

Article

Conditional Random Field-Guided Multi-Focus Image Fusion

Odysseas Bouzos *, Ioannis Andreadis and Nikolaos Mitianoudis *

Department of Electrical and Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece

* Correspondence: obouzos@ee.duth.gr (O.B.); nmitiano@ee.duth.gr (N.M.)

Abstract: Multi-Focus image fusion is of great importance in order to cope with the limited Depth-of-Field of optical lenses. Since input images contain noise, multi-focus image fusion methods that support denoising are important. Transform-domain methods have been applied to image fusion, however, they are likely to produce artifacts. In order to cope with these issues, we introduce the Conditional Random Field (CRF) CRF-Guided fusion method. A novel Edge Aware Centering method is proposed and employed to extract the low and high frequencies of the input images. The Independent Component Analysis—ICA transform is applied to high-frequency components and a Conditional Random Field (CRF) model is created from the low frequency and the transform coefficients. The CRF model is solved efficiently with the α -expansion method. The estimated labels are used to guide the fusion of the low-frequency components and the transform coefficients. Inverse ICA is then applied to the fused transform coefficients. Finally, the fused image is the addition of the fused low frequency and the fused high frequency. CRF-Guided fusion does not introduce artifacts during fusion and supports image denoising during fusion by applying transform domain coefficient shrinkage. Quantitative and qualitative evaluation demonstrate the superior performance of CRF-Guided fusion compared to state-of-the-art multi-focus image fusion methods.

Keywords: multi-focus; image fusion; transform domain; graphical model



Citation: Bouzos, O.; Andreadis, I.; Mitianoudis, N. Conditional Random Field-Guided Multi-Focus Image Fusion. *J. Imaging* **2022**, *8*, 240. <https://doi.org/10.3390/jimaging8090240>

Academic Editor: Raimondo Schettini

Received: 23 July 2022

Accepted: 2 September 2022

Published: 5 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The limited Depth-of-Field of optical lenses allows only parts of the scene within a certain distance from the camera sensor to be captured well-focused each time, while the remaining parts of the scene stay out-of-focus or blurred. Multi-focus image fusion algorithms are thus of vital importance in order to cope with this limitation. Multi-focus image fusion methods merge multiple input images captured with different focus settings into a single image with extended Depth-of-Field. More precisely, the well-focused pixels of the input images are preserved in the fused image and the out-of-focus pixels of the input images are discarded. Consequently, the fused image should have extended Depth-of-Field and thus more information than each one of the input images and should not introduce artifacts during fusion.

The problem of multi-focus image fusion has been explored widely in the literature. Lately, a number of multi-focus image fusion methods have been proposed. Liu et al. [1] classified the multi-focus image fusion methods in four categories: spatial-domain methods, transform-domain methods, combined methods and deep learning methods. In spatial-domain methods, the fused image is estimated as the weighted average of the input images. Spatial-domain methods are also classified as block-based, region-based, and pixel-based. In block-based methods, the image is decomposed into blocks of fixed size, and the activity level is estimated individually for each of these blocks.

However, since blocks are likely to contain both well-focused and out-of-focus pixels, the block-based methods are likely to have blocking artifacts near the boundaries of well-focused and out-of-focus pixels. Thus, the fused image has lower quality near their boundary. Region-based methods, use a whole region of irregular shape in order to estimate the saliency of the included pixels. Although region-based methods provide

higher flexibility than block-based methods, a region may also contain simultaneously both well-focused and out-of-focus pixels. As a result, region-based methods also produce artifacts and have lower fused image quality near the boundaries of well-focused and out-of-focus pixels. In order to overcome these issues, pixel-based methods have lately gained more popularity. In these methods, activity level estimation is carried out at pixel level. Pixel-based methods do not have blocking artifacts and have better accuracy near the boundary of well-focused and out-of-focus pixels, however, they are likely to produce noisy weight maps, which also lead to fused images of lower image quality. Popular spatial domain-based multi-focus image fusion methods include: Quadtree [2], Boundary Finding [3], dense Sift [4], guided filtering [5], PCNN [6] and Image Matting [7]. Singh et al. [8] used the Arithmetic optimization algorithm (AOA) in order to estimate the weight maps for image fusion, which were refined with weighted least square optimization (WLS). The fused image is extracted through pixel-wise weighted average fusion. In [9], the fusion method FNMRA was presented, which used the modified naked mole-rat algorithm (mNMRA) in order to generate the weight maps, which were refined with weighted least-squares optimization. Pixel-wise single-scale composition was used in order to create the fused image.

In transform-domain methods, a forward transform is firstly applied to the input images. A fusion rule is then applied in order to combine the transform coefficients. Finally, an inverse transform is applied to the fused coefficients in order to return the fused image to the spatial domain. An advantage of dictionary-based transform-domain methods is the support of image denoising during fusion, by applying shrinkage methods, such as [10], which can be used to remove the noisy transform-domain coefficients. An issue of transform-domain methods lies in the imperfect forward-backward transforms that result in visible artifacts due to the Gibbs phenomenon. Since both the selection of the transform domain and the manual design of the fusion rule highly impact the quality of the fused image a number of transform domain-based multi-focus image fusion methods have been introduced. Typical transform domain-based multi-focus image fusion methods include: ICA [11], ASR [12], CSR [13], NSCT [14], NSCT-SR [15], MWGF [16] and DCHWT [17]. Qin et al. [18] proposed a new image fusion method combining discrete wavelet transform (DWT) and sparse representation (SR). Jagtap et al. [19] introduced information preservation-based guided filtering in order to decompose the input images into base and detail images. Low-rank representation was used in order to estimate the focus map and perform a fusion of the detailed images. In [20], the authors used weight maps based on local contrast, and the fused image was estimated with multi-scale weighted average fusion based on pyramid decomposition.

The methods that lie in the combined methods category employ both the merits of spatial and transform domain methods. Nonetheless, each method uses different domains. Bouzos et al. [21] combined the advantages of both the ICA domain and spatial domain. Chai et al. [22] combined advantages of multi-scale decomposition and spatial domain. He et al. [23] combined the Meanshift algorithm and NSCT domain. An issue of the aforementioned methods is that they do not support image denoising during fusion. Singh et al. [24] proposed the Discrete Wavelet Transform-bilateral filter (DWTBF) method, which combined the Discrete Wavelet Transform (DWT) and the bilateral filter. In [25], the authors combined a multi-resolution pyramid and the bilateral filter in order to predict the fused image.

Lately, deep learning-based methods have gained more popularity. According to the study [26], deep learning-based methods, are classified into decision map-based methods and end-to-end methods. In decision-map-based methods, the deep learning networks predict a decision map, with a classification-based architecture. Post-processing steps, including morphological operations, are usually employed to refine the decision map. The decision map is later used to guide the fusion of the input images, by selecting the respective pixels from the input images. Typical decision map-based deep learning multi-focus image fusion methods include: CNNFusion [27], ECNN [28] and p-CNN [29]. On the other hand,

end-to-end deep learning-based networks, directly predict the fused image without the intermediate step of the decision map. Typical end-to-end based deep learning networks for multi-focus image fusion include: IFCNN [30] and DenseFuse [31]. Ma et al. [32] introduced a multi-focus image fusion method based on an end-to-end multi-scale generative adversarial network (MsGAN). Wei et al. [33] combined advantages of sparse representation and CNN networks in order to estimate the fusion weights for the multi-focus image fusion problem. Since the sensitivity of the aforementioned deep learning-based methods to noise was not studied, the methods are likely to be sensitive to noise. In addition, these deep learning-based multi-focus image fusion methods do not support image denoising during fusion.

In this manuscript, we introduce CRF-Guided fusion, which is a novel transform domain-based method that uses the Conditional Random Field model, in order to guide the fusion of the transform-domain ICA method. Due to various sources, input images are likely to contain noise, thus multi-focus methods that are robust to noise and support fusion and denoising during fusion are of great importance. Since CRF-Guided fusion is a dictionary-based method (ICA), the method is robust to Gaussian noise and supports image denoising during fusion by applying the shrinkage coefficient method [10]. A novel Edge Aware Centering method (EAC) is also introduced and is used, instead of the typical centering method, and alleviates artifacts caused by the centering procedure. The combination of EAC and the proposed CRF-Guided fusion method produce fused images of high quality, without introducing artifacts for both clean images and images that contain Gaussian noise, while also supporting denoising during fusion.

The main contributions of this manuscript and improvements over our previous method [21] are:

1. the development of the novel EAC method, which preserves the strong edges of the input images, instead of the typical centering method.
2. the design of a novel framework, based on a CRF model, that is suitable for transform-domain image fusion.
3. the design of a novel transform-domain fusion method that produces fused images of high visual quality, preserves via CRF optimization, the boundary between well-focused and out-of-focus pixels, and does not introduce artifacts during fusion.
4. the introduction of a novel transform-domain fusion rule, based on the labels extracted from the CRF model, that produces fused images of higher image quality without the transform-domain artifacts
5. the robustness of the proposed method against Gaussian noise and the support of denoising during fusion, by applying the transform-domain coefficient shrinkage method [10].

2. Proposed Method Description

The proposed framework of the CRF-Guided fusion is summarised in Figure 1. An outline of the method is now provided: Firstly, Edge Aware Centering is applied to the input images, in order to extract the low and high-frequency components. The Forward ICA transform is then applied to the high frequencies of the input images. Then, the Low frequency and ICA coefficients are used to compute the Unary U and Smoothness V potentials and thus construct the CRF model. Consequently, the CRF model is solved efficiently with the α -expansion method based on GraphCuts [34]. The predicted labels are then employed to fuse the low frequencies leading to the fused low-frequency image. In addition, they are also used to guide the fusion of the transform-domain ICA coefficients. Lastly, the inverse ICA transform is applied to the fused transform coefficients in order to return the fused high-frequency component. Finally, the fused image F is estimated by the addition of the fused low-frequency and the fused high-frequency components. More details of the aforementioned steps of the proposed framework are included in the following subsections. Figure 2 includes two source input images for multi-focus image fusion that will be used during the steps of the CRF-Guided fusion.

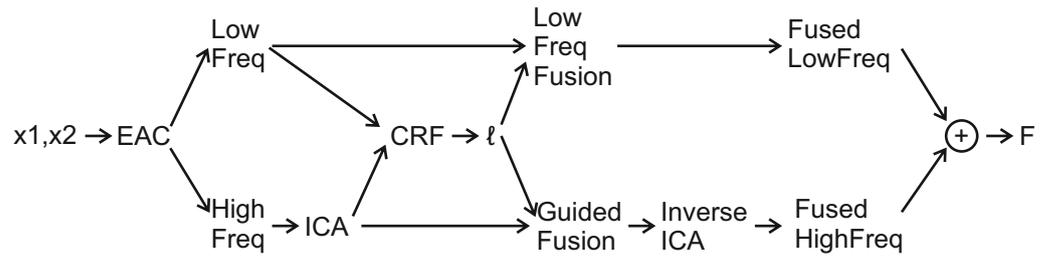


Figure 1. CRF-Guided fusion Framework for input images x_1, x_2 , labels ℓ as estimated from the CRF minimization, and the fused image F is constructed by the addition of the low-frequency and high-frequency fusion results.

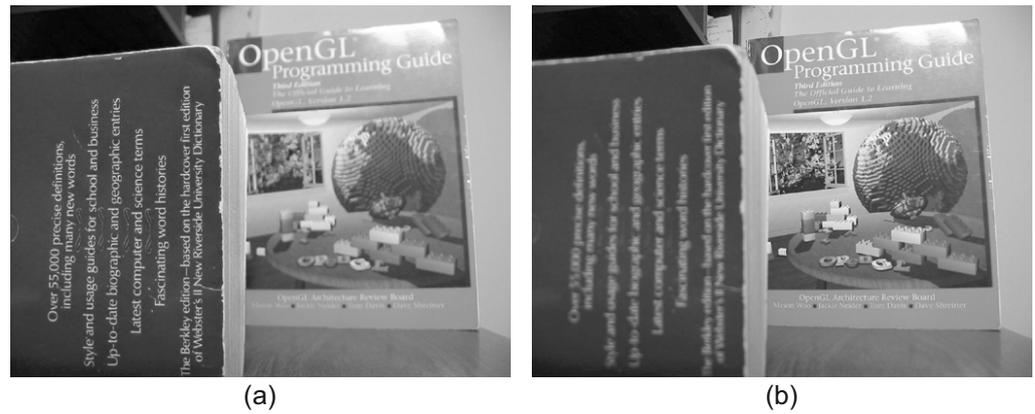


Figure 2. Source input images: (a) Near focused image, (b) Far focused image.

2.1. Edge Aware Centering

In this section, we introduce the *Edge Aware Centering* (EAC) method, which is used instead of the typical centering method, in order to estimate the low frequency of the multi-focus input images. EAC consists of a spatially varying Gaussian filter that preserves the strong edges of the input images. More precisely,

$$w_{i,j} = \exp\left(-\frac{(x_{i,j} - \mu_{i,j})^2}{2\langle (x_{m,n} - \mu_{i,j})^2 \rangle}\right) \tag{1}$$

where $w_{i,j}$ is the weight at spatial location (i, j) , $\mu_{i,j}$ is the mean value of a 7×7 block with central pixel at (i, j) , x is the input image and $m \in [i - 3, i + 3], n \in [j - 3, j + 3]$. In addition, the $\langle \cdot \rangle$ operator implies averaging over the all m, n values. Finally, the filtered image f in spatial locations (i, j) is estimated as:

$$f_{i,j} = \frac{\sum_{m,n} w_{m,n} I_{mn}}{\sum_{m,n} w_{m,n}} \tag{2}$$

EAC is applied to both input images in order to estimate the low frequency of each image. Figure 3 includes the low-frequency images, as computed by applying the proposed EAC to the input images of Figure 2. It is evident that the EAC preserves accurately the strong edges of the input images.



Figure 3. Low frequency of input images using the EAC: (a) Low frequency of near focused image, (b) Low frequency of far focused image. It is evident that the EAC preserves the strong image edges.

By subtracting the low-frequency images from the input images, we extract the high-frequency images as demonstrated in Figure 4. The forward ICA transform is then applied to the high-frequency images in order to get the transform domain coefficients. For more information on the estimation of the ICA transform, and its application on images for fusion, please refer to [11].



Figure 4. High frequency of input images: (a) High frequency of near focused image, (b) High frequency of far focused image.

2.2. Energy Minimization

In order to model the multi-focus image fusion problem and solve it efficiently, we construct an energy minimization equation. Since solvers of graph cuts can reach a global or close-to-global optimum solution, we formulate the energy minimization problem of multi-focus image fusion as a graph cut problem. More precisely, we introduce the Conditional Random Field (CRF) equation that describes our multi-focus image fusion problem, which is solved efficiently with the inference method of α -expansion reaching a global or close-to-global optimum solution. The solution of the proposed energy minimization leads to the optimum labels of the decision that is used to guide the fusion of low frequency and transform coefficients.

In order to guide the fusion of the low frequency and the transform coefficients, we formulate the Conditional Random Field (CRF) equation, as follows:

$$\ell = \arg \min_{\ell} \left[\sum_{i=1}^N U(\ell_i) + \sum_{(m,n) \in C} V_{m,n}(\ell_m, \ell_n) \right] \quad (3)$$

where ℓ are the estimated labels, U is the unary potential function, V is the pairwise potential function, i are spatial locations, and m, n adjacent pixels in the C which is the N8-neighborhood. The energy minimization equation is optimized using the α -expansion method, based on GraphCuts [34].

2.3. Inference α -Expansion Method

In the α -expansion, the optimization problem is divided into a sequence of binary-valued maximization problems. Given a current label configuration h and a fixed label $\alpha \in U$, with U being the set of all label values. In the α -expansion move, each pixel i gets a binary decision, to either retain its old value or change it to label α . The expansion move starts with the initial set of labels h^0 and then based on some order, computes the optimal α -expansion moves for the labels α . Only the moves that lead to the increase of the objective function are accepted.

2.4. Unary Potential Estimation

Let us assume that x_1, x_2 are the input images, P_L is the probability of low frequency, P_H the probability of high frequency, P the probability of the input images, and U unary potential function. Figure 5 depicts the method of estimating the unary potential. More precisely, EAC is firstly applied to the images to extract low and high frequencies. The 2nd Laplacian is applied to both low-frequency components and the probability of the low frequency, P_L is estimated by:

$$P_L(\ell_n) = \begin{cases} \frac{S_0}{S_0+S_1}, & \ell_n = 0 \\ \frac{S_1}{S_0+S_1}, & \ell_n = 1 \end{cases} \tag{4}$$

where, S_0, S_1 are the second Laplacian of the low frequencies of the first and the second image respectively.

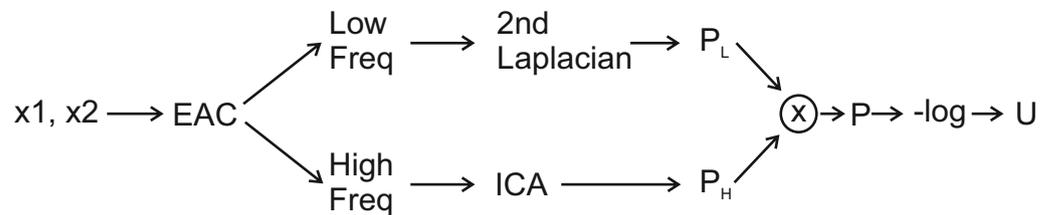


Figure 5. Unary potential estimation for CRF-Guided method.

The probability of the high frequency P_H is extracted by the ICA coefficients and is estimated as follows:

$$P_H(\ell_n) = \begin{cases} \frac{\|C_0\|}{\|C_0\|+\|C_1\|}, & \ell_n = 0 \\ \frac{\|C_1\|}{\|C_0\|+\|C_1\|}, & \ell_n = 1 \end{cases} \tag{5}$$

where $\|C_0\|$ is the L2-norm of ICA coefficients of the first image, $\|C_1\|$ is the L2-norm of ICA coefficients of the second image. In order to determine the probability that each one of the input images i should contribute to the spatial location n of the guidance map, we compute the combined probability of high and low frequencies for each image. This probability we call the probability of input image that corresponds to label ℓ . Thus probability of each input image $P(\ell_n)$ is estimated as follows:

$$P(\ell_n) = P_H(\ell_n) * P_L(\ell_n) \tag{6}$$

Finally, the Unary potential function U is estimated by the negative likelihood of the predicted probabilities:

$$U(\ell_n) = -\log P(\ell_n) \tag{7}$$

2.5. Smoothness Term

The smoothness potential function V is estimated from the low-frequency image, as follows:

$$V_{pq} = \frac{|l_0^p - l_1^q| + |l_1^p - l_0^q|}{|l_0^p - l_0^q| + |l_1^p - l_1^q|} \tag{8}$$

where p, q are adjacent pixels in the N8-neighborhood and l_0, l_1 are the first and second low-frequency images respectively. Finally, the labels ℓ of the CRF model in (3) are estimated efficiently using the α -expansion method [34].

Figure 6 demonstrates the labels, as estimated from the direct minimization of the unary term U and the labels, as estimated from the CRF minimization (3). The predicted labels are then used to fuse the low frequency of the input images.

$$L_F(i) = (1 - \ell_i) * L_0(i) + \ell_i * L_1(i) \tag{9}$$

where L_F is the low-frequency fused image, i is the spatial location, L_0 is the low frequency of the first image and L_1 is the low-frequency of the second image.



Figure 6. Predicted labels, black pixels correspond to $\ell = 0$, white pixels correspond to $\ell = 1$, (a) $\ell = \arg \min U$, (b) $\ell = \arg \min(CRF)$.

2.6. Transform-Domain CRF Fusion Rule

A sliding window with size 7×7 is applied to the decision map of the predicted probabilities. The transform coefficients that correspond to each 7×7 block are then fused according to the label of the central pixel of the block by selecting the respective coefficients from the input images that correspond to that label. Inverse ICA is then applied to the fused transform coefficients in order to return the fused high frequency. Figure 7 depicts the fused low-frequency component and the fused high-frequency component.



Figure 7. (a) Fused low frequency, (b) Fused high frequency.

Finally, the fused image is estimated by the addition of the low and high-frequency components. Figure 8 demonstrates the final fused image.

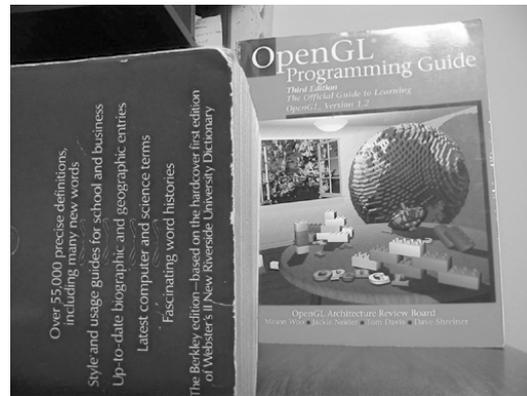


Figure 8. Final fused image by the proposed method.

3. Fusion and Denoising

A major advantage of the proposed CRF-Guided fusion is the robustness against Gaussian noise and the support of denoising during fusion. In the case of Gaussian noise, the coefficient shrinkage method [10] is applied to the transform coefficients of both input images. More precisely,

$$C(k) = 0, \text{ if } |C(k)| < 1.95 * \sigma_n \tag{10}$$

where $C(k)$ is the k -th transform coefficient in the ICA domain and σ_n is the standard deviation of the noise, which is estimated by areas of the image where there is low activity. Low activity areas contain no strong edges, therefore may contain only noise and thus can be used to estimate the noise standard deviation σ_n . The denoised transform coefficients are then employed to estimate the P_H of both input images. Consequently, Guided fusion from the CRF labels is performed on the denoised transform coefficients. Then, the inverse ICA transform is used to return the denoised high-frequency image. Lastly, the final denoised fused image is formed by the addition of the denoised high-frequency and the fused low-frequency images.

Figure 9 includes the noisy input images with Gaussian noise $\mathcal{N}(0, \sigma^2)$, $\sigma = 5$ and the denoised fused image F . The fused image F is successfully denoised during the fusion, as is demonstrated in Figure 9c.



Figure 9. (a) Near-focused image with Gaussian noise $\sigma_n = 5$, (b) Far-focused image with Gaussian noise $\sigma_n = 5$, (c) Denoised fused image.

Figure 10 includes the noisy input images with Gaussian noise $\mathcal{N}(0, \sigma^2)$, $\sigma = 10$ and the denoised fused image F . The proposed CRF-Guided fusion framework can successfully produce the denoised fused image Figure 10c, with denoising performed during fusion.

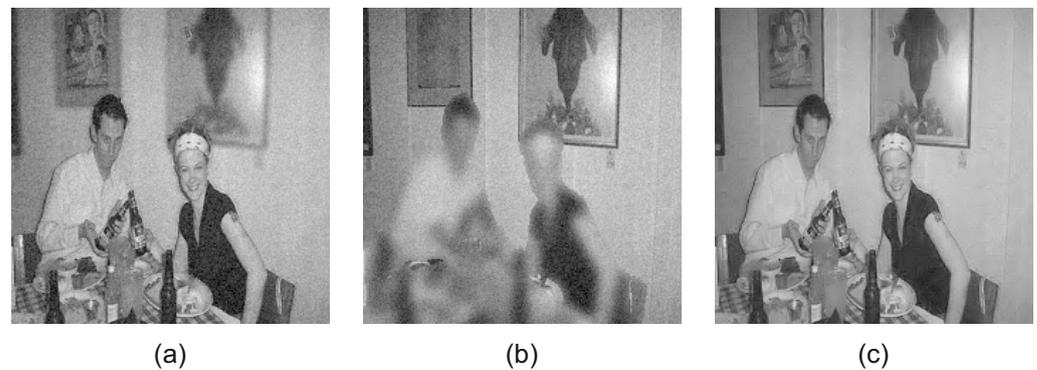


Figure 10. (a) Near focused image with Gaussian noise $\sigma = 10$, (b) Far focused image with Gaussian noise $\sigma = 10$, (c) Denoised fused image.

4. Experimental Results

The proposed CRF-Guided fusion method is compared to 13 state-of-the-art image fusion methods in the two public datasets: the Lytro dataset [35], which consists of 20 color input image pairs and the grayscale dataset [3], which consists of 17 grayscale input image pairs. The state-of-the-art compared methods are: GBM [36], NSCT [14], ICA [11], DCHWT [17], ASR [12], IFCNN [30] and DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] and Joint [42]. Both quantitative and qualitative results are included in order to evaluate the performance of CRF-Guided fusion and the compared multi-focus image fusion methods.

4.1. Quantitative Evaluation

In [43,44] Singh et al. made a review of multiple image fusion algorithms along with the image fusion performance metrics. In order to assess the quality of the fused images of the compared multi-focus image fusion methods, eight metrics are used. More precisely the metrics used are: Mutual Information (MI) [45], $Q_{ab/f}$ [46], Q_g [47], Q_y [48], CB [49], SSIM [50], NIQE [51] and Entropy.

4.1.1. Mutual Information—MI

Mutual Information—MI is an information theory-based metric and the objective measure of the mutual dependence of two random variables. For two discrete random variables U and V , MI is defined as follows:

$$MI(U; V) = \sum_{v \in V} \sum_{u \in U} p(u, v) \log_2 \frac{p(u, v)}{p(u)p(v)} \tag{11}$$

4.1.2. Yang’s Metric Q_Y

Yang et al. [48] proposed the image structural similarity-based metric Q_Y . For input images A, B and fused image F , it is defined as follows:

$$Q_Y = \begin{cases} \lambda(w)SSIM(A, F|w) + (1 - \lambda(w))SSIM(B, F|w), & SSIM(A, B|w) \geq 0.75 \\ \max\{SSIM(A, F|w), SSIM(B, F|w)\}, & SSIM(A, B|w) < 0.75 \end{cases} \tag{12}$$

$$\lambda(w) = \frac{s(A|w)}{s(A|w) + s(B|w)} \tag{13}$$

where $s(A|w)$ is a local salience measure of image A within a window w . A higher value of Q_Y indicates better-fused image quality and higher structural similarity of the fused images and the input images.

4.1.3. Chen-Blum Metric—CB

The Chen-Blum Metric CB [49] is a human perception-inspired fusion metric that features the following five steps:

1. Contrast sensitivity filtering: Filtered image $I'_A(m, n) = I_A(m, n)S(r)$, where $S(r)$ is the CSF filter in polar form and $r = \sqrt{m^2 + n^2}$.
2. Local contrast computation:

$$C(i, j) = \frac{\phi_k(i, j) * I(i, j)}{\phi_{k+1}(i, j) * I(i, j)} - 1 \tag{14}$$

$$\phi_k(x, y) = \frac{1}{(\sqrt{2\pi}\sigma_k)^2} e^{-\frac{x^2+y^2}{2\sigma_k^2}} \tag{15}$$

where $\sigma_k = 2$.

3. Contrast preservation calculation: For input image I_A the masked contrast map is estimated as:

$$C'_A = \frac{t(C_A)^p}{h(C_A)^q + Z} \tag{16}$$

where t, h, p, q, Z are real scalar parameters that determine the shape of the nonlinearity of the masking function [49].

4. Generation of saliency map: The saliency map for image I_A is:

$$\lambda_A = \frac{C_A'^2}{C_A'^2 + C_B'^2} \tag{17}$$

The value of information preservation is:

$$Q_{AF} = \begin{cases} \frac{C'_A}{C'_F} & \text{if } C'_A < C'_F \\ \frac{C'_F}{C'_A} & \text{otherwise.} \end{cases} \tag{18}$$

5. The global quality map is defined as:

$$Q_{GQM}(i, j) = \lambda_A(i, j)Q_{AF}(i, j) + \lambda_B(i, j)Q_{BF}(i, j) \tag{19}$$

The value of metric *CB* is the average of the global quality map:

$$CB = \text{mean}_{i,j}Q_{GQM}(i, j) \tag{20}$$

4.1.4. Gradient Based Methods— $Q_G, Q_{AB/F}$

Xydeas et al. [47] proposed a metric to measure the amount of edge information from source images to the fused image. Q_G is a gradient-based method. Firstly, a Sobel operator is applied to input image *A* to extract edge strength $g_A(i, j)$ and orientation $\alpha_A(i, j)$.

$$g_a(i, j) = \sqrt{s_A^x(i, j)^2 + s_A^y(i, j)^2} \tag{21}$$

$$\alpha_A(i, j) = \tan^{-1}\left(\frac{s_A^y(i, j)}{s_A^x(i, j)}\right) \tag{22}$$

where s_A^x, s_A^y are the outputs of the convolution application of the horizontal and vertical Sobel templates respectively. The relative strength between input image *A* and fused image *F* is:

$$G^{AF}(i, j) = \begin{cases} \frac{g_F(i, j)}{g_A(i, j)}, & \text{if } g_A(i, j) > g_F(i, j) \\ \frac{g_A(i, j)}{g_F(i, j)}, & \text{otherwise.} \end{cases} \tag{23}$$

The orientation values Δ^{AF} between input image *A* and fused image *F* are:

$$\Delta^{AF}(i, j) = 1 - \frac{|\alpha_A(i, j) - \alpha_F(i, j)|}{\pi/2} \tag{24}$$

The edge strength value is estimated as:

$$Q_g^{AF}(i, j) = \frac{\Gamma_g}{1 + e^{k_g(G^{AF}(i, j) - \sigma_g)}} \tag{25}$$

The orientation preservation value is estimated as:

$$Q_\alpha^{AF}(i, j) = \frac{\Gamma_\alpha}{1 + e^{k_\alpha(\Delta^{AF}(i, j) - \sigma_\alpha)}} \tag{26}$$

The constants Γ_g, k_g, σ_g and $\Gamma_\alpha, k_\alpha, \sigma_\alpha$ are used to define the shape of the sigmoid functions used for the edge strength and orientation preservation values [47].

$$Q^{AB/F} = \frac{\sum_{n=1}^N \sum_{m=1}^M [Q^{AF}w^A + Q^{AB}w^B]}{\sum_{n=1}^N \sum_{m=1}^M [w^A + w^B]} \tag{27}$$

and

$$Q^{AF} = Q_g^{AF} Q_\alpha^{AF} \tag{28}$$

where $Q^{AF}(i, j)$ denotes the edge similarity at position (i, j) between input image *A* and fused image *F*, Q_g^{AF} the edge strength similarity and Q_α^{AF} the orientation similarity.

4.1.5. Structural Similarity Index—SSIM [50]

The structural similarity index—SSIM for two images *A, B* is defined as:

$$SSIM(A, B) = \frac{(2\mu_A\mu_B + C_1)(2 * \sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{29}$$

where μ_A, μ_B are the mean intensity values of images A, B , σ_A, σ_B are the standard deviation of images A, B and σ_{AB} is the square root of covariance of A, B . C_1, C_2 are constants. Due to the lack of ground truth image, the $SSIM$ for input images A, B and fused image F in the experiments is defined as follows:

$$SSIM = \frac{SSIM(A, F) + SSIM(B, F)}{2} \tag{30}$$

where A and B are the two input images and F is the fused image.

4.1.6. Niqe [51]

$NIQE$ is a blind image quality metric based on the Multivariate Gaussian Model (MVG). The quality of the fused image is defined as the distance between the quality aware natural scene statistic (NSS) model and the MVG fit, extracted from features of the distorted image:

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{\left((v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2) \right)} \tag{31}$$

where v_1, v_2 and Σ_1, Σ_2 are the mean vectors and covariance matrices of the natural multivariate Gaussian model [51] and the multivariate Gaussian model that is fit to the fused image.

4.1.7. Entropy

The entropy of an image I is defined as:

$$E_I = - \sum_{j=1}^{2^L-1} p(s_j) \log_2(p(s_j)) \tag{32}$$

where L is the number of gray levels, $p(s_j)$ is the probability of occurrence of gray level s_j in image I .

Table 1 includes the objective evaluation of the compared methods for the Lytro dataset [35].

For the Lytro dataset [35], the proposed CRF-Guided fusion method has the highest value for the metrics $MI, Q_g, Q^{AB/F}, Q_Y, CB$, the lowest value for the $NIQE$ metric and the second highest score for $SSIM$ and entropy. These results indicate that the fused quality of the proposed fused image is better than the compared state-of-the art methods. Since CRF-Guided has the highest Mutual Information [45], the proposed method preserves best the information of the input images. In addition, CRF-Guided has the highest Q_g [47] and $Q^{AB/F}$ [46] values, which indicate that the proposed method preserves best the edge information from the input images to the fused image. In order to assess the quality of the structural similarity of the fused images, Yang’s metric Q_Y [48] and the structural similarity index measure $SSIM$ [50] are employed. The proposed method has the highest Q_Y value and the second highest according to $SSIM$, which indicates high fused image quality, regarding structural similarity. DenseFuse [31] has highest $SSIM$ value for the Lytro dataset. The proposed CRF-Guided method has the highest value on the human perception inspired fusion metric CB [49], which implies that perceptually the produced results by the method are the most pleasing to the human eye. According to the blind image quality metric $NIQE$ [51], CRF-Guided has the lowest value and thus the best fused image quality. Lastly, for the blind image quality Entropy, GBM [36] has the highest score and CRF-Guided has the second highest score. Overall for the Lytro dataset [35] of perfectly registered color input images, the proposed CRF-Guided method outperforms the compared state-of-the art image fusion methods in most metrics.

Table 1. Objective evaluation for the Lytro dataset [35]. Lower values for *NIQE* indicate better fused image quality, while for rest metrics higher values indicate better fused image quality.

Methods	<i>MI</i> [45]	<i>Qg</i> [47]	$Q^{AB/F}$ [46]	Q_y [48]	<i>CB</i> [49]	<i>SSIM</i> [50]	<i>NIQE</i> [51]	<i>Entropy</i>
ASR [12]	7.1310	0.7510	0.7013	0.9691	0.7264	0.8437	3.4591	7.5217
NSCT [14]	7.1986	0.7502	0.6960	0.9649	0.7527	0.8432	3.4479	7.5309
GBM [36]	3.8813	0.7172	0.6202	0.8554	0.6159	0.7932	3.0434	7.5684
ICA [11]	6.8769	0.7393	0.6741	0.9512	0.7088	0.8534	3.3915	7.5267
IFCNN [30]	7.0400	0.7337	0.6628	0.9522	0.7292	0.8440	3.4623	7.5319
DenseFuse [31]	6.2048	0.5532	0.4694	0.8141	0.6037	0.8651	3.3953	7.4681
dchwt [17]	6.7298	0.7184	0.6078	0.9202	0.6924	0.8526	3.2976	7.5205
acof [37]	7.2675	0.5287	0.5112	0.9475	0.6387	0.8260	4.6501	7.4901
cfl [38]	5.6254	0.6576	0.5746	0.8827	0.6323	0.8158	3.4033	7.5734
ConvCFL [39]	5.9742	0.6916	0.5864	0.8869	0.6643	0.8396	3.7099	7.5581
DTNP [40]	6.7854	0.7431	0.6779	0.9566	0.7347	0.8390	3.4198	7.5298
mlcf [41]	6.4414	0.5377	0.5147	0.8593	0.6259	0.8564	3.8699	7.4906
joint [42]	6.9991	0.7435	0.6970	0.9621	0.7176	0.8426	3.3935	7.5200
CRFGuided	7.3639	0.7534	0.7143	0.9851	0.7557	0.8601	3.0336	7.5697

Table 2 includes the quantitative evaluation of the compared methods for the grayscale dataset [3]. The CRF-Guided fusion method outperforms the compared state-of-the-art methods, in terms of metrics *MI* [45], *Qg* [47], $Q^{AB/F}$ [46], Q_y [48], *CB* [49] and *SSIM* [50] and has the second lowest score for the *NIQE* [51] metric and the second highest Entropy value. More precisely, since CRF-Guided has the highest Mutual Information [45], it preserves better the original information compared to the other methods. The highest value of CRF-Guided in Q_g [47] and $Q^{AB/F}$ [46] indicate that the proposed method preserves better the edges of the input images, compared to the state-of-the-art methods. Moreover, the structural information of the original images is best preserved in the CRF-Guided method, since both Q_y [48] and *SSIM* [50] have the highest value for the proposed method. According to the human perception inspired fusion metric *CB* [49], CRF-Guided has the best fused image quality. For the *NIQE* [51] metric, the method dchwt [17] has the lowest score and the proposed method has the second lowest value. The method GBM [36] has the highest entropy value for the grayscale dataset. Overall, the proposed method has the highest fused image compared to the state of the art methods for the grayscale dataset [3].

Table 2. Objective evaluation for the grayscale dataset [3]. Lower values for *NIQE* indicate better fused image quality, while for rest metrics higher values indicate better fused image quality.

Methods	<i>MI</i> [45]	<i>Qg</i> [47]	$Q^{AB/F}$ [46]	Q_y [48]	<i>CB</i> [49]	<i>SSIM</i> [50]	<i>NIQE</i> [51]	<i>Entropy</i>
ASR [12]	6.3790	0.7192	0.6721	0.9541	0.7057	0.8150	5.5111	7.3262
NSCT [14]	6.2947	0.7074	0.6593	0.9439	0.7284	0.8161	5.3080	7.3451
GBM [36]	3.5292	0.6729	0.5826	0.8275	0.6005	0.7503	5.0053	7.5298
ICA [11]	6.0174	0.6945	0.6507	0.9313	0.6996	0.8302	5.2144	7.3449
IFCNN [30]	5.9641	0.6743	0.6074	0.9118	0.6725	0.8230	5.4436	7.3435
DenseFuse [31]	6.0467	0.6139	0.5798	0.8517	0.6275	0.8351	5.2584	7.3739
dchwt [17]	5.9965	0.6781	0.5810	0.8997	0.6752	0.8244	4.9713	7.3396
acof [37]	6.5748	0.5594	0.5543	0.8691	0.6183	0.8098	5.1625	7.3088
cfl [38]	4.8158	0.5985	0.5327	0.8548	0.6138	0.7966	5.5156	7.4403
ConvCFL [39]	5.3014	0.6510	0.5619	0.8640	0.6558	0.8234	5.5023	7.3895
DTNP [40]	6.0911	0.6966	0.6357	0.9296	0.7056	0.8119	5.2817	7.3496
mlcf [41]	6.3294	0.5912	0.5890	0.9274	0.6594	0.8040	5.2670	7.3176
joint [42]	6.6541	0.7212	0.6775	0.9553	0.7234	0.8102	5.4543	7.3239
CRFGuided	6.6740	0.7290	0.6903	0.9798	0.7337	0.8356	5.0001	7.3928

In summary, according to the 8 metrics used for quantitative evaluation, the proposed CRF-Guided method has the best performance compared to 13 state-of-the-art image fusion methods for both public datasets: the Lytro dataset [35] and the grayscale dataset [3].

4.2. Qualitative Evaluation

In this section, we perform a visual comparison between the tested methods. Figure 11 includes the fused results of the compared methods for the scene ‘Lab’ of the grayscale dataset [3]. The compared methods GBM [36], NSCT [14], ICA [11], DCHWT [17], ASR [12], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] and Joint [42], all feature visible artifacts in the area of the head. Moreover, these methods cannot accurately preserve the boundary of the clock in the red rectangle. MLCF cannot accurately capture the boundaries of the well-focused and out-of-focus pixels. NSCT [14], ICA [11], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39] also feature artifacts around the arm, included in the red rectangle area. The proposed CRF-Guided method has the highest fused image quality for the area of the head, without introducing artifacts during fusion. Furthermore, the boundary of the clock is best preserved in the CRF-Guided fusion method, compared to the state-of-the-art methods. Moreover, the CRF-Guided fusion method does not introduce artifacts in the area of the red rectangle around the arm. The proposed CRF-Guided method does not have artifacts during fusion and has the highest visual image quality for the ‘Lab’ scene.

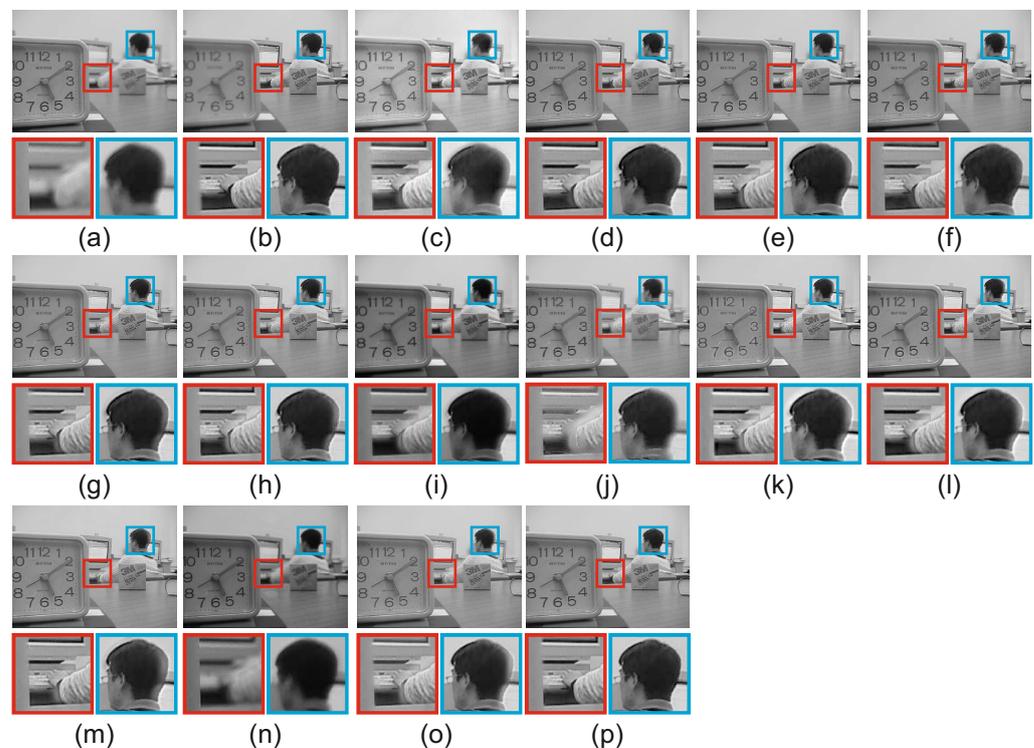


Figure 11. Fused results for the scene ‘Lab’ of the grayscale dataset [3]. (a) Source 1, (b) Source 2, (c) GBM [36], (d) NSCT [14], (e) ICA [11], (f) DCHWT [17], (g) ASR [12], (h) IFCNN [30], (i) DenseFuse [31], (j) acof [37], (k) CFL [38], (l) ConvCFL [39], (m) DTNP [40], (n) MLCF [41], (o) Joint [42], (p) CRFGuided.

Figure 12 includes the resulting fused images of the proposed and the compared methods for the scene ‘Temple’ of the grayscale dataset [3]. Two regions are selected for magnification to assess with qualitative evaluation. GBM [36], NSCT [14], ICA [11], DCHWT [17], ASR [12], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] and Joint [42] all have visible artifacts in both regions of the red and the blue rectangles. Moreover, they cannot accurately preserve the boundary of the well-focused and out-of-focus pixels. The proposed CRF-Guided method preserves accurately the boundaries between the well-focused and out-of-focus pixels for both regions without introducing artifacts, compared to the other multi-focus image fusion methods. CRF-Guided features the best fused image quality for the scene ‘Temple’. Qualitative evaluation

indicates that the proposed CRF-Guided method has the best visual fused image quality, without introducing artifacts during fusion, compared to 13 state-of-the-art methods.

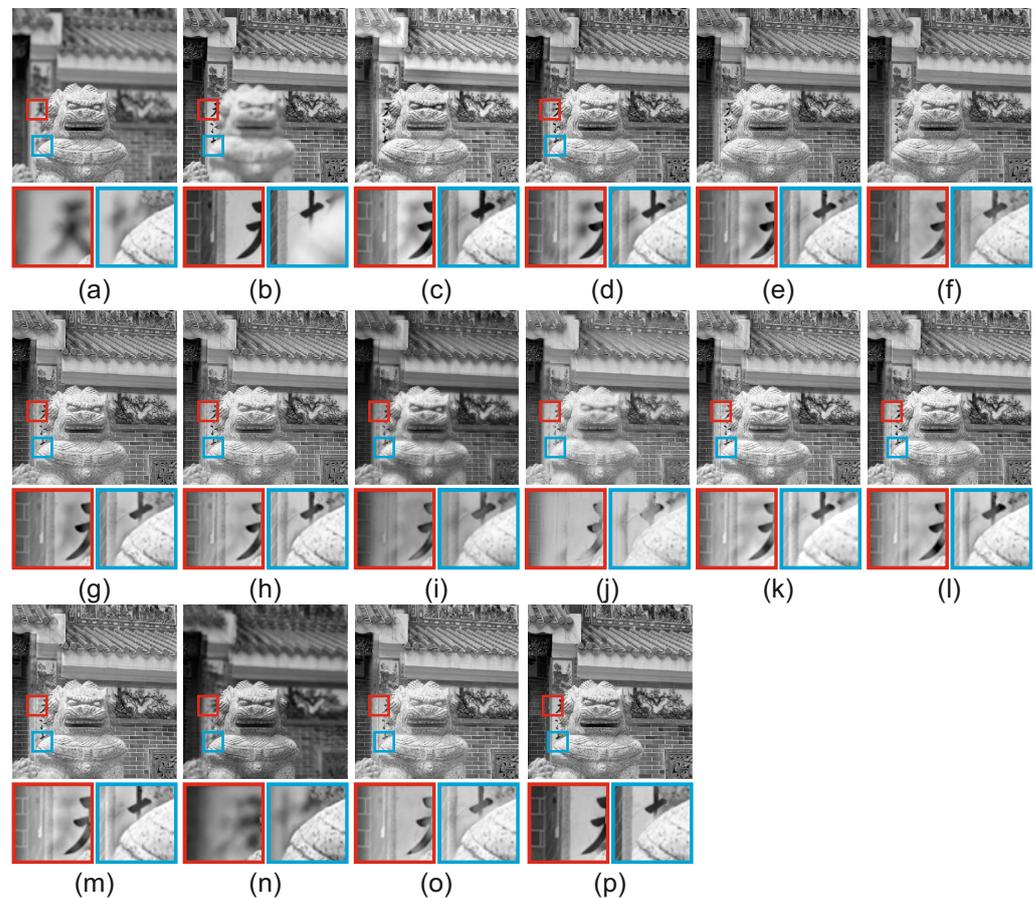


Figure 12. Fused results for the scene ‘Temple’ of the grayscale dataset [3]. (a) Source 1, (b) Source 2, (c) GBM [36], (d) NSCT [14], (e) ICA [11], (f) DCHWT [17], (g) ASR [12], (h) IFCNN [30], (i) DenseFuse [31], (j) acof [37], (k) CFL [38], (l) ConvCFL [39], (m) DTNP [40], (n) MLCF [41], (o) Joint [42], (p) CRFGuided.

Figure 13 includes the qualitative evaluation of the compared methods for the scene ‘Golfer’ of the Lytro dataset [35]. CFL [38] and ConvCFL [39] produce artifacts around the boundary of well-focused and out-of-focus pixels in both regions. The boundary of the well-focused pixels isn’t well preserved in GBM [36], NSCT [14], ICA [11], DCHWT [17], ASR [12], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] and joint [42], while on the proposed CRFGuided the fused image is better preserved. Methods acof [37], mlcf [41] cannot accurately capture the boundary of well-focused and out-of-focus pixels in both regions. NSCT [14], DenseFuse [31], acof [37], DTNP [40] and MLCF [41] cannot preserve accurately the boundaries between the well-focused and out-of-focus pixels in the area of the red rectangle. The proposed CRFGuided method has the highest visual quality for both regions for the ‘Golfer’ scene of the Lytro [35] dataset, preserving best the boundary of well-focused and out-of-focus pixels, without introducing artifacts during fusion.

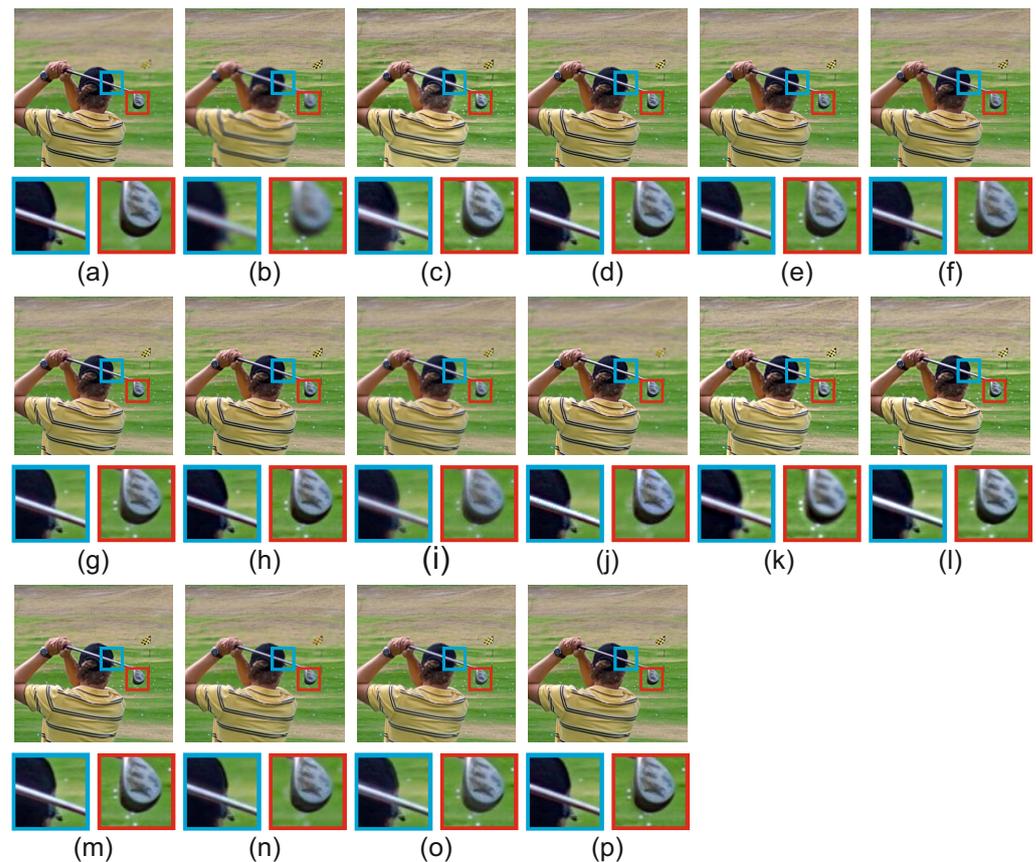


Figure 13. Fused results for the scene ‘Golfer’ of the Lytro dataset [35]. (a) Source 1, (b) Source 2, (c) GBM [36], (d) NSCT [14], (e) ICA [11], (f) DCHWT [17], (g) ASR [12], (h) IFCNN [30], (i) DenseFuse [31], (j) acof [37], (k) CFL [38], (l) ConvCFL [39], (m) DTNP [40], (n) MLCF [41], (o) Joint [42], (p) CRFGuided.

Figure 14 features the qualitative evaluation for the ‘Volley’ scene of the Lytro [35] dataset. Two regions were selected with magnification. For the blue region, the boundaries of well-focused and out-of-focus pixels in methods GBM [36], NSCT [14], ICA [11], DCHWT [17], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] are not accurately preserved. Methods acof [37] and MLCF [41] can not preserve accurately the boundaries of well-focused and out-of-focus pixels in both regions. For the red region, Joint [42] produces color degradation and lower contrast. Moreover, GBM [36], NSCT [14], ICA [11], DCHWT [17], ASR [12], IFCNN [30], DenseFuse [31], acof [37], CFL [38], ConvCFL [39], DTNP [40], MLCF [41] can not preserve well the boundary of well-focused and out-of-focus pixels and the back shoe is not well-focused. The proposed CRFGuided preserves best the boundary between well-focused and out-of-focus pixels for both regions for the ‘Volley’ scene of the Lytro dataset [35], resulting in a fused image of high quality without introducing artifacts during fusion.

According to the previous qualitative evaluation, the proposed CRF-Guided fusion method produces fused images of high quality, preserving best the boundary of well-focused and out-of-focus pixels without introducing artifacts during fusion.



Figure 14. Fused results for the scene ‘Volley’ of the Lytro dataset [35]. (a) Source 1, (b) Source 2, (c) GBM [36], (d) NSCT [14], (e) ICA [11], (f) DCHWT [17], (g) ASR [12], (h) IFCNN [30], (i) DenseFuse [31], (j) acof [37], (k) CFL [38], (l) ConvCFL [39], (m) DTNP [40], (n) MLCF [41], (o) Joint [42], (p) CRFGuided.

4.3. Complexity

We analyzed the computational complexity of the proposed and compared image fusion methods. The average execution time on the Lytro dataset of the compared methods are included in Table 3. The included times were computed on an Intel® Core™ i9 2.9GHz processor with 16 GB RAM and a 64-bit operating system. IFCNN [30] and DenseFuse [31] were executed on an NVIDIA GeForce RTX 2080 with Max-Q Design.

Table 3. Average running time of compared methods for input image pairs of size [520 × 520].

Methods	Time (s)
GBM [36]	2.43 s
NSCT [14]	87.27 s
ICA [11]	24.02 s
DCHWT [17]	18.59 s
ASR [12]	1204.92 s
IFCNN [30]	0.22 s
DenseFuse [31]	0.41 s
acof [37]	9.91 s
CFL [38]	23.69 s
ConvCFL [39]	138.42 s
DTNP [40]	420 s
MLCF [41]	53.11 s
Joint [42]	83.09 s
CRF-Guided	31.00 s

The two deep learning-based approaches IFCNN and DenseFuse have very small execution times, due to their parallel implementation on a GPU. The remaining methods were implemented on MATLAB v2021b. The proposed CRF-Guided was implemented on MATLAB without any code optimization. Nonetheless, its average execution time of 31 s compares favorably with the fastest methods, being the 6th method (excluding the IFCNN and DenseFuse), but with the best overall qualitative performance. Thus, the best qualitative fortunately implies a medium computational complexity.

5. Conclusions

A novel transform domain multi-focus image fusion method is introduced in this paper. The proposed CRF-Guided fusion takes advantage of the CRF minimization and the labels are used to guide the fusion of both low frequency and the ICA transform coefficients and thus the high frequency. CRF-Guided fusion supports image denoising during fusion, by applying coefficient shrinkage. Quantitative and qualitative evaluation demonstrate that CRF-Guided fusion outperforms state-of-the-art multi-focus image fusion methods. Limitations of the proposed CRF-Guided fusion method include the selection of the transform domain and the hand-crafted design of the unary and smoothness potential functions for the energy minimization problem. Future work includes the application of CRF-Guided fusion in different transform domains and learning the unary and smoothness potential function with deep learning networks.

Author Contributions: Conceptualization, O.B.; methodology, O.B.; software, O.B.; validation and writing, O.B.; reviewing and editing, I.A. and N.M.; supervision, I.A. and N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [\[CrossRef\]](#)
2. Bai, X.; Zhang, Y.; Zhou, F.; Xue, B. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Inf. Fusion* **2015**, *22*, 105–118. [\[CrossRef\]](#)
3. Zhang, Y.; Bai, X.; Wang, T. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure. *Inf. Fusion* **2017**, *35*, 81–2535. [\[CrossRef\]](#)
4. Liu, Y.; Liu, S.; Wang, Z. Multi-focus image fusion with dense SIFT. *Inf. Fusion* **2015**, *23*, 139–155. [\[CrossRef\]](#)
5. Qiu, X.; Li, M.; Zhang, L.; Yuan, X. Guided filter-based multi-focus image fusion through focus region detection. *Signal Process. Image Commun.* **2019**, *72*, 35–46. [\[CrossRef\]](#)
6. Li, M.; Cai, W.; Tan, Z. A region-based multi-sensor image fusion scheme using pulse-coupled neural network. *Pattern Recognit. Lett.* **2006**, *27*, 1948–1956. [\[CrossRef\]](#)
7. Li, S.; Kang, X.; Hu, J.; Yang, B. Image matting for fusion of multi-focus images in dynamic scenes. *Inf. Fusion* **2013**, *14*, 147–162. [\[CrossRef\]](#)
8. Singh, S.; Singh, H.; Mittal, N.; Hussien, A.G.; Sroubek, F. A feature level image fusion for Night-Vision context enhancement using Arithmetic optimization algorithm based image segmentation. *Expert Syst. Appl.* **2022**, *209*, 118272. [\[CrossRef\]](#)
9. Singh, S.; Mittal, N.; Singh, H. A feature level image fusion for IR and visible image using mNMRA based segmentation. *Neural Comput. Appl.* **2022**, *34*, 8137–8154. [\[CrossRef\]](#)
10. Hyvärinen, A.; Hurri, J.; Hoyer, P.O. Independent Component Analysis. In *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*; Springer: London, UK, 2009; pp. 151–175.
11. Mitianoudis, N.; Stathaki, T. Pixel-based and region-based image fusion schemes using ICA bases. *Inf. Fusion* **2007**, *8*, 131–142. [\[CrossRef\]](#)
12. Liu, Y.; Wang, Z. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process.* **2015**, *9*, 347–357. [\[CrossRef\]](#)

13. Liu, Y.; Chen, X.; Ward, R.K.; Wang, Z.J. Image Fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* **2016**, *23*, 1882–1886. [[CrossRef](#)]
14. Zhang, Q.; Guo, B.I. Multifocus image fusion using the nonsubsampling contourlet transform. *Signal Process.* **2009**, *89*, 1334–1346. [[CrossRef](#)]
15. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
16. Zhou, Z.; Li, S.; Wang, B. Multi-scale weighted gradient-based fusion for multi-focus images. *Inf. Fusion* **2014**, *20*, 60–72. [[CrossRef](#)]
17. Shreyamsha Kumar, B.K. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal Image Video Process.* **2013**, *7*, 1125–1143. [[CrossRef](#)]
18. Qin, X.; Ban, Y.; Wu, P.; Yang, B.; Liu, S.; Yin, L.; Liu, M.; Zheng, W. Improved Image Fusion Method Based on Sparse Decomposition. *Electronics* **2022**, *11*, 2321. [[CrossRef](#)]
19. Jagtap, N.S.; Thepade, S.D. High-quality image multi-focus fusion to address ringing and blurring artifacts without loss of information. *Vis. Comput.* **2021**, *37*, 1–9. [[CrossRef](#)]
20. Singh, H.; Cristobal, G.; Bueno, G.; Blanco, S.; Singh, S.; Hrisheekesha, P.N.; Mittal, N. Multi-exposure microscopic image fusion-based detail enhancement algorithm. *Ultramicroscopy* **2022**, *236*, 113499. [[CrossRef](#)]
21. Bouzos, O.; Andreadis, I.; Mitianoudis, N. Conditional random field model for robust multi-focus image fusion. *IEEE Trans. Image Process.* **2019**, *28*, 5636–5648. [[CrossRef](#)]
22. Chai, Y.; Li, H.; Li, Z. Multifocus image fusion scheme using focused region detection and multiresolution. *Opt. Commun.* **2011**, *284*, 4376–4389. [[CrossRef](#)]
23. He, K.; Zhou, D.; Zhang, X.; Nie, R. Multi-focus: Focused region finding and multi-scale transform for image fusion. *Neurocomputing* **2018**, *320*, 157–170. [[CrossRef](#)]
24. Singh, S.; Singh, H.; Gehlot, A.; Kaur, J.; Gagandeep, A. IR and visible image fusion using DWT and bilateral filter. *Microsyst. Technol.* **2022**, *28*, 1–11. [[CrossRef](#)]
25. Singh, S.; Mittal, N.; Singh, H. Multifocus image fusion based on multiresolution pyramid and bilateral filter. *IETE J. Res.* **2020**, *68*, 2476–2487. [[CrossRef](#)]
26. Zhang, X. Deep learning-based Multi-focus image fusion: A survey and a comparative study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4819–4838. [[CrossRef](#)]
27. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* **2017**, *36*, 191–207. [[CrossRef](#)]
28. Amin-Naji, M.; Aghagolzadeh, A.; Ezoji, M. Ensemble of CNN for multi-focus image fusion. *Inf. Fusion* **2019**, *51*, 201–214. [[CrossRef](#)]
29. Tang, H.; Xiao, B.; Li, W.; Wang, G. Pixel convolutional neural network for multi-focus image fusion. *Inf. Sci.* **2018**, *433–434*, 125–141. [[CrossRef](#)]
30. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [[CrossRef](#)]
31. Li, H.; Wu, X.J. DenseFuse: A fusion approach to infrared and visible images. *IEEE Trans. Image Process.* **2019**, *28*, 2614–2623. [[CrossRef](#)]
32. Ma, X.; Wang, Z.; Hu, S.; Kan, S. Multi-focus image fusion based on multi-scale generative adversarial network. *Entropy* **2022**, *24*, 4819–4838. [[CrossRef](#)] [[PubMed](#)]
33. Wei, B.; Feng, X.; Wang, K.; Gao, B. The multi-focus-image-fusion method based on convolutional neural network and sparse representation. *Entropy* **2021**, *23*, 827. [[CrossRef](#)] [[PubMed](#)]
34. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [[CrossRef](#)]
35. Nejati, M.; Samavi, S.; Shirani, S. Multi-focus image fusion using dictionary-based sparse representation. *Inf. Fusion* **2015**, *25*, 72–84. [[CrossRef](#)]
36. Paul, S.; Sevcenco, I.S.; Agathoklis, P. Multi-exposure and multi-focus image fusion in gradient domain. *J. Circuits Syst. Comput.* **2016**, *25*, 1650123. [[CrossRef](#)]
37. Zhu, R.; Li, X.; Huang, S.; Zhang, X. Multimodal medical image fusion using adaptive co-occurrence filter-based decomposition optimization model. *Bioinformatics* **2021**, *38*, 818–826. [[CrossRef](#)]
38. Veshki, F.G.; Ouzir, N.; Vorobyov, S.A.; Ollila, E. Multimodal image fusion via coupled feature learning. *Signal Process.* **2022**, *200*, 108637. [[CrossRef](#)]
39. Veshki, F.G.; Vorobyov, S.A. Coupled Feature Learning Via Structured Convolutional Sparse Coding for Multimodal Image Fusion. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 2500–2504.
40. Li, B.; Peng, H.; Wang, J. A novel fusion method based on dynamic threshold neural P systems and nonsubsampling contourlet transform for multi-modality medical images. *Signal Process.* **2021**, *178*, 107793. [[CrossRef](#)]
41. Tan, W.; Thitøn, W.; Xiang, P.; Zhou, H. Multi-modal brain image fusion based on multi-level edge-preserving filtering. *Biomed. Signal Process. Control.* **2021**, *64*, 102280. [[CrossRef](#)]

42. Li, X.; Zhou, F.; Tan, H. Joint image fusion and denoising via three-layer decomposition and sparse representation. *Knowl.-Based Syst.* **2021**, *224*, 107087. [[CrossRef](#)]
43. Singh, S.; Mittal, N.; Singh, H. Review of various image fusion algorithms and image fusion performance metric. *Arch. Comput. Methods Eng.* **2021**, *28*, 3645–3659. [[CrossRef](#)]
44. Singh, S.; Mittal, N.; Singh, H. Classification of various image fusion algorithms and their performance evaluation metrics. In *Computational Intelligence for Machine Learning and Healthcare Informatics*; De Gruyter: Berlin, Germany, 2020; pp. 179–198.
45. Hossny, M.; Nahavandi, S.; Creighton, D. Comments on ‘Information measure for performance of image fusion’. *Electron. Lett.* **2008**, *44*, 1066–1067. [[CrossRef](#)]
46. Xydeas, C.S.; Petrovic, V. Objective image fusion performance measure. *Electron. Lett.* **2000**, *36*, 308–309. [[CrossRef](#)]
47. Xydeas, C.S.; Petrovic, V.S. Objective pixel-level image fusion performance measure. In *Proceedings of the Sensor Fusion: Architectures, Algorithms, and Applications IV*, Orlando, FL, USA, 24–28 April 2000; SPIE: Bellingham, DC, USA, 2000; Volume 4051, pp. 89–98. [[CrossRef](#)]
48. Yang, C.; Zhang, J.Q.; Wang, X.R.; Liu, X. A novel similarity based quality metric for image fusion. *Inf. Fusion* **2008**, *9*, 156–160. [[CrossRef](#)]
49. Chen, Y.; Blum, R.S. A new automated quality assessment algorithm for image fusion. *Image Vis. Comput.* **2009**, *27*, 1421–1432. [[CrossRef](#)]
50. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
51. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” image quality analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [[CrossRef](#)]