Photography image enhancement by image fusion

Marcin Denkowski[*], Paweł Mikołajczak

Department of Information Technology, Institute of Computer Science, Maria Curie Skłodowska University, pl. M. Curie-Skłodowskiej 5, 20-031 Lublin, Poland

Abstract

This paper presents an overview of two-dimensional image fusion methods using convolution filters and discrete wavelet transform developed by the authors. Image fusion is a process of combining two or more images of the same object into one extended composite image. It can extract features from source images and provide more information than one image. This technique can be easily used for image restoration and enhancement. In this article the authors focus on multi-exposure images, high dynamic range improvement and depth-of-field enhancement.

1. Introduction

Image fusion is a process of combining a set of images of the same scene into one composite image. The main objective of this technique is to obtain an image that is more suitable for visual perception. This composite image has reduced uncertainty and minimal redundancy while the essential information is maximized. In other words, image fusion integrates redundant and complementary information from multiple images into a composite one but also decreases dimensionality. There are many methods discovered and discussed in literature that focus on image fusion. They vary with the aim of application used and can be mainly categorized due to algorithms used into pyramid techniques [1,2], morphological methods [3-5], discrete wavelet transform [6-8] and neural network fusion [9]. Different classification of image fusion involves pixel, feature and symbolic levels [10]. Pixel-level algorithms are low level methods and work either in the spatial or transform domain. This kind of algorithms works as a local operation despite transform used and can generate undesirable artifacts. These methods can be enhanced by using multiresolution analysis [1] or by complex wavelet transform [8]. Feature-based methods use segmentation algorithms to divide images into relevant patterns and then combine them to

---

[*]Corresponding author: *e-mail address*: denmar@goblin.umcs.lublin.pl

create output image by using various properties [11]. High-level methods combine image descriptions, typically, in the form of relational graphs [12].

In this work we present an overview of our methods that are capable of obtaining image with enhanced features from a series of digital photographs. Two fusion methods are taken into consideration: Multiresolution Convolution Fusion (MCF) and Discrete Wavelet Transform Fusion (DWTF). We apply these methods for two different aims of image enhancement. In the first case, we enhance depth-of-field (DOF) by fusing multifocus macro-photography images, and in the second case, we create Dynamic Range Increased Image (DRI, HDR) from images taken with different exposure values.

For both fusion aims, as an input, we have created a series of photographs of the same scene, but in the DOF case the images were obtained with different focus lengths and in the DRI case with different exposition. In the first step of our method they have to be registered together to create a properly aligned stack of images. The next step is to fuse them into one composite image. For that purpose we propose the enhanced multiscale convolution and morphology method presented in [13]. As an effect of fusing algorithm we obtain a height map and the reconstructed image with enhanced features.

2. Image alignment

In the first step a set of photographs of the desire object is obtained. Unfortunately, during acquiring a series of photographs small movements of the camera are possible even when using tripods for stabilization. To make the reconstruction method more robust, we can make use of an image registration procedure. The main idea behind the image registration is to find perfect geometric alignment between a set of overlapping images. The quality of match measure represents the matching function parameterized by the geometric transformation. In our method we use the rigid (translations and rotation) or the affine transformation model (rigid + scaling and shears). In most cases it is sufficient to use the simplified rigid transformation (translations only). But when images are acquired without stabilization devices the use of complete affine transformation is a necessity. In our approach we use the normalized mutual information [14] as the matching function:

$$NMI(FI, RI) = \frac{h(FI) + h(RI)}{h(FI, RI)},$$ (1)

where *RI* represents the reference image and *FI* represents the floating image.

$$h(FI) = -\sum p_{FI}(x) \log p(_{FI}(x)),$$
$$h(RI) = -\sum p_{RI}(x) \log (p_{RI}(x)),$$ (2)

$$h(FI,RI) = -\sum\sum p_{FI,RI}(x,y)\log\left(p_{FI,RI}(x,y)\right),$$

where $h(FI)$, $h(RI)$ and $h(FI,RI)$ are the single and joint entropies [14], pFI and pRI are the probabilities of each intensity in the intersection volume of both data sets and pFI,RI is a probability distribution of a joint histogram. For the minimization of the selected similarity measure we use Powell's algorithm [15]. As a result of the registration procedures we obtain a set of geometrically matched images that can be used in the next stages of our reconstruction algorithm.

3. Methods

At this stage we assume that images on the image stack are aligned to each other. In our work we present two methods of fusing images. Both methods are based on the same algorithm scheme and differ only in the decomposition process and reconstruction rules. Fig. 1 shows the diagrams for our Multiresolution Convolution Fusion (MCF) and Discrete Wavelet Transform Fusion (DWTF). Our fusion method is also capable of working with color images.
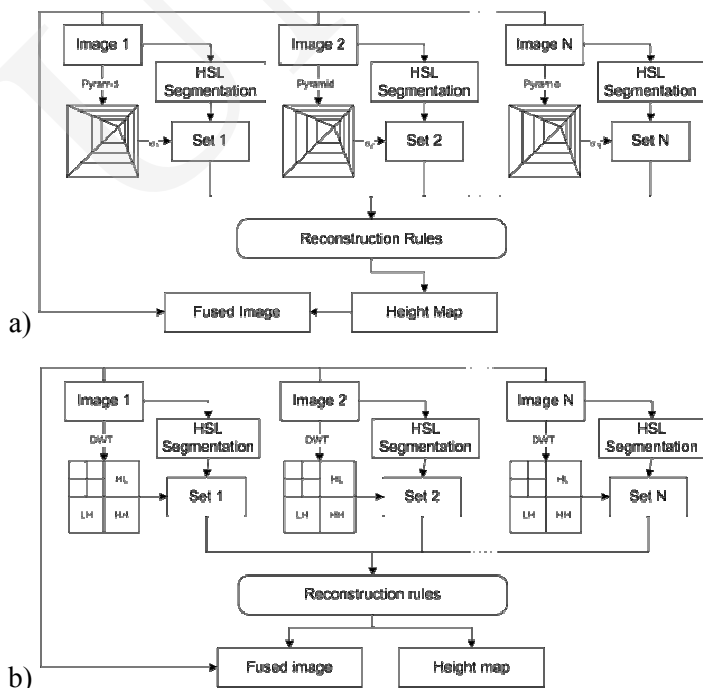

Fig. 1. Image reconstruction scheme using (a) Pyramid Decomposition, (b) Discrete Wavelet Decomposition and HSL Segmentation

The whole algorithm can be divided into 4 stages:

1. Segmentation of every image on the stack by using pyramid segmentation. For this process we convert images into HSL color model [16] to separate luminance (contrast) information contained in the luminance channel from color description in hue and saturation channels.
2. Creation of n-level pyramid for every input image. In the MCF method we use median filter to downscale images instead of high pass and low pass filters, while in the DWTF method we use Discrete Wavelet Transform Decomposition to create appropriate pyramids.
3. Reconstruction rules wich depend on the method used, discussed later in this chapter.
4. Creation of fused image. The value of fused image pixel $f(x, y)$ is equal to the pixel $f^{(z)}(x, y)$ from $z - th$ input image on the stack, where z is a value taken from the created height map $HM(x, y)$.

Creation of segmentation maps gives us spatial information about the whole scene while the Multiscale Convolution Decomposition or Discrete Wavelet Decomposition process gives frequency information. Spatial domain is especially important because the fusion process "likes" to create halo effects near the edges of objects. To resolve this problem we use segmentation maps to determine edges and we are able to mark pixels near these edges properly.

Reconstruction rules for the Multiresolution Convolution Fusion (MCF) method:

Step-1 Calculation of local standard deviation SD at local region R for every pixel $f(x, y)$ at each pyramid level L for every image on the stack (z):

$$SD_R^{(L)}(x, y, z) = \sqrt{\frac{1}{N_R} \sum_{(x,y) \in R, z} \left(f(x, y) - \overline{f_R} \right)^2} \; . \qquad (3)$$

The color RGB components are converted to its graylevel intensity according to $G_f = 0.299R + 0.587G + 0.114B$.

Step-2 For the lowest level of pyramid, pixels with the maximum $SD_{max}^{(0)}(x, y, z)$ are marked as focus and labeled in the height map $HM(x, y)$ with z value. If $abs\left(SD_{max}^{(0)}(x, y) - SD_{max}^{(0)}(x, y) \right) < T_s$, where T_s is a threshold value, pixel is marked as unresolved because it usually belongs to smooth region. These pixels are taken care of at subsequent steps.

Step-3 Every pixel is checked with the segmentation map. If it is not near any edge and its $SD_R(x, y, z)$ value drastically differs from $SD_R(x, y, z)$ of average pixel value for its region R, it is marked with the $SD_R(x, y, z)$ value of the median pixel. It prevents from marking false or noise pixels.

Step-4 For every i-th pyramid level, starting with i=1, if $SD_R^{(i)}(x,y,z)$ of actual pixel is not equal to $SD_R^{(i-1)}(x,y,z)$ from the previous pyramid level, then:

   a) if the pixel is near some edge marked on the segmentation map, the pixel with $max(SD_R^{(i)}(x,y,z),SD_R^{(i-1)}(x,y,z))$ value is taken and labeled in the height map $HM(x, y)$ with (i) or $(i-1)$ value,

   b) else, the height map $HM(x, y)$ is labeled as:

$$HM(x,y) = HM^{(i-1)}(x,y) + \frac{HM^{(i)}(x,y) - HM^{(i-1)}(x,y)}{2}. \tag{4}$$

Step-5 Labeling remaining pixels. If unresolved pixel belongs to region with many other unresolved pixels, it is marked as a background, else the median value from the region is taken.

Reconstruction rules for the Discrete Wavelet Transform Fusion (DWTF) method:

Step-1 Calculation of the priority *P* at the local region *R* in every image *z* and for every pyramid level *(L)*. We use simple activity measure *d(x,y)* taking the absolute value of the detailed wavelet coefficients:

$$d(x,y) = 2HH(x,y) + HL(x,y) + LH(x,y),$$

$$P_R^{(L)}(x,y,z) = \frac{1}{N_R} \sum_{(x,y)\in R} |d_z(x,y)|. \tag{5}$$

Step-2 For every level of pyramid, the coefficients with the maximum $P_{max}^{(L)}(x,y,z)$ are written to the output pyramid. If the ratio $abs\left(P_{max}^{(L)}(x,y,z)/P_{min}^{(L)}(x,y,z)\right) < T_s$, where T_s is a threshold value, the pixel is marked as unresolved because it usually belongs to smooth region. These pixels are taken care of at subsequent steps.

Step-3 Every pixel is checked with the segmentation map. If it is not near any edge and its $P_R(x, y, z)$ value drastically differs from $P_R(x, y, z)$ of average detailed coefficient value for its region *R,* it is marked with a value of the median coefficient. It prevents from marking single false pixels.

Step-4 Labeling remaining pixels. If unresolved pixel belongs to region with many other unresolved pixels, it is marked as a background, else the median value from the region is taken.

Step-5 Inverse Discrete Wavelet Transform compose created output pyramid into the final fused image.

Step-6 Height map is created based on the region priorities P_R chosen at the previous steps.

3.1. Depth-of-field reconstruction

Macro photography is a type of close-up photography with magnification ratios from about 1:1 to about 10:1 [17]. The most crucial parameter of macro photography is the depth of field (DOF). Because it is very difficult to obtain high values of DOF for extreme close-ups, it is essential to focus on the most important part of the subject. Any other elements that are even a millimeter farther or closer may appear blurred in the acquired photo. The depth of field can be defined as the distance in front of and behind the subject appearing in focus. Only a very short range of the photographed subject will appear in the exact focus. Being more precise, for a specific film format, the depth of field is described as a function parameterized by: the focal length of the lens, the diameter of the lens opening (the aperture), and the distance between the subject and the camera. Let D be the distance at which the camera is focused, F the focal length (in millimeters) calculated for an aperture number f and k - the "circle of confusion" for a given film format (in millimeters), then the depth of field (DOF) can be defined as:

$$DOF_{1,2} = \frac{D}{1 \pm \dfrac{1000 \times D \times k \times f}{F^2}} \, , \qquad (6)$$

where DOF_1 is the distance from the camera to the far depth of field limit, and DOF_2 is the distance from the camera to the near depth of field limit. As formula (3) shows, there are three main factors controlling the depth-of-field for a given film format. The aperture controls the effective diameter of the lens opening. Reducing the aperture size increases the depth of field, however, it also reduces the amount of light transmitted (see Figure 2). Lenses with a short focal length have a greater depth-of-field than long lenses. A greater camera-to-subject distance results in a greater depth-of-field. Our goal is to achieve the deepest possible depth-of-field using the standard digital camera images and image processing algorithms. To achieve this, we create a series of macro photograph images of the same subject with different focus lengths. The next step is to register them together. Following this we create a height map using the MCF or DWTF method. In the last step we reconstruct the focused image with a very deep depth-of-field using a generated height map. The reconstructed images are shown in Fig 3. and the numerical differences between these two methods are shown in Table 1.

Table 1. The $Q^{AB/F}$ quality measure for five tested series of images for both methods of image fusion for both aims

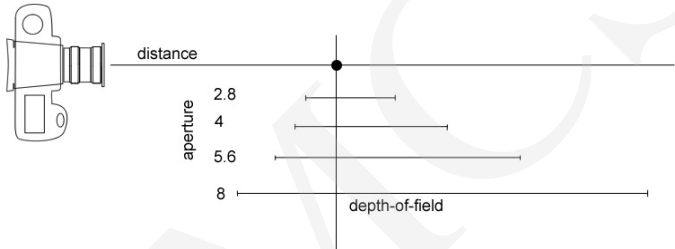|		Series 1	Series 2	Series 3	Series 4	Series 5
DOF	MCF	**0.84**	0.62	**0.79**	**0.67**	0.87
	DWTF	0.76	**0.65**	0.73	0.61	**0.93**
DRI	MCF	0.87	0.48	**0.74**	**0.86**	**0.84**
	DWTF	0.87	**0.53**	0.71	0.84	0.85

Fig. 2. Example diagram of how the f-number (aperture) affects depth-of-field

Fig. 3. Multifocus Image Fusion example sets. Small images are part of the original multifocus image set; last small, blurred image is the Height Map; large image is the fused image

3.2. Dynamic range enhancement

While photographing a scene we obtain a two dimensional array of values taken by a sensor. In digital cameras, both CCD (charge coupled device) and CMOS (complementary metal oxide semiconductor) sensor arrays which image the scene, convert the light into electric charge and process it into electronic signals. These signals are usually represented with 10 or 12 bits resolution and by using nonlinear mapping are converted and stored as colour impressions. We can imagine that while taking photos of sunlit scenes or scenes with shiny materials and artificial light sources, we often receive extreme differences in radiance values that are impossible to be captured entirely by a sensor. We must then decide which part of the scene we want to be exposed properly – the highlights or the darks – losing some information. The obvious reason for that case is a limited range of values that can be received by the sensor. This range is often referred to as a dynamic range. The dynamic range can be defined as a ratio of the lightest and the darkest parts of an image and is usually expressed in the exposure values (EV). Zero EV is defined by the combination of an aperture of f/1 and a shutter speed of 1s at ISO 100. EV is increased by 1 when the amount of light halves. The dynamic range ratio of a typical digital sensor reaches approximately 500:1 (~8EV). However, in a real world these values can reach as much as 100 000:1 (~17EV) (or even more), while human eye perceives 10 000:1 (~14EV). To cover the full dynamic range of that scene, we can take a series of photographs with different exposures. But this creates a problem: how can we combine these separate images into a composite image with full information? There are two most common methods: High Dynamic Range (HDR) and Dynamic Range Increasing (DRI). The HDR image is created by increasing bit depth up to 32 per channel and by merging a set of photographs in such a way that all tonal ranges, from lights to darks, are properly recorded. But that still leaves some problems because common imaging devices like LCD screens or printers are unable to show that tonal range. There must be another method to convert 32 bits to 8 bits resolution available by imaging devices. The second technique, DRI, often called tonal mapping, does not increase bit depth of image but overlaps masked layers created from photos. In this article we focus on the DRI technique. It is worth mentioning that more often these two techniques are just called HDR imaging, which is only simplification.

Our goal in this case is to obtain image where all regions are properly exposed. To achieve this, we create a series of images of the same scene with different values of exposition. The next step is to register them together. Following this we create a mask which designates regions that are in a suitable tonal range. For that purpose, both the MCF and DWTF methods are used. In the last step algorithm creates a composite image by choosing pixels from input

images on the basis of the mask values. The reconstructed images are shown in
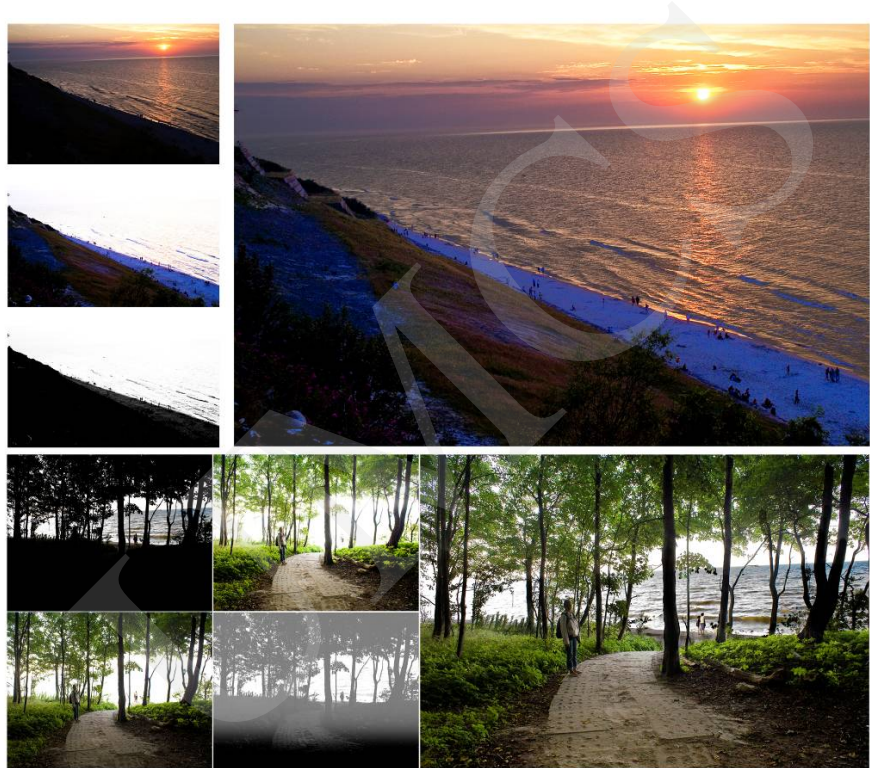Fig 4. and the numerical differences between these two methods are shown in
Table 1.

Fig. 4. Multiexposured Image Fusion example sets. Small images are part of the original
multiexposured image set; the last small, blurred image is the Height Map; the large,
color image is the fused image

4. Results

The proposed methods have been implemented on the Linux platform in C++
language using the SemiVis framework [18]. For the testing procedure we
prepared five image stacks from macro-photography and five stacks for
multiexposure part. The images were photographed using a typical DSLR
camera and preprocessed manually. In all cases, the procedure is performed in
the following order.

At first, the registration process aligns multifocus images to each other to
minimize misregistration. Then all images are segmented and the pyramid is
created up to three levels. Finally, the reconstruction process combines image
stack into the height map and the fused image. The examples of multifocus

images with their height maps and reconstructed fused images with the extended depth-of-field are shown in Fig. 3. Fig. 4 shows the qualitative results of our method for the multiexposured set of images.

Table 1 includes the numerical results of comparison between the two methods of image fusion MCF and DWTF for both cases: DOF and DRI. We utilize the widely used metric $Q^{AB/F}$ that measures quality of the image fusion. This measure was proposed by Xydeas and Petrovič in [19]. In this case, a per-pixel measure of information preservation is obtained between each input and the fused image which is aggregated into a single score $Q^{AB/F}$ using a simple local importance assignment. The results show that there is no significant difference between the two methods proposed in this article, however they, point to a little MCF advantage. It is also noticeable that the DWTF method often creates fused images with more visible noise or artifacts.

Conclusions

This paper presents an attempt to the problem of fusing images from a set of images. The two methods were introduced: Multiresolution Convolution Fusion and Discrete Wavelet Transform Fusion for two applications of image fusion: enhancing depth-of-field and increasing dynamic range of the image.

The presented results are very promising, but as for now, there are still many problems that need to be solved. Future work could include improvements in segmentation and edge detection to help in automatic detection of the background plane. Secondly, there should be more complex methods used to identify smooth regions of objects. We think that in both cases pattern recognition algorithms should improve effectiveness of our method. Also Feature-based fusion methods such as [11] could generate more accurate height maps.

References

[1] Burt P.J., *The pyramid as a Structure for Efficient Computation*. Multiresolution Image Processing and Analysis. Springer-Verlag, Berlin, (1984) 6.
[2] Toet A., *Image Fusion by Rati of Low-Pass Pyramid*. Pattern Recognition Letters, 9(4) (1989) 245.
[3] Ishita D., Bhabatosh C., Buddhajyoti C., *Enhancing Effective Depth-of-Field by Image Fusion Using Mathematical Morphology*. Image and Vision Computing, 24 (2006) 1278.
[4] Mukopadhyay S., Chanda B., *Fusion of 2d Gray Scale Images Using Multiscale Morphology*. Pattern Recognition, 34 (2001) 1939.
[5] Matsopoulos G.K., Marshall S., Brunt J.N.M., *Multiresolution Morphological Fusion of mr and ct Images of the Human Brain*. IEEE Proceedings Vision, Image and Signal Processing, 141(3) (1994) 137.
[6] Li H., Manjunath H., Mitra S., *Multisensor Image Fusion Using the Wavelet Transform*. Graphical Models and Image Processing, 57(3) (1995) 235.

[7] Chibani Y., Houacine A., *Redundant Versus Orthogonal Wavelet Decomposition for Multisensor Image Fusion*. Pattern Recognition, 36 (2003) 879.

[8] Lewis L.J., O'Callaghan R., Nikolov S.G., Bull D.R., Canagarajah N., *Pixel and Region-Based Image Fusion with Complex Wavelets*. Information Fusion, 8 (2007) 119.

[9] Ajjimarangsee P., Huntsberger T.L., *Neural Network Model for Fusion of Visible and Infrared Sensor Outputs*. Sensor Fusion, Spatial Reasoning and Scene Interpretation, The International Society for Optical Engineering, 1003, SPIE, Bellingham, USA, (1988) 152.

[10] Goshtasby A.A., *Guest Editorial: Image Fusion: Advances in the State of the Art*. Information Fusion, 8 (2007) 114.

[11] Piella G., *A General Framework for Multiresolution Image Fusion: from Pixels to Regions*. Information Fusion, 4 (2003) 259.

[12] Wiliams M.L., Wilson R.C., Hancock E.R., *Deterministic Search for Relational Graph Matching*, Pattern Recognition, 32 (1999) 1255.

[13] Denkowski M., Chlebiej M., Mikołajczak P., *Depth of Field Reconstruction Method Using Partially Focused Image Sets*. Polish Journal of Environmental Studies, 16(4A) (2007) 62.

[14] Studholme C., et al., *An Overlap Invariant Entropy Measure of 3D Medical Image Alignment*. Pattern Recognition, 32(1) (1999) 71.

[15] Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T., *Numerical Recipes in C*, Cambridge University Press, second edition, (1992).

[16] Gonzalez R.C, Woods R.E., *Digital Image Processing*, Pearson Education, Inc., (2008).

[17] Constant A., *Close-up Photography*. Butterworth-Heinemann, (2000).

[18] Denkowski M., Chlebiej M., Mikołajczak P., *Development of the Cross-Platform Framework for the Medical Image Processing*. Annales UMCS, Sectio AI Informatica, III (2005) 159.

[19] Xydeas C., Petrovič V., *Objective Image Fusion Performance Measure*, Electronics Letters, 36(4) (2000) 308.