Infrared Physics & Technology 88 (2018) 228–238

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Regular article

A rapid and accurate infrared image super-resolution method based on zoom mechanism

Chao Sun[a,*], Junwei Lv[a], Jianwei Li[b], Rongchao Qiu[a]

[a] Department of Control Engineering, Naval Aeronautical University, Yantai 264001, China
[b] Department of Electronic and Information Engineering, Naval Aeronautical University, Yantai 264001, China

HIGHLIGHTS

- An effective infrared image super resolution method is proposed.
- We design a novel network architecture and introduce the zoom mechanism.
- The proposed method owns a rapid running speed and can process videos in real-time.
- We apply the method from natural images to infrared images by transfer learning.

ARTICLE INFO

Article history:
Received 5 September 2017
Revised 23 November 2017
Accepted 24 November 2017
Available online 24 November 2017

Keywords:
Super-resolution
Deep learning
Convolutional neural network
Zoom mechanism
Transfer learning

ABSTRACT

In view that the existing methods for infrared image super-resolution could not have good performance both in speed and accuracy simultaneously, with the advantage of deep learning method, this paper presents a rapid and accurate infrared image super-resolution method based on zoom mechanism. In the meantime, we design a novel network architecture. First, we take low-resolution images as network inputs directly, and employ a convolution layer to extract and represent features. Then we introduce the combination of a deconvolution layer and a pooling layer into the network, what we name as zoom mechanism. The zoom mechanism could not only enlarge and shrink obtained feature maps successively to extract features that are more sensitive to the results, but also implement the nonlinear mapping in a more effective way than other methods. Moreover, we depend on a sub-pixel convolution layer to realize the features map and images fusion in a single step. Subsequently, a plenty of natural images are utilized as the auxiliary training samples to obtain the pre-trained network. Finally, the pre-trained network is fine-tuned again by adding a few infrared images as objective training samples, which could make the network more suitable for infrared image super-resolution. The proposed approach is tested on both of natural and infrared images. It can achieve more satisfactory results both in objective criterion and subjective perspective compared to other state-of-the-art methods. In addition, it can process more than 24 images in size of 320×240 pixel per second. The experimental results show that the proposed method can not only generate images in higher quality, but also satisfy the requirement of real-time video super-resolution.

1. Introduction

Infrared imaging has several outstanding capabilities, such as passivity, strong anti-interference, adaptability to environment and so on. For these advantages, infrared imaging technology plays a great role in the military, video surveillance, night navigation and other areas. To some extent, the amount of information contained in the infrared image depends on the resolution. Especially in the military field, the infrared image in higher resolution means a more powerful ability to detect, identify and attack military targets, which is of great help to enhance combat effectiveness. So in order to improve its visual effects and make it more in line with the requirements of human or computer processing, it is an urgent task to improve the resolution of the infrared image at present. Due to the limits of the fabrication technique and material properties, it is difficult to rely solely on manufacturing high density infrared focal plane array (IRFPA) to improve the resolution of the infrared

* Corresponding author.
E-mail address: lemony1314@163.com (C. Sun).

https://doi.org/10.1016/j.infrared.2017.11.033
1350-4495/

image. Therefore, it is particularly significant to achieve the goal in the way of image processing by software technology.

Single image super-resolution (SR) aims at recovering a high-resolution (HR) image from a given low-resolution (LR) image [1]. The existing SR methods can be mainly divided into three categories, namely, interpolation-based, reconstruction-based and learning-based methods. The interpolation-based method is the most basic and simplest means, but it is difficult to reproduce the details of the image effectively, and the generated image is relatively vague. The reconstruction-based method utilizes the prior knowledge and recovers the HR image on the basis of several existing degradation models such as down-sampling, motion blurring, optical distortion. In recent years, the methods based on example-learning have attracted much attention. This kind of algorithm tries to reconstruct the missing details via a great deal of LR/HR example-pairs. It contains neighbor embedding (NE) method [2], anchored neighborhood regression (ANR) method [3], adjusted anchored neighborhood regression (A+) method [4] and the following. Yang et al. [5,6] presented an image super-resolution method based on sparse representation. The method makes use of sparse-code (SC) to express HR and LR image patches as HR and LR dictionaries and establishes the mapping relationships between the LR/HR dictionaries. Besides, with the development of deep learning, SR methods based on deep learning have achieved many impressive results. Compared with the traditional methods, the methods based on deep learning can generate images in higher quality and have better performance on SR. Dong et al. [7] put forward an algorithm for SR using convolutional neural networks (SRCNN). Kim et al. [8] were enlightened by the Res-Net and then introduced a deeper network for super-resolution (VDSR). Moreover, they constructed a network suitable for different upscale factors, which could reduce the network burden and accelerate the learning rate. Kim et al. [9] then augmented the receptive field of network by more convolution layers and introduced recurrent neural network (DRCN) to avoid more parameters. These two methods both adopted very deep network and achieved excellent performance while at the cost of computational time. Lei et al. [10] proposed a method named local-global-combined networks (LGCNet) by leveraging the multi-level data representation ability of deep learning.

However, the above methods based on deep learning owned vast parameters and heavy work of calculation, which resulted in a slow computing speed and inability to recover SR video in real-time. For this reason, Dong et al. [11] proposed a fast super-resolution by CNN method (FSRCNN) on the basis of SRCNN. In the meantime, Dong accordingly proposed a simplified method named FSRCNN-s. Shi et al. [12] proposed a method named ESPCN to obtain SR image by rearranging the last feature maps, which considerably accelerated the computing speed. However, compared to VDSR and DRCN, the performance of FSRCNN and ESPCN degraded while reducing parameters and simplifying network.

Moreover, infrared image SR based on deep learning is also faced with great difficulties compared with natural image SR. Training by deep learning requires a large number of images of different scenes or targets as samples. But it is difficult in reality to collect vast infrared images in high quality coming up to the above requirements. With the continuous development of deep learning, the transfer learning has also attracted a lot of attention. The core of transfer learning is to solve different problems in related areas by utilizing the existing knowledge. Xu [13] successfully recovered SAR images based on natural images by transfer learning. Nima [14] analyzed medical images through pre-trained convolution neural network and acquired better results than re-training network. Yanai et al. [15] improved the accuracy of recognition by fine-tuning the existing network when trying to identify food images.

Therefore, in order to seek a method that can not only solve the problem of insufficient samples of HR infrared images, but also recover images effectively in real-time, this paper proposes a rapid and accurate infrared image super-resolution method based on zoom mechanism (SRZM). This method has more excellent performance on SR compared to other real-time super-resolution methods. Besides, it does a great job both in accuracy and speed in comparison with other SR methods based on CNN, second only to VDSR and DRCN.

2. Related work

The existing SR methods based on deep learning are mainly divided into four steps. ① Build the essential external image database for training; ② Structure the appropriate network architecture. ③ Utilize the database to train and optimize the network, achieving the representation of features and prior knowledge of images. ④ Take the LR image as the input of optimized network and finally obtain the SR image by fusing. Among these steps, the structure and training of the network are quite crucial steps. Most of the network architectures adopt convolution neural network (CNN) model. However, the existing methods still have the following four shortcomings when applied in infrared image super-resolution:

(1). The low-resolution image is usually upscaled to the desired size by bicubic interpolation before reconstruction and then taken as the input of network. Such as SRCNN [7], VDSR [8], DRCN [9] methods shown in Fig. 1(a)–(c), these methods initially implement a predefined up-sampling operation at the beginning of the network. It increases the computational burden of the methods in multiple in comparison with those taking LR image as input directly. For example, when the upscale factor is f, the network parameters of those taking the upscaled image as input increase by f^2-1 times compared to taking LR image as input. Besides, the time spent in training or testing will also increase by the way.

(2). When the existing SR methods work, the size of the image in the process of the network can generally be divided into two cases. ① Remain unchanged in general: take the upscaled image as the network input, and obtain SR image in the size as same as input after several convolution layers, as shown in Fig. 1(a)–(c). ② Magnify in multiple at the end of network: take LR image directly as input, and pass through a serial of convolution layers in the middle of network, obtain the SR image in the required size after upscaled by some way at the last layer, as ESPCN, FSRCNN shown in Fig. 1(d) and (e). The former case would result in too many parameters in the network, heavy computational work and slow processing speed. Therefore, it will lead to unable to restore video in real time. The latter could reduce the spent time, but the performance of SR would degrade compared to the former due to the simple network architecture.

(3). In order to ensure that the size of the output image conforms to the requirements in (2), the network architecture usually just contains convolution layer, the activation layer and the loss layer, without deconvolution layer, pooling layer and so on, such as SRCNN, VDSR as shown in Fig. 1. This is because the deconvolution layer is equivalent to the inverse operation of the convolution layer. The deconvolution layer will magnify the size of the feature obtained in the previous layer, while the pooling layer will shrink the size of the feature map. These two layers will greatly change the size of the final output image of the network.