A hybrid multiple exposure image fusion approach for HDR image synthesis

Ioannis Merianos and Nikolaos Mitianoudis
Electrical and Computer Engineering Dep.
Democritus University of Thrace
67100 Xanthi, Greece
Email: nmitiano@ee.duth.gr

***Abstract*—The latest advancement in imaging applications has increased the need for more High Definition Range (HDR) imaging, which is not easily attainable by common imaging sensors. However, the use of multiple exposure images, that cover multiple exposure settings for the captured scene, and their combination in a single image via image fusion has been proposed in the literature and seems a viable solution. In this paper, the authors combine two image fusion methods to perform multiple exposure fusion. They use Mitianoudis and Stathaki [1] method to fuse the luminance channel and the Mertens et al [2] method to fuse the color channels. The derived fusion output outperforms both individual methods and other state-of-the-art methods.**

I. Introduction

The latest achievement in the field of imaging sensors have led to the development of high-specification image cameras. Thus, one might assume that capturing photos of natural scenes, as perceived exactly by the human visual system, will be possible in the near future. However, the maximum-to-minimum light intensity that can be recorded by a modern camera imaging sensor is around $2^8 - 2^{14}$. This ratio is called *dynamic range* and its logarithm with base 2 is measured in *stops*. A conventional DSLR camera can capture correctly scenes with 8 - 12 stops dynamic range, while professional movie cameras, such as the Red Epic, can capture scenes with 14 stops. The human visual system responds correctly to all the scenes of the natural world, where the dynamic range exceeds 24 stops. Consequently, in high dynamic range scenes, a picture taken by a imaging camera will often result into several areas being under or overexposed, i.e. darker or whiter colors than in real life.

The solution to this problem is to use *multi-bracketing photography*. That is to say, the scene is captured using the same imaging sensor taking multiple pictures at different exposures. In this manner, the different areas can well be captured in one (or more) of the different exposure photos. Commonly, a three-exposure setup (one under-exposed, one normally-exposed and one over-exposed) is sufficient for capturing efficiently most high dynamic range images. The task is then to combine the useful information from all input images to create a single composite image that resembles as accurately as possible the image attained by the human visual system. This task is often tackled by *image fusion* algorithms [1], however, we encounter the term *exposure fusion* in the literature [2], since we deal with the problem of fusing multiple exposures of the same scene.

There has been a lot of work lately on the field of *Multiple Exposure Fusion* (MEF). In [2], Mertens et al proposed a MEF scheme, based on Laplacian pyramid analysis of the input images, which are then fused using weights created from contrast, saturation and well-exposedness of the input images. In [3], Vonikakis et al proposed a MEF scheme, where well-exposedness is estimated via an illumination estimation filtering method. These initial estimations are then combined to create fusion weights via fuzzy membership functions, which will favor and promote well-exposed pixels to the fused image. In [4], Tico et al proposed a mechanism to deal with the motion-blur that may exist in longer exposures, using photometric calibration. In [5], Jinho and Okuda proposed novel weighting functions, which retain their independence on different exposure areas with little overlap. They also address possible motion-blur and occlusions by using maximum a posteriori estimation.

There also exist traditional and more general image fusion methods. In [6], Mitianoudis and Stathaki proposed a self-trained Image Fusion framework based on Independent Component Analysis, where the analysis transformation is estimated from a selection of images of similar content [6]. Several fusion rules were proposed under this framework in [6]. The analysis framework is projecting the images into localized patches of relatively small size. The local mean value of the patches is subtracted and stored in order to reconstruct the local means of the fused image. In [6], an average of the stored means was used to reconstruct the fused image. In the case of multi-modal inputs, a gradient algorithm that optimises the Piella and Heijmans Fusion Quality index [7] was derived in [1] to infer an optimal means choice for the fused image.

In this paper, the authors propose a multiplr-exposure system, where the luminance channels are fused via the traditional ICA-based image fusion of [1] and the color channels are fused using the Mertens et al method [2].

II. The proposed Multiple Exposure Image Fusion system

The proposed system is described by Fig. 1. Next, we are going to discuss its individual system blocks.

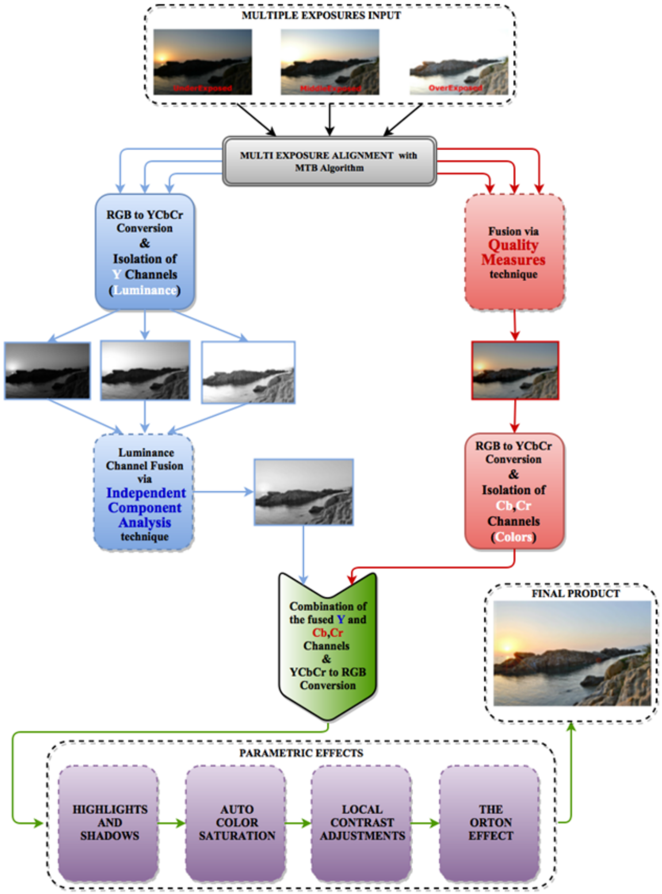

Fig. 1. Flow diagram of the proposed HDR synthesis approach.

A. Image Alignment - Color System Conversion

In our system, we used the common three exposure images setup (under-exposed, normally-exposed, over-exposed), however, the system can work with an arbitrary number of inputs. The first step was to check for possible registration errors between the three input images. To address this problem, we used the Median Threshold Bitmap (MTB) method of Ward [8], which deals with translational registration errors and scales linearly with the number of pixels. Since, in a real-life situation, the registration error between successive shots of a camera will be most probably translational, the MTB method offers a fast and viable solution that can be used in real-time applications.

The next step was to separate the luminance from the three RGB color channels. To achieve this, we converted the input RGB images to the YCbCr system, which offers this type of color separation. Then, the system used the Mitianoudis and Stathaki algorithm [1] to fuse the three Y channels from the input images and the Mertens et al algorithm [2] to fuse the three Cb and the three Cr channels. The inspiration comes from the human visual system. As it is also noted in the JPEG compression scheme, humans tend to pay greater attention to details that exist in the luminance/brightness channel and are less attentive to errors in the color channels. The Mitianoudis and Stathaki fusion algorithm yields very good performance at out-of-focus fusion cases, thus it was selected to fuse the luminance channels. However, since its performance may vary with images of different input intensity range, we chose the Mertens et al algorithm [2] to fuse the color channels, as a more linear blending color solution. Details of two image fusion algorithms follow in the next section.

B. Luminance channels fusion

For the fusion of luminance channels, we used the fusion algorithm proposed by Mitianoudis and Stathaki [1], which performs fusion in the Independent Component Analysis (ICA) domain. This technique is a transform-based fusion method, where the transform is learned by similar content images, using a statistical signal processing technique, called ICA. The analysis is performed at local image patches of size $N \times N$. The training algorithm extracts a random population of these patches (in the order of 10000 [9]) are randomly selected from similar-content training images. These patches are transformed to vectors $\mathbf{x}_w(t)$ using lexicographic ordering and the mean value of each vector is subtracted from the vector. The idea is to find a set of projection bases b_i (arranged in matrix B) that can lead to a sparse representation of the input image $\mathbf{u}(t)$.

$$\mathbf{x}_w(t) = B\mathbf{u}(t) \tag{1}$$

$$\mathbf{u}(t) = B^{-1}\mathbf{x}_w(t) = A\mathbf{x}_w(t) \tag{2}$$

The training procedure needs to be performed only once, as the estimated transform can be used for fusing images with similar content to the training images, as explained in more detail in [6]. First, we perform Principal Component Analysis (PCA) on the selected patches, in order to select the $K < N^2$ most important bases. Then, the ICA update rule in [6] for a chosen $L \times L$ neighborhood is iterated until convergence. In each iteration, the bases are orthogonalised using a symmetric decorrelation scheme. In the case of multimodal inputs, sample patches from all inputs are selected to train the ICA bases.

1) Fusion in the ICA domain: After estimating an ICA transform, Image fusion using ICA bases is performed. Every possible $N \times N$ patch is isolated from each image $x_k(i, j)$ and is consequently re-arranged to form a vector $\mathbf{x}_k(t)$. These vectors $\mathbf{x}_k(t)$ are normalized to zero mean and the subtracted local mean $MN_k(t)$ is stored for the reconstruction process. Each of the input vectors $\mathbf{x}_k(t)$ is transformed to the ICA or Topographic ICA domain representation $\mathbf{u}_k(t)$, using equation (2). Optional denoising in the ICA representation is also possible, by applying sparse code shrinkage on the coefficients in the ICA domain [9], assuming Laplacian (generally sparse) priors for the ICA representation. The corresponding coefficients $\mathbf{u}_k(t)$ from each image are then combined to construct a composite image representation $\mathbf{u}_f(t)$ in the ICA domain. The next step is to move back to the spatial domain, using the synthesis kernel B. The optimal means $MN_f(t)$ are estimated using the gradient rule in [1]. In the case of images of similar contrast, one can use the average means as an optimal choice, as this is usually the answer of the gradient rule in [1]. The optimal means are then added to the corresponding image patch. The image $f(i,j)$ is synthesised by spatially averaging the image patches $\mathbf{u}_f(t)$ in the same order they were selected during the analysis step.

There are a number of fusion rules that can be employed in ICA-based fusion [6]. Fusion by the *absolute maximum* rule simply selects the greatest in absolute value of the corresponding coefficients in each image ("max-abs" rule). This process seems to convey all the information about the edges to the fused image, however, the intensity information in constant background areas seems to be distorted. In contrast, fusion by the *averaging* rule averages the corresponding coefficients ("mean" rule). This process seems to preserve the correct contrast information, however, the edge details seem to get oversmoothed, since averaging is generally a "low-pass" filtering process. Finally, a *Weighted Combination* (WC) pixel-based rule uses weights $w_k(t)$ to combine the different inputs in the ICA representation. The weights should emphasize sources with more intense activity, as represented by the L1-norm. In our study, we used the "max-abs" rule, which seemed to produce the best results in our problem.

C. Color channels fusion

For the selection of the HDR image colors, we present the second fusion algorithm of multiple exposures into a high resolution image by Mertens et al [2]. This technique merges the input exposures in a manner guided by simple measures of quality such as the saturation, the contrast and the level of good exposure. First, we calculate the values of quality measures in each pixel of the multiple exposures and then we attach a weight that depends on all quality measures. Finally, we combine these weights properly to produce an image with enhanced features that includes all the information of multiple entry exposures. The three quality measures are calculated as follows:

- Contrast: applying a Laplacian filter to the gray version of each exposure, we take the absolute value of the filter response. This results in a simple index C for the contrast. This method tends to define large weights on important elements, such as edges and texture.
- Color saturation: As a picture is subjected to a greater exposure, colors lose their sharpness and get saturated. Saturated colors must be preserved as the image becomes more vivid. For this reason, a measurement of color saturation S is used, which calculates the standard deviation of the R, G, and B channels for every pixel.
- Well-exposedness: The channel intensities reveal how well each pixel is exposed. The purpose of this metric is to keep the intensities that are not close to zero (underexposed intensities) or one (overexposed). Each pixel intensity x is weighted by a weight that depends on how close is to 0.5, using the Gaussian curve: $\exp((-(x-0.5)^2)/(2\sigma^2))$ where σ is equal to 0.2. To take account of the three color channels, the Gaussian curve is applied to each channel and is multiplied resulting to the measure E.

The information from the different metrics is combined into a single weight map for each pixel using multiplication. This is performed, because we need to take into account all these factors simultaneously. We can control the effect of each metric using a power function:

$$W_{ij,k} = (C_{ij,k})^{w_C} \times (S_{ij,k})^{w_S} \times (E_{ij,k})^{w_E} \tag{3}$$

where C, S and are the contrast, color saturation and well-exposedness, respectively, and w_C, w_S, w_E are weighting exponents of the metrics. The indices i, j, k refer to pixel (i, j) of the k exposure. The fusion of K images can be done with a weighted average for each pixel, using the weights calculated from the quality metric. For a sensible result, the values of K weight maps are normalized so that they sum to 1 for each pixel (i, j).

In several cases, this simple weighted mixing of the input image may produce negative results. When the weights vary quickly, invalid combinations of the input images may appear. To solve this problem, the input images are decomposed in a Laplacian pyramid, which consists of band-pass filtered versions at different scales and then mixing is carried out for each level separately. Finally, the pyramid is reconstructed to obtain the fused image. The blending of multiple scales is quite effective in preventing strange assembly lines, because it combines the characteristics of the image instead of raw intensities.

D. Reconstruction and Post-Processing

The outcome of the two fusion algorithms are three images: one representing the fused luminance Y channel and two representing the fused color Cb, Cr channels. The three channels are combined and transformed to the RGB system, which gives the final fused image.

Modern HDR applications offer a wide range of post-processing options that can enhance the color balance in the fused image, even create surrealistic scenes containing very vivid colors. The proposed system features a number of optional post-processing steps that can be applied to the fused image: a) edge and detail processing (Highlights and shadows), b) color enhancement (Auto Color Saturation) and c) sharpness improvement (Local Contrast Adjustment). Edge and detail processing is achieved using bi-lateral filtering [10]. Auto Color Saturation is performed by boosting saturation using a patented method, described in [11]. For sharpness improvement, the common unsharp mask is used. Finally, the Orton effect can be implemented with a series of elementary image processing filter, following the method described in [12].

III. Experiments

In this section, we evaluate the performance of the proposed hybrid multiple exposure image fusion system. We compare the proposed model with four advanced fusion algorithms: 1) Multiple exposure fusion based on Illumination Estimation [3], 2) Image fusion via quality metrics (Mertens) [2], 3) Fusion based on bilateral filtering (Raman) [13] and 4)

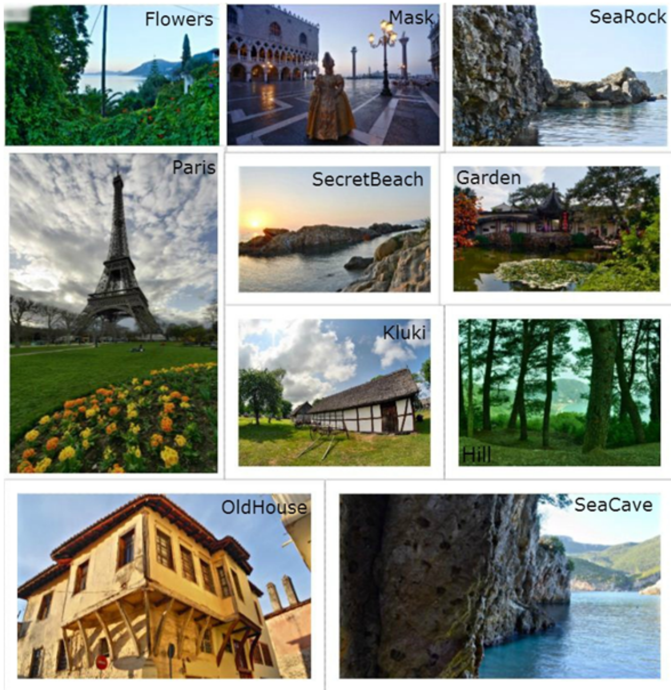

Fig. 2. Ten images from various databases and personal photography that were used in our experiments.

Dynamic Photo HDR, which is a commercial application. The proposed algorithm is called Hybrid HDR, since it combines two methods to produce the fused image. The method's development and performance comparison were performed in the MATLAB 2015a programming environment. For the proposed HybridHDR system, we used a neighborhood of 7×7 local patches for the ICA fusion framework, without dimensionality reduction and no post-processing. The developed system can be downloaded online[1]. MATLAB implementations for all the other methods were traced online, either at the authors' website or at the Mathworks resource center website.

To evaluate and compare the different algorithms with the proposed Hybrid HDR method, we applied the Image Quality Assessment (IQA) model of Ma et al [14], which is based on the principle of the structural similarity approach and a novel measure of patch structural consistency. This model gives a maximum score of 1 to the best fusion result. We have chosen 10 sequences of exposures, covering diverse content including exterior views, natural landscapes and artificial architectures. Some of them were selected from common multiple-exposure datasets, whereas some others were photographed by the authors and are freely available for download[2]. All input images are shown in Figure 2 and they contain 3 exposures (under-exposure, over-exposure and normal exposure) between the cases. The results are presented in Table I.

As it can be seen from the results, the fusion method based on the illumination estimation has on average the best score.

[1]http://utopia.duth.gr/nmitiano/download.htm
[2]http://utopia.duth.gr/nmitiano/download.htm

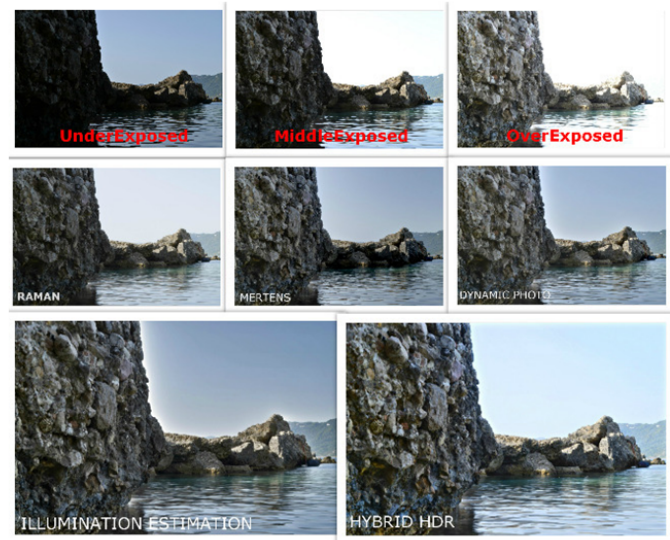

Fig. 3. Comparison of the four MEF algorithms using the "SeaRock" images.

The proposed Hybrid HDR nearly reaches the best quality, followed by the Dynamic Photo HDR application and Mertens algorithm. The method, which is based on bilateral filtering (Raman) gives the worst results, which are also reflected perceptually, if we compare visually all the results. The latter method fails to adequately convey small details and color information of the input images in the fused image. Subjective evaluation can be performed by looking at some indicative fusion results depicted in Fig. 3, 4 and 5. It is promising to see that the proposed HybridHDR outperforms the Mertens method, which implies that the ICA fusion for the luminance channel outperforms and improves the traditional Mertens method.

Therefore, we conclude that the proposed method gives, in accordance with the quality metric evaluation of the previous paragraph, better fusion results compared to many techniques that have been developed to date. Finally, it should be noted that although the algorithm based on the Illumination Estimation has a better average score than the proposed model (Hybrid HDR), it tends to show artificial objects around the edges, which results to intense brightness fluctuations (Fig. 3, 4 and 5). This is due to the fact that the algorithm doesnt have a perceptually correct representation of the physical scene at these points, but we reckon that it shows a higher rating, because the quality assessment model applied does not take account of the brightness components.

IV. Conclusions

In this paper, the authors propose a Hybrid multiple exposure algorithm, by combining a traditional image fusion algorithm and the Mertens et al algorithm. The first is used to fuse luminance input channels, whereas the latter is used to fuse color channels Cb and Cr. The proposed algorithm outperforms Mertens algorithm in most cases, and scores favourably with the Illumination Estimation method. The authors are looking to

TABLE I
QUALITY ASSESSMENT OF DIFFERENT MULTIPLE EXPOSURE FUSION ALGORITHMS.

Method	Flowers	Mask	SeaRock	Paris	SecretBeach	Garden	Kluki	Hill	OldHouse	SeaCave	Average
Illumination Estimation [3]	0.9675	0.9531	0.9562	0.9643	0.9624	0.9536	0.9605	0.9665	0.9669	0.9706	0.9621
HybridHDR	0.9720	0.9748	.09444	0.9662	0.9381	0.9588	0.9552	0.9671	0.9770	0.9641	0.9617
DynamicPhoto	0.9705	0.9777	0.8822	0.9698	0.9360	0.9724	0.9652	0.9686	0.9753	0.9491	0.9566
Mertens [2]	0.9642	0.9323	0.9316	0.9522	0.9508	0.9369	0.9601	0.9386	0.9743	0.9336	0.9474
Raman [13]	0.9064	0.9163	0.8964	0.853	0.9270	0.9010	0.8988	0.9166	0.9591	0.8813	0.9088

Fig. 4. Comparison between the HybridHDR and the Illumination Estimation method for three image sets. One can see more halo artifacts in the Illumination Estimation method.

Fig. 5. Comparison of the four MEF algorithms using the "Venice" images

expand the method, by following a strategy to remove possible halo artifacts from the fused image.

REFERENCES

[1] N. Mitianoudis and T. Stathaki, "Optimal Contrast Correction for ICA-based fusion of Multimodal Images," *IEEE Sensors Journal*, vol. 8, no. 12, pp. 2016 – 2026, 2008.

[2] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure Fusion," in *Proc. Pacific Graphics*, Maui, Hawaii, USA, 2007.

[3] V. Vonikakis ans O. Bouzos and I. Andreadis, "Multi exposure image fusion based on illumination estimation," in *Proc. SIPA*, Heraklion, Greece, 2011, pp. 135–142.

[4] M. Tico, N. Gelfand, and K. Pulli, "Motion-blur-free exposure fusion," in *Proc. ICIP*, Hong Kong, 2010, pp. 3321–3324.

[5] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *IEEE Trans. on Image Processing*, vol. 21, no. 1, pp. 358–365, 2012.

[6] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Elsevier Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.

[7] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, pp. 259–280, 2003.

[8] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," *Journal of Graphics Tools*, vol. 8, pp. 17–30, 2003.

[9] A. Hyvärinen, P. O. Hoyer, and E. Oja, "Image denoising by sparse code shrinkage," in *Intelligent Signal Processing*, S. Haykin and B. Kosko, Eds. IEEE Press, 2001.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International IEEE Conference on*, 1998, p. 839 846.

[11] A. Sarkar, J.E. Caviedes, and M. Subedar, "Joint Enhancement of Lightness, Color and Contrast of Images and Video, Patent US 8,477,247 B2," 2013.

[12] M. Orton, "The orton effect, http://www.michaelortonphotography.com/ortoneffect.html," 2012.

[13] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009, pp. 1–4.

[14] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.