Supplementary Materials to “Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach”

Kede Ma, Student Member, IEEE, Hui Li, Hongwei Yong, Zhou Wang, Fellow, IEEE, Deyu Meng, Member, IEEE, and Lei Zhang, Senior Member, IEEE

In the supplementary file, we present in detail the parameter settings of SPD-MEF.

I. PARAMETER SETTINGS OF SPD-MEF

SPD-MEF has eight parameters in total, including
1) a small positive constant $\epsilon$ in Eq. (10);
2) the exponent parameter $p$ to determine the weight of the structure vector component;
3-4) two Gaussian spread parameters $\sigma_g$ and $\sigma_l$ to determine the weight of the mean intensity component;
5-6) two thresholds $T_s$ and $T_m$ that binarize the structural consistency map;
7-8) the patch size $N$ and its associated stride $D$.

In the following, we discuss the selection of these parameters in detail. Note that all the parameters are fixed when testing the full SPD-MEF algorithm in all experiments in the manuscript.

A. Constant $\epsilon$

The value of $\epsilon$ is inherited from the corresponding normalization term of the structural similarity (SSIM) index [14] and is equal to $\frac{1}{2}(0.03L_d)^2$, where $L_d$ is the maximum intensity value of the source sequence (For a normalized sequence, $L_d = 1$).

B. $p$, $\sigma_g$ and $\sigma_l$

The exponent parameter $p$ and two Gaussian spread parameters $\sigma_g$ and $\sigma_l$ in the baseline SPD-MEF algorithm are jointly determined by maximizing MEF-SSIM [7] on 5 held-out static source sequences using a grid search method. The possible values of $p$, $\sigma_g$ and $\sigma_l$ are chosen to be $p \in \{1, 2, \cdots, 10\}$, $\sigma_g \in \{0.1, 0.2, \cdots, 1\}$ and $\sigma_l \in \{0.1, 0.2, \cdots, 1\}$, respectively. In other words, there are 1,000 possible parameter combinations and the one that achieves the highest MEF-SSIM value on average is selected, which turns out to be \{$p = 4, \sigma_g = 0.2, \sigma_l = 0.5$\}. We select MEF-SSIM to optimize the parameter setting in SPD-MEF because it is designed specifically for MEF and is in closer agreement with human perception of fused image quality compared with other objective quality measures for general purpose fusion applications [7].

C. $T_s$ and $T_m$

The two thresholds $T_s$ and $T_m$ are crucial for SPD-MEF to work with dynamic scenes in the presence of camera and object motion. Both $T_s$ and $T_m$ have the same range [0, 1]. Ideally, the structural consistency map should be able to reject inconsistent motions w.r.t. the reference exposure while incorporating as many consistent patches as possible to make full use of all valid information for fusion. Since the consistency map $\hat{B}_k$ induced by $T_m$ works as a supplement to $\hat{B}_k$ induced by $T_s$, we first conduct an experiment to analyze the sensitivity of the map versus $T_s$ by disabling $\hat{B}_k$. Fig. S1 shows the generated structural consistency maps as a function of $T_s$, from which we have several observations. First, when $T_s$ is relatively small (e.g., $T_s < 0.5$), the structural consistency check is unable to fully reject moving objects, especially for the exposures that are far from the reference exposure (e.g., the first and second exposures w.r.t. the fourth reference exposure). On the other hand, when $T_s$ is too large (e.g., $T_s = 0.9$), the structural consistency check mistakenly considers some consistent motions across exposures as inconsistent ones and rejects them. Empirically, we find that $T_s = 0.8$ is capable of reliably identifying inconsistent motions across exposures while making full use of consistent motions for fusion. More results are shown in Figs. S2 and S3.

We then conduct another experiment to analyze the sensitivity of the map versus $T_m$ by fixing $T_s = 0.8$. In Fig. S4 and Fig. S5, we observe that when we disable its functionality by setting $T_m = 1$ or set a relatively large $T_m$ (e.g., $T_m = 0.5$), the ghost artifacts appear in the football and its shadow regions. This is because SPD-MEF considers these extremely dark parts as under-exposed regions\(^1\) and compensate for them by fusing parts of the couch from other images which are properly exposed. On the other hand, when $T_m$ is too small, SPD-MEF begins to reject consistent motions across exposures. We find

\(^1\)Note that no existing deghosting algorithm can robustly distinguish real under-exposed regions and dark regions of an object in the scene, especially when the number of input images is limited.
that $T_m = 0.1$ is a good choice for refining the structural consistency map and not rejecting consistent motion regions. More results are shown in Figs. S6 and Fig S7, from which we observe that many test sequences are robust to variations of $T_m$, conforming our claim that $T_m$ works as a supplement to $T_s$.

D. N and D

We now discuss the impact of patch size $N$ and the stride of the moving window $D$ on the fusion performance and computational time. The larger the $N$ is, the more robust the signal structure vector is in terms of structural consistency. However, the computational complexity also increases. Figs. S8, S9 and S10 show the perceptual quality of fused images and the computational time versus $N$ executed on a computer with 4G Hz CPU and 32G RAM. It can be observed that the computational time indeed increases with $N$, and a small $N$ results in some ghosting artifacts. We find that $N = 21$ provides a good balance between the performance and the complexity.

We also analyze the performance of SPD-MEF against the stride $D$ by fixing $N = 21$. From Figs. S11 and S12, we observe that the computational time drops as $D$ increases at the cost of possibly visible blocking artifacts. $D = \lfloor \frac{N}{10} \rfloor = 2$ is a good compromise between computational complexity and blocking artifacts.

\footnote{We fix $D = 2$ in this case to study the impact of $N$ solely.}
Fig. S1: Sensitivity analysis of the structural consistency map versus $T_s$. $B_h$ is disabled by setting $T_m = 1$, so as to solely observe the contribution of $T_s$. 

(a) Source image sequence by courtesy of Zhengguo Li [43] 

(b) Structural consistency maps generated by setting $T_s = 0.1$ 

(c) Structural consistency maps generated by setting $T_s = 0.5$ 

(d) Structural consistency maps generated by setting $T_s = 0.8$ 

(e) Structural consistency maps generated by setting $T_s = 0.9$
Fig. S2: Fusion results of SPD-MEF with different $T_s$ values on source sequences “Brunswick” and “Horse”, respectively.
Fig. S3: Fusion results of SPD-MEF with different $T_s$ values on source sequences “Russ1” and “Tate3”, respectively.
Fig. S4: Sensitivity analysis of the structural consistency map versus $T_m$ with $T_s$ fixed to 0.8.
Fig. S5: Fusion results of SPD-MEF with different $T_m$ values on the source sequence “Puppets”.

(a) $T_m = 1$

(b) $T_m = 0.5$

(c) $T_m = 0.1$

(d) $T_m = 0.05$
Fig. S6: Fusion results of SPD-MEF with different $T_m$ values on the source sequence “Lady”.

(a) $T_m = 1$

(b) $T_m = 0.5$

(c) $T_m = 0.1$

(d) $T_m = 0.05$
Fig. S7: Fusion results of SPD-MEF with different $T_m$ values on source sequences “Campus” and “Llandudno”, respectively.
Fig. S8: The perceptual quality of fused images and their execution time $t$ versus $N$ on the source sequence “Lady”.

(a) $N = 7, t = 5s$

(b) $N = 15, t = 8s$

(c) $N = 21, t = 12s$

(d) $N = 25, t = 15s$
Fig. S9: The perceptual quality of fused images and their execution time $t$ versus $N$ on the source sequence “Puppets”.

(a) $N = 7$, $t = 9s$
(b) $N = 15$, $t = 15s$
(c) $N = 21$, $t = 22s$
(d) $N = 25$, $t = 27s$
Fig. S10: The perceptual quality of fused images and their execution time $t$ versus $N$ on source sequences “Prof. JeonEighth” and “Office”, respectively.
Fig. S11: The perceptual quality of fused images and their execution time $t$ versus the stride of moving window $D$ on source sequences “Campus” and “Cliff”, respectively.
Fig. S12: The perceptual quality of fused images and their execution time $t$ versus the stride of moving window $D$ on source sequences “Square” and “Wroclaw”, respectively.