MULTI-EXPOSURE FUSION WITH CNN FEATURES

{ HUI LI AND LEI ZHANG }   DEPARTMENT OF COMPUTING
THE HONG KONG POLYTECHNIC UNIVERSITY

2018 IEEE International Conference on Image Processing
ICIP 2018

INTRODUCTION

Multi-exposure fusion (MEF) is to fuse a sequence of multi-exposure images to get a HDR-like image.

Motivation: Feature extraction plays a pivotal role in determining the weight map for MEF.

Our work: We investigate the effectiveness of convolutional neural network (CNN) features for MEF.

THE PROPOSED METHOD

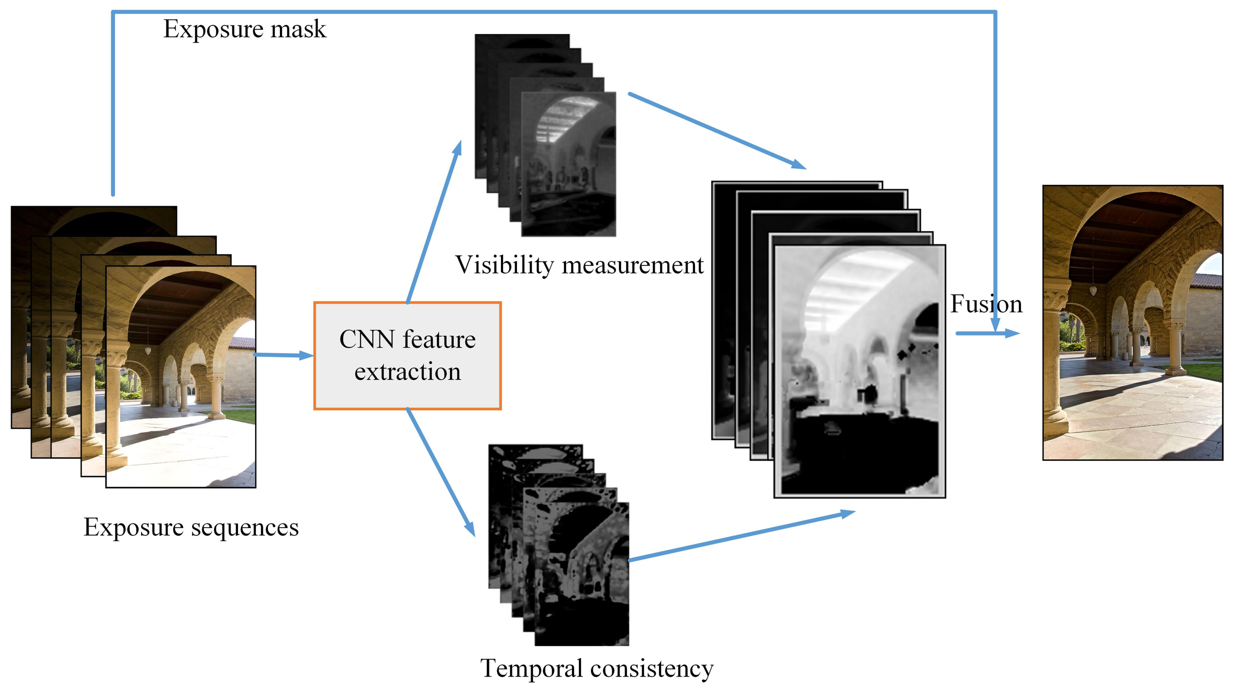

Figure 1: Flowchart of the proposed CNN feature based multi-exposure fusion method.

Feature extraction and visibility measurement:
The feature map of each source image is extracted as

$$\mathbf{F}_i(x,y) = CNN(I_i)(x,y), \quad (1)$$

We measure the visibility of pixel $I_i(x,y)$ as the L_1 norm of $V_i(x,y)$:

$$V_i(x,y) = \|\mathbf{F}_i(x,y)\|_1. \quad (2)$$

Temporal consistency:
The distance of two feature vectors at a pixel of two images is computed as:

$$s_{ij}(x,y)^2 = \|\bar{\mathbf{F}}_i(x,y) - \bar{\mathbf{F}}_j(x,y)\|_2^2, \quad (3)$$

We use a Gaussian kernel to map the similarity between $\bar{\mathbf{F}}_\mathbf{i}(\mathbf{x},\mathbf{y})$ and $\bar{\mathbf{F}}_\mathbf{j}(\mathbf{x},\mathbf{y})$ into the range of [0, 1]:

$$S_i(x,y) = \sum_{j=1}^{K} exp\frac{-s_{ij}(x,y)^2}{2\sigma^2}, \quad (4)$$

Fusion:
With the visibility and similarity weight maps V and S, we can get the final weight map $W_i(x,y)$ as follows:

$$W_i(x,y) = \frac{V_i(x,y) \times S_i(x,y) \times M_i(x,y)}{\sum_{j=1}^{K} V_i(x,y) \times S_i(x,y) \times M_i(x,y) + \alpha}, \quad (5)$$

The used mask is defined as follow:

$$M_i(x,y) = \begin{cases} 1, & \beta < I_i(x,y) < 1-\beta, \\ 0, & else, \end{cases} \quad (6)$$

Finally, we fuse the images as follows to produce the MEF output I_f:

$$I_f(x,y) = \sum_{i=1}^{K} I_i(x,y) \times W_i(x,y). \quad (7)$$

EXPERIMENTAL RESULTS

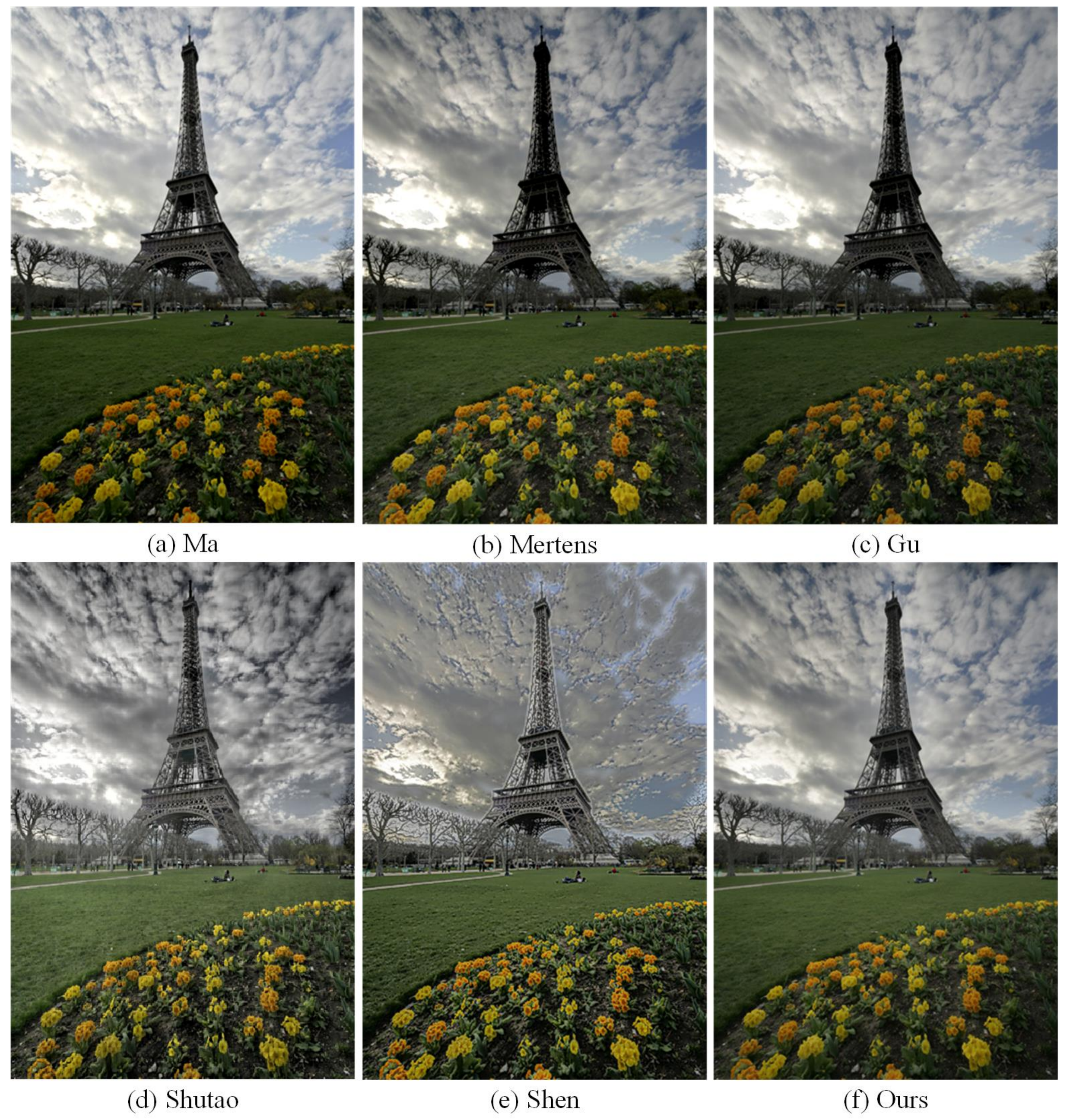

Figure 2: The MEF results by competing methods on a static scene

Figure 3: The MEF results by competing methods on a dynamic scene

Network Type	Denoising(DnCNN)				Super-resolution(VDSR)				Classification(VGG)			
Feature Layer	1	3	10	18	1	3	10	18	1	3	10	18
MEF-SSIM	0.869	0.970	0.969	0.965	0.867	0.957	0.846	0.930	0.969	0.620	0.610	0.560

Table 1: The MEF-SSIM scores by different networks at different layers on static scene dataset

Feature type	CNN	SIFT	Gabor
MEF-SSIM	0.969	0.952	0.900

Table 2: The MEF-SSIM scores of CNN and traditional features on the static scene dataset

Methods	Gu12	Li12	Shutao13	raman09	Shutao12	Shen14	Mertens09	Ma17	Ours
MEF-SSIM	0.910	0.944	0.965	0.852	0.960	0.753	0.975	0.977	0.969

Table 3: The average MEF-SSIM scores of different methods on the static scene dataset

CONCLUSION & FUTURE WORK

This paper made the first attempt to exploit the CNN features for weight design in MEF. We will investigate how to fine-tune the CNN features to make them more effective for MEF applications.

CONTACT INFORMATION

Name: Lei Zhang
Email: cslzhang@comp.polyu.edu.hk
Website: https://www4.comp.polyu.edu.hk/~cslzhang/