Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain

Sujoy Paul†§, Ioana S. Sevcenco*‡ and Panajotis Agathoklis‡||

†Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, U.S.A.
‡Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8P 5C2, Canada
§paul.sujoy.ju@gmail.com
*ioana.sevcenco@ece.uvic.ca
||pan@ece.uvic.ca

Received 20 August 2014
Accepted 4 April 2016
Published 8 June 2016

A multi-exposure and multi-focus image fusion algorithm is proposed. The algorithm is developed for color images and is based on blending the gradients of the luminance components of the input images using the maximum gradient magnitude at each pixel location and then obtaining the fused luminance using a Haar wavelet-based image reconstruction technique. This image reconstruction algorithm is of $O(N)$ complexity and includes a Poisson solver at each resolution to eliminate artifacts that may appear due to the nonconservative nature of the resulting gradient. The fused chrominance, on the other hand, is obtained as a weighted mean of the chrominance channels. The particular case of grayscale images is treated as luminance fusion. Experimental results and comparison with other fusion techniques indicate that the proposed algorithm is fast and produces similar or better results than existing techniques for both multi-exposure as well as multi-focus images.

Keywords: Multi-focus image fusion; multi-exposure image fusion; gradient domain image fusion; image reconstruction from gradients.

1. Introduction

In applications such as computer vision, medical imagery, photography and remote sensing, there is a need for algorithms to merge the information acquired by either single or multiple image sensors at the same or different time instants. Generally speaking, image fusion integrates information from a stack of images into a single

*This paper was recommended by Regional Editor Kshirasagar Naik.
†Corresponding author.
image that has more details than the individual images it is made of. In static image
fusion, it is assumed that the input images are aligned and there exist no differences
in terms of depth or viewpoint of the imaged scenes. In dynamic image fusion, the
imaged scenes in the input images contain some perturbations and are not exactly
the same in terms of depth or viewpoint. Many researchers\textsuperscript{1,2} tend to first identify the
perturbations and then align all the images by image registration to produce a static
sequence of images having similar geometry. After registration, the algorithms for
static fusion can be applied to these images. There are some algorithms\textsuperscript{3} in which the
two steps of registration and fusion are integrated. Such algorithms can handle
motion of some of the objects in the source images provided that the position of the
camera is kept constant.

Static image fusion algorithms can be classified in terms of the way in which the
image is processed into pixel-based\textsuperscript{5,4} or region-based algorithms.\textsuperscript{5,6} In pixel-based
methods, the simplest way of obtaining a fused image is by taking a weighted sum at
each pixel location, with weights depending on some function of the input images. In
region-based techniques, the input images are represented in a multi-resolution
framework, using pyramid or wavelet transforms, and then operations such as taking
the maximum or averaging the resulting coefficients are used to integrate the in-
formation into a more comprehensive image. Naidu\textsuperscript{7} proposed a multi-resolution
singular value decomposition (SVD)-based image fusion technique. The images to be
fused are decomposed into approximation and detail coefficients, a similar structure
to that of wavelet decomposition. Then, at each decomposition level, the largest
absolute values of the detail coefficients are selected and an average of the approx-
imation coefficients is used to obtain the fused image. Zheng \textit{et al.}\textsuperscript{8} proposed a fusion
rule based on principal component analysis (PCA) for multi-scale decomposed
images. Lewis \textit{et al.}\textsuperscript{9} presented a comparative study of pixel- and region-based
fusions and indicated that for most cases the region-based techniques provide better
results.

Image fusion can be applied to multi-focus or multi-exposure images. In the multi-
focus case, the input images are those in which only some portion of the image is well
focused, whereas other portions appear blurred. Haghighat \textit{et al.}\textsuperscript{10} proposed a multi-
focus image fusion technique that operates in the discrete cosine transform (DCT)
domain. They compute the variance of the $8 \times 8$ DCT coefficients of each image, and
the fused blocks are those having the highest variance of DCT coefficients. Song
\textit{et al.}\textsuperscript{11} proposed a wavelet decomposition-based algorithm for multi-focus image fu-
sion. They fuse the wavelet coefficients using an activity measure which depends on
the gradients of the wavelet coefficients. A multiresolution approach was also adopted
in the algorithms developed by Li and Wang in Ref.\textsuperscript{12} and by Biswas \textit{et al.} in Ref.\textsuperscript{13}.
A survey on multi-focus image fusion techniques can be found in Ref.\textsuperscript{14}. More recent
research\textsuperscript{15,16} makes use of edge detection techniques for color image fusion.

In the multi-exposure case, the input images have different exposures. These
images have details only in a part of the image while the rest of the image is either
under- or over-exposed. Fusion of such images is done to integrate the details from all images into a single, more comprehensive result. Mertens et al.\textsuperscript{17} proposed such an algorithm, in which the images are decomposed into Laplacian pyramids and then they are combined at each level using weights depending on the contrast, saturation and well-exposedness of the given images. A technique for image contrast enhancement using image fusion has been presented in Ref. \textsuperscript{18} and is similar to Ref. \textsuperscript{17}. In Ref. \textsuperscript{18}, the input images to the fusion algorithm are obtained from the original image after applying local and/or global enhancements. Shen et al.\textsuperscript{19} use a probabilistic model based on local contrast and color consistency to combine multi-exposure images. Li et al.\textsuperscript{3} fuse the multi-exposure images using a weighted sum methodology based on local contrast, brightness and color dissimilarity. They use a pixel-based method instead of a multi-resolution approach to increase the speed of execution. In Ref. \textsuperscript{20}, the input images are first divided into blocks and the blocks corresponding to maximum entropy are used to obtain the fused image. The genetic algorithm (GA) is used to optimize block size, and this may require a considerable amount of time to converge.

Image fusion in the gradient domain has been recently studied by some researchers. Socolinsky and Wolff\textsuperscript{21} proposed an image fusion approach which integrates information from a multi-spectral image dataset to produce a one band visualization of the image. They generalize image contrast, which is closely related to image gradients, by defining it for multi-spectral images in terms of differential geometry. They use this contrast information to reconstruct the optimal gradient vector field, to produce the fused image. Later, Wang et al.\textsuperscript{22} fused the images in gradient domain using weights dependent on local variations in intensity of the input images. At each pixel position, they construct an importance-weighted contrast matrix. The square root of the largest eigenvalue of this matrix yields the fused gradient magnitude, and the corresponding eigenvector gives the direction of the fused gradient. Recently, Hara et al.\textsuperscript{23} used an inter image weighting scheme to optimize the weighted sum of the gradient magnitude and then reconstruct the fused gradients to produce the fused image. The optimization step tends to slow down this technique. Additionally, their technique comprises a manually thresholded intra image weight saliency map, requiring user intervention. An interesting block-based approach was recently proposed by Ma and Wang in Ref. \textsuperscript{24}. This approach is unique in the way in which it processes color images. Specifically, the RGB color channels of an image are processed together, and instead the images are split into three “conceptually independent components: signal strength, signal structure and mean intensity”.\textsuperscript{24} This idea was inspired by the increasingly popular structural similarity (SSIM) index,\textsuperscript{25} developed by the same main author as an objective measure of similarity between two images.

In this paper, a gradient-based image fusion algorithm is proposed. The algorithm proposed here works for the fusion of both color as well as grayscale images. In the case of color images, one of the key ideas of the fusion algorithm proposed here is that
it treats the luminance and chrominance channels of the images to be fused in a
different manner. This different treatment of the channels is motivated by the fact
that the luminance channel contains a major part of information about image details
and contrast, whereas the chrominance channels contain only color information, to
which the human visual system is less sensitive. The fusion of the luminance channels
is done in the gradient domain, by taking the gradients with the maximal magnitude
of the input images at each pixel location. The luminance channel of the fused image
is then obtained by integrating the fused gradients. This done by using a wavelet-
based method,\textsuperscript{26} which includes a Poisson solver\textsuperscript{27} at each resolution. This algorithm
is known\textsuperscript{28} to produce good results, free from artifacts, when the gradient field is a
nonconservative field, as is the case when gradients of different images are combined.
Next, for the chrominance part of the color images, fusion is done by taking a
weighted sum of the input chrominance channels, with the weights depending on the
channel intensities, which conveys information about color. Grayscale images may be
dealt with in the same way as the luminance component of color images. The pro-
posed algorithm can be applied for multi-exposure as well as multi-focus images.

The rest of the paper is organized as follows. In Sec. 2, the proposed algorithm is
presented. In Sec. 3, experimental results and comparisons with other image fusion
algorithms are presented. Finally, in Sec. 4, the main conclusions are drawn.

2. Image Fusion in Gradient Domain

In this section, a new image fusion algorithm is proposed. The proposed algorithm
can be applied to fuse together a sequence of either color or grayscale images
(minimum two images). A flowchart of the algorithm in its most general case (i.e.,
fusion of multiple color images) is illustrated in Fig. 1.

The proposed algorithm operates in the YCbCr color space.\textsuperscript{a} The luminance (Y)
channel represents the image brightness information and it is in this channel where
variations and details are most visible, since the human visual system is more sen-
sitive to luminance (Y) than to chrominance (C_b, C_r). This important observation
has two main consequences for the proposed fusion algorithm. Firstly, it indicates
that the fusion of the luminance and chrominance channels should be done in a
different manner, and that it is in the luminance channel where the most advanced
part of the fusion is to be performed. Secondly, it reveals that the same procedure
used for the luminance channels fusion can be used to fuse single channel images (i.e.,
images in grayscale representation).

In what follows, the proposed luminance fusion technique is described, followed by
chrominance fusion.

\textsuperscript{a}Rec. ITU-R BT.601-5, Studio encoding parameters of digital television for standard 4:3 and wide-screen
2.1. Luminance fusion

As mentioned in the previous sections, the luminance fusion can be carried out on grayscale images, or on color images that are in the YCbCr color coordinate system. If the input images are in RGB representation, conversion to YCbCr should be performed first.
Luminance fusion is performed in the gradient domain. This domain choice is motivated by the fact that the image gradient depicts information on detail content, to which the human visual system is more sensitive under certain illumination conditions. For example, a blurred, over- or under-exposed region in an image will have a much lower gradient magnitude of the luminance channel than the same region in an image with better focus or exposure. This observation implies that taking the gradients with the maximal magnitude at each pixel position will lead to an image which has much more detail than any other image in the stack.

Let the luminance channels of a stack of $N$ input images be $I' = \{I_1, I_2, \ldots, I_N\}$, where $N \geq 2$. According to a commonly employed discretization model, the gradient of the luminance channel of the $n$th image in the stack may be defined as:

$$\Phi^x_n(x,y) = I_n(x+1,y) - I_n(x,y), \quad (1)$$

$$\Phi^y_n(x,y) = I_n(x,y+1) - I_n(x,y), \quad (2)$$

where $\Phi^x_n$ and $\Phi^y_n$ are the gradient components in the $x$- and $y$-directions. The magnitude of the gradient may be defined as

$$H_n(x,y) = \sqrt{\Phi^x_n(x,y)^2 + \Phi^y_n(x,y)^2}. \quad (3)$$

Let the image number having the maximum gradient magnitude at the pixel location $(x,y)$ be $p(x,y)$. It may be mathematically represented as

$$p(x,y) = \max_{1 \leq n \leq N} H_n(x,y). \quad (4)$$

Using (4), the fused luminance gradient may be represented as

$$\Phi^x(x,y) = \Phi^x_{p(x,y)}(x,y), \quad (5)$$

$$\Phi^y(x,y) = \Phi^y_{p(x,y)}(x,y), \quad (6)$$

where $\Phi^x_{p(x,y)}(x,y)$ and $\Phi^y_{p(x,y)}(x,y)$ denote the values of the $x$ and $y$ gradient components of the image with index $p(x,y)$, at pixel position $(x,y)$. So, the fused luminance gradient is $\Phi = [\Phi^x, \Phi^y]^T$. It may be noted that the fused luminance gradient has details from all the luminance channels from the stack and in order to get the fused luminance channel, reconstruction is required from the gradient domain. The relationship between the fused gradient ($\Phi$) and the fused luminance channel ($I$) may be represented as

$$\nabla I = \Phi, \quad (7)$$

where $\nabla = [d/dx, d/dy]^T$. We need to solve for $I$ in (7) in order to get the fused luminance, which may not have a solution if the fused gradient violates the zero curl condition. This is due to the fact that the fused gradient is not the gradient of a single
luminance channel, but it is a combination of several luminance gradients. Thus it may not be a conservative field, or in other words, integrals along a closed path may be nonzero. A common approach to solve this problem is to formulate the reconstruction as a $l^2$ optimization problem, which leads to solving the Poisson equation,

$$\nabla^2 I = \nabla^T \Phi. \tag{8}$$

One way to solve Eq. (8) numerically is by using a large system of linear equations. Some other researchers project the given gradient to another space, in which the zero curl condition is enforced. In Ref. 32, a method for gradient integration is presented, where the least square objective function for surface reconstruction is expressed in terms of matrix algebra and it is shown that the minimizer can be obtained as the solution to a Lyapunov equation. In this paper, a gradient reconstruction technique by Sevcenco et al. is used. This technique is inspired by Hampton et al. and is based on the Haar wavelets. The basic idea of this reconstruction algorithm is the relationship between the Haar wavelet filters and the gradient model. In this technique, the Haar wavelet decomposition coefficients of the luminance channel can be directly computed from the fused luminance gradient. Then, synthesis of these coefficients is done to produce the fused luminance channel. During synthesis, an iterative Poisson solver based on (9) is used at each resolution level to overcome the artifacts that might occur due to the fact that the fused gradients do not satisfy the zero curl condition, as they are from different luminance gradients. The recursion formula may be represented as

$$I(k + 1) = I(k) - \frac{1}{4} \left( \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix} \otimes I(k) + \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \otimes \Phi^x(k) \right) + \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \otimes \Phi^y(k), \tag{9}$$

where $k$ is the iteration index and $\otimes$ represents 2D convolution. A very small number of iterations are required at each resolution, because a good initial point is provided thus leading to fast convergence. This reconstruction algorithm is based on a modified version of the wavelet transform and, as a result of this, has a low complexity $O(N)$, where $N$ is the number of samples in the signal to be reconstructed. A detailed discussion regarding the computational complexity can be found in Sec. III B of Ref. 33.

After obtaining the image from the gradient domain, some pixels may have intensity values outside the standard range of the luminance component (16/235). This is due to the fact that the fused gradient is obtained by merging multiple image gradients, and as a result, high differences between neighboring gradient values exist,
possibly leading to a reconstructed image with a high dynamic range of pixel intensities. A linear mapping of the pixel intensities of the reconstructed luminance channel can be done such that the resultant intensities lie within the required range. The caveat of this approach, however, is that it leads to a loss of contrast. For this reason, a nonlinear mapping similar to gamma correction is used. The resultant image may be obtained using

\[ I(i, j) = \left( \frac{I(i, j) - \min_{i,j} I(i, j)}{\max_{i,j} I(i, j) - \min_{i,j} I(i, j)} \right)^\gamma \times R_C + L, \]

where \( \gamma = \log_e(R_C)/\log_e(R_I), \) \( R_I \) is the range of values present in the reconstructed luminance component, \( R_C = H - L \) and \( H \) and \( L \) are the maximum and minimum intensity values in the channel. In the case of the luminance component of a color image, \( H = 235 \) and \( L = 19 \), thus \( R_C = 216 \). After using Eq. (10), the details in the image are preserved and the result does contain more details than the input images.

At the end, local histogram equalization is applied on the luminance component. This is done in order to distribute the intensities properly throughout the entire range of display. It may be noted that grayscale images can be fused in the same way as the luminance component of a color image. In case of grayscale images, \( H = 255, \) \( L = 0, \) thus \( R_C = 255. \)

### 2.2. Chrominance fusion

Chrominance fusion is to be carried out for the fusion of the input chrominance channels of color images in YCbCr representation (i.e., grayscale images). If the input images are in RGB representation, conversion to YCbCr should be performed first, to obtain the luminance (\( Y \)) and chrominance (\( C_b, C_r \)) channels representation. If the input images are in single channel (e.g., grayscale representation), this step does not apply.

Inherently different than luminance fusion, chrominance fusion operates directly on chrominance values, as opposed to their gradients. Specifically, the chrominance fusion is done by taking a weighted sum of the input chrominance channels. The values in the chrominance channels have a range from 16/240 and carry information about color. These channels are such that when both \( C_b \) and \( C_r \) are equal to 128, the image is visually similar to a grayscale image, and thus carries the least amount of color information. This motivates selecting the weights for the chrominance channels such that at each pixel position they depend on how far from 128 the chrominance value is. Let us denote the chrominance channels of the input images by \( C'_b = \{C'_b, C'_b, \ldots, C'_b\} \) and \( C'_r = \{C'_r, C'_r, \ldots, C'_r\}. \)

The fused chrominance channels may be represented as follows:

\[ C_b(i, j) = \sum_{n=1}^{N} w^n_b(i, j). \left( C'^n_b(i, j) - 128 \right) + 128 \]
where

\[ w^n_b(i,j) = \frac{|C^n_b(i,j) - 128|}{\sum_{k=1}^{N} |C^k_b(i,j) - 128|} \]  

(12)

\[ C_r(i,j) = \sum_{n=1}^{N} w^n_r(i,j).(C^n_r(i,j) - 128) + 128 \]  

(13)

where

\[ w^n_r(i,j) = \frac{|C^n_r(i,j) - 128|}{\sum_{k=1}^{N} |C^k_r(i,j) - 128|} \]  

(14)

where \(|\cdot|\) returns the absolute value. If all chrominance values at a pixel position in all images from the stack are equal to 128, the corresponding weights will be zero. It may be noted that the fusion of the chrominance channels done by Eqs. (11)–(14) is a pixel-based approach, and is thus less computationally intensive than luminance fusion, which is gradient-based.

3. Performance Evaluation and Comparison

In this section, the performance evaluation of the proposed algorithm on different types of images is presented. The results are compared with the ones of four other image fusion algorithms, namely — DCT, SVD, multi-exposure fusion (MEF) and the gradient weighting (GrW) method. The source codes of the DCT, SVD and MEF methods are available at Refs. 35–37, respectively. The input images used in the comparison are grouped into four different classes according to the type of fusion performed (i.e., multi-focus and multi-exposure, grayscale and color) and are presented in the following subsections.

The performance analysis begins with a visual comparison of the results produced by each of the studied algorithms. In passing we note that, to the best of our knowledge, this kind of evaluation (i.e., subjective evaluation) continues to dominate the chart of quality assessment measures for image fusion algorithms. The use of objective measures will be discussed later. The codes for the algorithm proposed in this paper are available at Ref. 38.

3.1. Multi-focus grayscale images

Clock and pepsi, presented in Figs. 2 and 3, are the two multi-focus grayscale images used for comparison. The fused results produced by the proposed algorithm are presented in Figs. 2(c) and 3(e). For visual comparison, we consider the results using two methods presented in the literature for multi-focus grayscale images, the DCT (Figs. 2(c) and 3(c)) and SVD (Figs. 2(d) and 3(d)) methods. It may be noted in Fig. 2(f), that the DCT method produces undesirable blocking artifacts. The SVD method also produces artifacts that are more clearly visible in Fig. 2(h), on the
zoomed in object edges. On the other hand, the proposed algorithm produces a good fusion of the two multi-focus images and is free from visual artifacts.

### 3.2. Multi-focus color images

Figure 4 presents an example of multi-focus fusion done with the proposed method for a color image named *foot*. None of the four algorithms used here for comparison is proposed by their authors for multi-focus color images and thus the proposed method is not compared to any of them.
3.3. Multi-exposure grayscale images

Two multi-exposure grayscale images named igloo\textsuperscript{40} and monument\textsuperscript{41} are presented in Figs. 5 and 6. The fused results of the proposed algorithm are presented in Figs. 5(h) and 6(e), respectively. GrW\textsuperscript{23} is an algorithm for image fusion, where the
authors have used multi-exposure grayscale images to test their method. It is a gradient domain fusion method and requires reconstruction to get the fused image. As the authors of the GrW algorithm have not mentioned any specific method for reconstruction, the wavelet-based reconstruction procedure\textsuperscript{26} has been used to yield the fused image. The saliency map used by the authors of GrW is not used here, because no automated way of selecting the threshold for the map has been mentioned in their paper. The fused results produced by the GrW method are presented in Figs. 5(g) and 6(d). It may be observed from Fig. 5 that the details inside the igloo building are more visible in the result produced using the method proposed in this paper than in the one produced by the GrW method. Again, in Fig. 6, the sky-cloud portion is more visible in the image fused by the proposed algorithm than in the image fused by the GrW method.
3.4. Multi-exposure color images

Door\textsuperscript{37} and house\textsuperscript{37} are the two multi-exposure color images presented in Figs. 7 and 8. It may be observed that for the door image, the details within the door are not visible in the first input image and the details outside the door are not visible in the last input image. The proposed algorithm fuses all input images properly, as may be

---

Fig. 6. (a)–(c) are the input images (monument). (d) and (e) are the images fused by GrW and the proposed algorithm, respectively.
observed from the results presented in Figs. 7(h) and 8(f) for the door and house images, respectively. A technique for MEF for color images presented in the literature is MEF\textsuperscript{17}. This method uses a saturation measure defined only for color images. The results produced by the MEF method are presented in Figs. 7(g) and 8(e). It can be seen that the MEF algorithm performs in a similar fashion to the proposed method.

In addition to visual comparison, efforts have been made for quantitative comparison using objective metrics. To the best of our knowledge, in literature there exists no objective quality measure to evaluate the results of image fusion techniques. One of the main reasons behind this appears to be the fact that in most frameworks there exists no ideal fused image that can be used as benchmark. This
has led researchers to develop metrics like edge content (EC),\(^{18,23}\) second order entropy (SOE),\(^{18}\) blind image quality (BIQ),\(^{42}\) and others. These metrics do not require an ideal fused image for comparison, but are prone to give misleading results in the presence of noise and/or blocking artifacts. For example, EC is an average measure of the gradient magnitudes of an image and methods producing blocking artifacts lead to higher EC values. Similar problems are associated with SOE and BIQ, as they are both variations of information and entropy of the image. Thus, we have refrained from comparing the results quantitatively using such metrics.

Comparison with respect to computational time is presented in Table 1 (using Intel® Core™ i3-3110M @ 2.4 GHz and 4 GB RAM). It should be noted that the time presented in the table is normalized with respect to the total number of pixels present in the image and an average over 100 executions of each algorithm. It can be observed that for all images considered, the proposed algorithm consumes the least computational time with respect to the other methods. The filled in entries indicate the average execution times offered by the analyzed algorithms. Specifically, the filled in entries in the DCT and SVD columns represent the times needed to fuse grayscale multi-focus images, whereas the filled in entries in the MEF and GrW columns are the times needed to fuse color multi-exposure images, in agreement with the authors’ intended applications. The filled in entries in the “proposed method” column represent the average times it took the proposed algorithm to perform grayscale, color, multi-focus or MEF tasks in the same configuration. The entries left

![Fig. 8. (a)–(d) are the input images (house). (e) and (f) are the images fused by MEF and the proposed algorithm, respectively.](image-url)
blank in Table 1 indicate that the respective method was not applied for the respective task.

The results presented in this section indicate that the proposed method performs well for both multi-focus and multi-exposure images, for color as well as for grayscale images. It consistently leads to better and faster results over the other analyzed methods.

4. Conclusion

In this paper, a new gradient-based image fusion algorithm is proposed. Fusion of luminance and chrominance channels is dealt with differently. The fusion of the luminance channel is done in the gradient domain and the fused luminance is obtained using a wavelet-based gradient integration algorithm. The fusion of the chrominance channels is based on a weighted sum of the chrominance channels of the input images. The efficiency of the gradient reconstruction algorithm with complexity $O(N)$ and the simplicity of the chrominance fusion leads to a fast execution speed. Experiments indicate that the proposed algorithm leads to very good results for both multi-exposure as well as multi-focus images.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and MITACS.

References

2. G. Xiao, K. Wei and Z. Jing, Improved dynamic image fusion scheme for infrared and visible sequence based on image fusion system, *Int. Conf. Information Fusion* (Cologne, 2008), pp. 1–6.

1650123-18