling of occlusion infers occluded regions by computing motion bidirectionally to find out regions which are inconsistent. Methods such as symmetric estimation [11], motion symmetry with adaptive weighting of the data fidelity [12] have been proposed. However, such methods detect occlusions as a postprocessing step [10] and suffer from no correspondence mappings in occluded regions [1].

Therefore, explicit modeling of occlusion has more advantages in obtaining reliable occluded features than implicit modeling [15]. Typically, the classification-based method is a common choices, it combines different types of features for occluded pixels, such as statistical [16] and appearance features [5], to represent and detect occlusion. But such method may lead to request of complex optimization for the label problem of large numbers. Hence, most current methods concentrate on specific optical flow visibility constraint to simplify this problem for efficient performance

* Corresponding author.
E-mail addresses: chen.pg@mail.scut.edu.cn (P. Chen), cszxm@scut.edu.cn (X. Zhang), pcyuen@comp.hkbu.edu.hk (P.C. Yuen), ahmao@scut.edu.cn (A. Mao).

http://dx.doi.org/10.1016/j.neucom.2016.06.025
0925-2312/

[4,1,13,14]. However, directly applying optical flow visibility constraint for occlusion modeling may not achieve convincing performance, since they usually set a threshold for the residual by artificial setting [13] or a weight function [14], which may be invalid in the complex motion.

Moreover, since the essential content of the motion field, such as occlusion, can be designed to be efficiently preserved and represented with sparse representation [17], the fidelity evaluation of the occlusion modeling is depended upon reasonable regularized optimization for ℓ_1[4,1], which is helpful to improve performance. As a result, jointing sparse representation for modeling occlusion with optical flow constraint for more significantly representing and reliably measuring is required.

Recently, the application of sparse representation [4,1,18–20] has received a lot of interest as a promising method for motion estimation in optical flow models. However, the sophistication of sparse representation in the model integration is still limited. On one hand, sparse models with sparsifying transform [17,19,20,21], such as gradient, DCT, Fourier and wavelet transform, are based on image patches and more robust against noise. However such method usually cannot be directly deployed on larger image blocks due to the computational complexity, such as [20]. To handle the images with arbitrary sizes, a predefined or prelearned dictionary is often required for significantly representing the overall image during the optimization [21,17,19], where the dictionary learning process is computationally expensive. On the other hand, sparse models without sparsifying transform work for an overall image treatment [4,1,18]. However, such method may degrade the performance in dealing with complex motion, since the representation is assumed directly on the image which is sensitive to image noise and outliers.

Concerning these issues, we focus on explicit occlusion modeling with optical flow, which is capable of sparse representations both in the spatio-temporal domain and the transform domain for significantly representing occlusions. Moreover, for better computational efficiency during dictionary learning, we present an unified framework to directly construct an overall dictionary for the sparsifying transform model via the sparse model without transform. Our framework is based on the Stein–Weiss analysis function [22] acting as a novel regulariser and a sparsifying transform function respectively in variational and sparsity models.

The contributions of this paper are in the following ways:

(1) We propose to apply the Stein–Weiss analysis function into the optical flow model. Since the Stein–Weiss analysis function is smoothly harmonic and inverse, we use it rather than the gradient operator of both sparse models of [1] and [20] as a novel regulariser and a sparsifying transform function respectively. We make use of the Stein–Weiss analysis function to play different roles in variational and sparsity models that facilitates to construct an unified framework for the benefit of guaranteed global and fast optimality.

(2) By extending the idea of Ayvaci et al. [4,1] that detects occlusion by identifying the sparse flow-field mismatch, we represent the sparse mismatch by a new combination of the spatio-temporal domain with the transform domain. The proposed method has the advantages of both information from sparse representations in the spatio-temporal domain and the Stein–Weiss analysis function transform domain. Compared with [4,1], our proposed approach improves performance significantly both on occlusion detection and optical flow estimation.

(3) We present an unified framework for solving the minimization problem accompanying with the sparsifying transform model. The key idea of our scheme is that, firstly, an overall dictionary for the sparsifying transform model is directly generated via the sparse model without transform. Then, small size dictionaries are computed with the overall dictionary. Hence, the proposed method does not require a predefined or prelearned dictionary. In comparison with [18,19,20], our method does not need of a predefined or prelearned dictionary and can handle images with arbitrary sizes.

2. Related work

Occlusion is of universal attention for optical flow computation. In the early work of optical flow detection, Horn and Schunck (HS) defined a quadratic formulation to handle data outliers and preserve flow inconsistency [6]. Subsequently, lot of functions compromising between the robustness and flow accuracy have been employed, such as Charbonnier's penalty [8] and the magnitude of the derivative of median flow [7]. Since inconsistent motions often occur at object boundaries [19], total variation (TV) norm of the residual is a common method [9]. Several work exploited higher order terms, structure–texture decomposition, median filtering or dynamical weighting to reduce the influence of outliers and over-smoothing. However, such methods without indicating occlusion may lead to inconsistency-preserving on motion fields [10].

To obtain better performance, methods that jointing optical flow estimation and occlusion detection have been proposed. The popular one is cross-checking by computing optical flow bidirectionally forward and backwards [11]. Furthermore, methods [12,10] add measurement of the data confidence to improve accuracy, by exploiting with different monotonically decreasing functions. However, such motion symmetry models suffer from the lack of correspondences in occluded regions, since occlusion is undefined. An alternative approach is the classification of occlusion feature. In [16,5], authors respectively use statistics and appearance features of occlusion for descriptor matching, which are computationally expensive. Several algorithms have been proposed to reduce the computational complexity of matching problem. Yang et al. [14] use visibility registration energy for representing occlusion. Sun et al. [13] proposed a probabilistic function to enhance robustness. However, such methods are based on thresholds for the residual, which are difficult to improve performance in handling complex motion.

Efforts have also been put into integrating sparse representation into optical flow models. Under specific transform domain, Shen and Wu [20] first explored the sparsity for optical flow estimation, where they use ℓ_1 norm of the flow gradient field and transform coefficient for sparsity constraint. Due to the implementation complexity, in practice the proposal is limited in handling small size images. Jia et al. [19] tackled ℓ_1 regularized problems for large image sizes via a sparse representation over the learned over-complete flow dictionary. Learning dictionary which facilitates a sparse representation is a common approach in compressive sensing community [17]. However, such method increases the computational complexity. An alternate approach is the one without sparsifying transform. Ayvaci et al. [4,1] integrated ℓ_1 regularized optimization that separates the matching errors into optical flow model. Chen et al. [18] concerned the regularization error in the model. However, such method defines sparse representation directly on the motion field and thus suffers from lack of robustness.

3. Combining sparse representations of spatio-temporal and transform domain

The proposed model combines sparse representations in the spatio-temporal domain and the transform domain to estimate