Pattern Recognition 000 (2017) 1–12

Contents lists available at ScienceDirect

Pattern Recognition

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

Learn to model blurry motion via directional similarity and filtering

Wenbin Li[a,b,*], Da Chen[c], Zhihan Lv[b,*], Yan Yan[d], Darren Cosker[e]

[a] Department of Computing, Imperial College London, UK
[b] Department of Computer Science, University College London, UK
[c] Department of Computer Science, University of Bath, UK
[d] Department of Information Engineering and Computer Science (DISI), University of Trento, Italy
[e] Centre for the Analysis of Motion, Entertainment Research and Applications (CAMERA), University of Bath, UK

A R T I C L E I N F O

Article history:
Received 21 July 2016
Revised 7 February 2017
Accepted 17 April 2017
Available online xxx

Keywords:
Optical flow
Convolutional Neural Network (CNN)
Video/image deblurring
Directional filtering

A B S T R A C T

It is difficult to recover the motion field from a real-world footage given a mixture of camera shake and other photometric effects. In this paper we propose a hybrid framework by interleaving a Convolutional Neural Network (CNN) and a traditional optical flow energy. We first conduct a CNN architecture using a novel learnable directional filtering layer. Such layer encodes the angle and distance similarity matrix between blur and camera motion, which is able to enhance the blur features of the camera-shake footages. The proposed CNNs are then integrated into an iterative optical flow framework, which enable the capability of modeling and solving both the blind deconvolution and the optical flow estimation problems simultaneously. Our framework is trained end-to-end on a synthetic dataset and yields competitive precision and performance against the state-of-the-art approaches.

1. Introduction

In the image space, the information observed by the dynamical behavior of the object of interest or by the motion of the camera itself is a decisive interpretation for representing natural phenomena. Dense motion, in particular optical flow estimation between a consecutive image pair is the most low-level characterization of such information, which is supposed to estimate a dense field corresponding to the displacement of each pixel. It has become one of the most active fields of computer vision because such characterizations can be extremely embedded into a large number of other higher-level computer vision fields and application domains. Indeed, one can be interested in tracking [1–3], 3D reconstruction [4], segmentation, as well as the general virtual reality, augmented reality and post-production [5,6].

A typical pipeline of optical flow estimation has been lied on solving a brightness energy with the assistance of patch detection, matching, constrained optimization and interpolation. For many state-of-the-art approaches – even the precision has reached a reasonable level – the related applications are still limited by the difficult photometric effects and low performance in runtime. In the recent years, the deep *Convolutional Neural Networks* (CNNs) grows rapidly, which makes a step forward to provide hidden features and end-to-end knowledge representation for many precentral issues e.g. motion and texture style *etc*. Such knowledge representation is able to improve the robustness and yields a rapid fashion in the typical optical flow pipeline.

Camera-shake blur is a common photometric effect in the real-world footage, which is often caused by the fast camera motion under a low light condition. Such effect may lead to an invariant blur information for each of the pixel, and may bring extra difficulties into typical optical flow estimation because the basic brightness constancy [7] is violated. However, the blur from a daily video footage (24 FPS) can be directionally characterized [8]. This observation enables an extra prior to enhance the camera-shake deblurring [9] and further recover precise optical flow from a blurry images. Such directional prior needs a strict pre-knowledge on the motion direction of the camera which can be obtained by an external sensor [8].

1.1. Contributions

In this paper, we study the issue of recovering accuracy optical flow from frames of a real-world video footage given a camera-shake blur. The main idea is to learn directional filters, encoded the angle and distance similarity between blur and camera motion. Such filters are further applied to enhance the optical flow estimation. Our proposed method only relies on the input images, and does not need any other information e.g. ground truth camera motion and blur prior.

* Corresponding authors.
E-mail address: wenbin.li@imperial.ac.uk (W. Li).

http://dx.doi.org/10.1016/j.patcog.2017.04.020

In overview, we propose a novel hybrid approach: **(1)** we conduct a CNN architecture using a learnable directional filtering layer. Our network is able to extract the blur&latent features from a blurry image, and further recover the blur kernel within an iterative deconvolutional fashion (Section 4); **(2)** we integrate our network into a variational optical flow energy, further optimized within a hybrid coarse-to-fine framework (Section 5).

In the evaluation (Section 6), we quantitatively compare our method to four baselines on the synthetic *Ground Truth* (GT) sequences. Those baselines include two blur-oriented optical flow approaches and two other publicly available state-of-the-art methods. We also give quality comparison given real-world blurry footages.

2. Related work

2.1. Image deblurring

Image blur is a common photometric effect for the daily capture. It is often caused by fast camera movement under a low light condition. Such global blur can be formulated as follows:

$$I = k * \ell + n$$

where an observed blurred image I can be represented as a combination of spatial noise n along with a convolution between the latent sharp image ℓ and a spatial-invariant blur kernel w.r.t. *Point Spread Function*. To solve the k and ℓ, a blind deconvolution is normally performed on I:

$$\operatorname*{argmin}_{k,\ell} \{\|I - k * \ell\| + \rho(k)\}$$

where ρ represents a regularization that penalizes spatial smoothness with a sparsity prior [10]. To solve this ill-posed problem, many approaches rely on additional priors regarding to properties of observed images [11–18]. Pan et al. [13], for example, propose a blind deconvolution method by taking advantages from the dark channel [19] regarding to the observation that the dark pixels in the observed image are normally averaged with neighboring pixels along the blur. Krishnan et al. [11] introduce a novel scale-invariant regularizer to generate a more stable kernel by fixing the attenuation of high frequencies.

By taking into account the efficient inference, several algorithms [9,10,20,21] are also proposed to solve the deblurring problem. Cho and Lee [10] adopts a predicted edge map as a prior and solve the blind deconvolution energy within a course-to-fine framework. Xu et al. [20], however, discuss a key observation that salient edges do not always help with blur kernel searching. These edges can greatly increase the blur ambiguity in many common scenes. Hence, instead of the use of edge map, they propose an automatic gradient selection scheme to eliminate the “noisy” edges for kernel initialization. Furthermore, Zhong et al. [9] introduce an approach to reduce the noise using a pre-filtering process. Such process preserves the useful image information by reducing the noise along a specific direction.

Both natural image properties based and efficient inference based methods mentioned above are able to provide highly accurate deblurring result for general invariant camera-shake blur. However, these methods often show difficulties given the cases under variant blur. A handful of approaches are proposed to solve such a problem [22–26]. Gupta et al. [22] propose a *Motion Density Function* to represent the camera motion which is further adopted to recover the spatially varying blur kernel. Hu et al. [25] consider the various depth information of the scene while most of the deblurring methods apply a constant depth for simplicity. They apply an unified layer-based model to jointly estimate the depth and deblurring result from the underlying geometric relationship caused by camera motions.

Since all the methods mentioned above have the specialty along with their limitation, there is no general solution for images blurred by mixed sources, with regard to mixture of fast camera and object movement and scene depth variance. In this case, the image blur is hard to represent by a global model. With the development of Convolutional Neural Network (CNN), some CNN based deblurring methods are proposed to solve such problem. Hradiš et al. [27] apply a CNN to restore the blurred text documents which is restricted by highly structured data. Xu et al. [28] propose a more general deblurring method. They design a neural network which is guided by traditional deconvolution schemes.

Those mentioned above usually involve a single blurred image as input. There are some hardware assisted methods which are supposed to improve the precision and performance of deblurring [29–32]. Levin et al. [29] propose a uniform method using the known camera arc motion. Such uniformly deblurred image can be estimated by controlling the camera movement along with a parabolic arc. As an extension of this work, Joshi et al. [30] propose to estimate the acceleration and angular velocity of camera by a inertial sensor, *i.e.* gyroscopes and accelerometers. Instead of the highly accurate sensor, Hu et al. [32] introduce a deblurring approach using the smartphone inertial sensors. These methods with extra camera motion information often yield higher performance comparing to those methods only rely on single blurred image as input. However, these methods require complex camera setup and precise calibration.

2.2. Optical flow

Dense motion estimation problem, in particular optical flow, has been widely studied as it can be adopted to many computer vision applications, e.g. video segmentation [33] and recognition [34] *etc.* Many estimation methods have obtained impressive performance in terms of reliability and accuracy showed on the Middlebury [35] and Sintel [36] benchmark. Most of this works are based on the pioneering optical flow method proposed by Horn and Schunck [7]. They combine a data term and a smoothness term into an energy function where the former term assumes the certain constancy of the image feature – typically according to Brightness Constancy Constraint (BCC) – and the latter term controls how the motion field is varied (such as the Motion Smoothness Constraint). This energy function is then optimized across the entire image to reach the global motion field. This original formula is generally applicable but often limited by many challenges such as large displacement, non-rigid motion, motion boundaries discontinuities, motion blur *etc.* [36]. Numbers of extensive works have been proposed to conquer these challenges by introducing additional constraints and more advanced optimization procedure [37–46]. Brox et al. [38] bring a gradient constancy assumption into the data term in order to reduce the dependency of BCC, and bring a discontinuity-preserving spatio-temporal smoothness constraint to deal with motion discontinuities. Xu et al. [40] propose a novel extended coarse-to-fine (EC2F) refinement framework by taking advantages of feature matching technique. Li et al. [41] propose to apply Laplacian mesh energy to adapt the non-rigid deformation in the scenes.

Moreover, some neural network based methods are recently popular. Revaud et al. [47] propose a edge preserving interpolation based on a sparse deep convolutional matching result. The sparse-to-dense interpolation result is then apply to initialize the optimization process for obtaining the final motion field. However, this method strongly relies on the quality of sparse matching where parameters are set manually. Dosovitskiy et al. [48] propose an automatic approach for matching and interpolation. Guiding by a correlation layer, their network can better predict the flow to initialize the refinement. Furthermore, Teney and Hebert [49] introduce

Please cite this article as: W. Li et al., Learn to model blurry motion via directional similarity and filtering, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.04.020