A VARIATIONAL APPROACH FOR THE FUSION OF BRACKETING IMAGES

By

Marcelo Bertalmío

and

Stacey Levine

IMA Preprint Series # 2371

(June 2011)

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

A Variational Approach for the Fusion of Bracketing Images

Marcelo Bertalmío and Stacey Levine

***Abstract*—In the very common situation of taking pictures in a night scene with artificial lighting, the light coming from the scene is not enough for most cameras. Using the exposure bracketing feature available in most photo cameras, we obtain a series of pictures taken in rapid succession with different exposure times with the implicit idea that the user picks from this set the better looking image. But often none of these images is *good enough*: in general, good color information is retained from longer exposure settings while sharp details are obtained from the shorter ones. In this work we propose a variational method for automatically recombining a bracketed pair of images into a single picture that reflects the optimal properties of each one. For this we introduce an energy functional consisting of two terms, one measuring the difference in edge information with the underexposed image, the other term measuring the local color difference with a warped version of the overexposed image. The method is able to handle camera and subject motion as well as noise, and the results compare favorably with the state of the art.**

I. Introduction

Exposure bracketing is a setting on many photo cameras in which a series of pictures is taken in rapid succession using varying shutter speeds. The idea is that the user picks from this set of images the one that has a better compromise between color information and sharp details. This can be a quite useful feature when taking a photograph under difficult lighting conditions. But often none of these images is *good enough*: longer exposure times preserve good intensity and color information but may result in blurred images; shorter exposure times don't afford good color information but sharp details are preserved. See figure 1 for an example.

In this work we propose a technique for automatically recombining a bracketed pair of images of different exposure times into a single picture that reflects the optimal properties of each one. The proposed technique is a concise variational approach. This method is intended for still images, however in the final section we suggest how to extend it to deal with motion pictures, which suffer from the same limitations.

There are several related works on image fusion of pairs of photographs. For example, [1], [2], [3], [4], [5] deal with fusing a pair of pictures taken both with and without a flash, or deblurring a long-exposure image with a blurry and non-blurry image pair. However, the current literature on fusing flash/no-flash image sets assume there is no motion. In general this is not the case, even if using a tripod, when the picture features human subjects and hence a small but noticeable amount of motion is present. Current work on picture deblurring assumes the blur only comes from camera motion and the scene is static, which again prevents this method from being applied to most pictures featuring people. Jia et al. [6] proposed a fusion method that combines histogram matching with spatial color matching yielding very good results, but they require the underexposed image to be noise-free, a requisite which does not usually hold in practice, and the spatial color matching of small regions is problematic. Denoising image bursts was addressed by Buades et al. in [7], which combines a stack of short-exposure images of the same scene to create a noise-free image. The results are quite nice, however this differs from the proposed work in that each image in the burst is noisy but sharp, so it's not necessary to account for blurring. There is extensive literature on color transfer methods based on histogram modification, see for instance Rabin et al. [8] and references therein. These types of approaches are global, i.e. they work on the color space and ignore the spatial distributions of color values, which may produce visible artifacts in the processed images. The method by Huang and Chen in [9] is capable of locally transferring color (and brightness) from a blurry, long-exposure image to a sharp, noisy one, allowing for camera and subject motion, but the results may suffer from color leakage artifact's, i.e. colors spill outside their corresponding region. In [10], the authors proposed a technique to handle both static and non-static scenes, with camera and/or subject motion. The technique uses stereo matching to extract the optimal color information from the overexposed image and Non-local means denoising [11] to suppress noise while retaining sharp details in the luminance of the underexposed image. Although this approach is effective, it requires a number of different steps and parameters that require some care in tuning. The technique proposed in the present article is a more concise, mathematically sound approach which also improves on the results in [10].

The paper is laid out as follows. In sections II and III, we discuss the proposed automated algorithm. Numerical results are presented in section IV that demonstrate the effectiveness of the algorithm. In section V, we draw some conclusions and discuss future work.

II. The proposed model

Our problem statement is as follows: given and underexposed image I_u and an overexposed image I_o, we want to create an image I that combines the sharpness of detail of I_u and the colors of I_o. Variational approaches provide a concise

M. Bertalmío is with the Departamento de Tecnologías de la Información y las Comunicaciones, Universitat Pompeu Fabra, Barcelona, Spain, marcelo.bertalmio@upf.edu.

S. Levine is with the Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, Pennsylvania, USA, sel@mathcs.duq.edu.

Fig. 1. Pictures taken with different exposure settings in low light conditions. Left: short exposure implies sharp objects but noise, low brightness and poor colors. Right: long exposure implies motion blur but no noise and good colors.

methodology to tackle several tasks simultaneously. Therefore, we will propose an image functional $E(I)$ with two terms, one measuring the difference in edge information between I and I_u, the other measuring the difference in colors between I and I_o. The minimization of $E(I)$ will produce the desired fusion result.

We deal with each color channel independently so in our notation I, I_u, I_o denote single-valued functions from the image domain Ω into the real interval $[0, 1]$.

A. First energy term: edge information

The fusion result I should retain the sharp details of I_u. This can be rephrased in the following way: the level lines of I should match those of I_u, so that the edges in both images coincide. Approximating edge direction by the orthogonal gradient direction, the intended matching of level lines may be achieved by locally matching the gradient *direction*:

$$E_{g,I_u}(I) := \int_\Omega \left(|\nabla I(x)| - \nabla I(x) \cdot \frac{\nabla I_u(x)}{|\nabla I_u(x)|} \right) dx. \quad (1)$$

Ballester et al. [12] first proposed this type of energy in the context of image inpainting: given a normalized gradient field θ, minimizing an energy such as (1) (where the normalized gradient of I_u plays the role of θ) propagates the values of I along the field. Notice that we use the normalized gradient of I_u so the strength of the gradient doesn't come into play: we want I to have the same level lines as I_u (which is given by the gradient direction), not the same brightness and contrast (which would be given by the gradient magnitude) because I_u is poor in these respects.

B. Second energy term: color matching

The second challenge is to match the colors of our fused image I to that of the overexposed image I_o. In [13], Sapiro and Caselles proposed a variational formulation for histogram modification. For a given image I they showed that by minimizing the following functional the histogram of I is *equalized*:

$$E_{histogram}(I) := \frac{1}{2} \int \left(I(x) - \frac{1}{2} \right)^2 dx$$
$$-\frac{1}{4} \int \int |I(x) - I(y)| \, dxdy.$$

They also suggest extending this to more general gray-value distributions. We have adapted this idea to our setting. If we want I to have the same histogram as I_o then the functional that we should minimize is:

$$E_{h,I_o}(I) := \frac{2}{WH} \int \int |I(x) - I_o(y)| \, dxdy$$
$$+\frac{1}{WH} \int \int |I(x) - I(y)| \, dxdy$$

where W and H are the dimensions (width and height respectively) of the image domain. Performing gradient descent we obtain the following update formula:

$$I_t(x,t) = 4(p(x,t) - p_o(x,t)), \quad (2)$$

where $p(x,t)$ is the percentage of pixels in the image I with values below $I(x)$ and $p_o(x,t)$ is the percentage of pixels in the image Io with values below $I(x)$, both at evolution time t. So we see that, upon convergence, $I_t = 0$ and the histograms of I and I_o coincide.

But this is a *global* histogram modification procedure, i.e. the color values are modified regardless of their spatial distributions, and as such it adolesces of the visual artifacts common in global histogram matching methods as mentioned in the introduction. Introducing locality into the histogram matching term should alleviate some of these issues:

$$E_{h,\omega,I_o}(I) := \frac{2}{WH} \int \int \omega(x,y) \, |I(x) - I_o(y)| \, dxdy$$
$$+\frac{1}{WH} \int \int \omega(x,y) \, |I(x) - I(y)| \, dxdy \quad (3)$$

where $\omega(x,y)$ is a decreasing function of the distance $||x-y||$ (in our case ω is a Gaussian of standard deviation σ pixels). There is still a problem using (3) in that if σ is large, then just as before we have global histogram matching. But if σ is small, ghosting artifacts appear wherever there is motion since the histograms of the corresponding regions may be too different.

To solve this we introduce a kind of "motion compensation" into the functional:

$$E_{h,\omega,\tilde{I}_o}(I) := \frac{2}{WH}\int\int \omega(x,y)\left|I(x)-\tilde{I}_o(y)\right| dxdy + \frac{1}{WH}\int\int \omega(x,y)\left|I(x)-I(y)\right| dxdy \quad (4)$$

where $\tilde{I}_o$ is a warped version of I_o that keeps the colors of I_o but where its shapes have been deformed so that its geometry is close to that of I_u. How you construct this warped I_o is not essential to the algorithm. In our case, we construct it with stereo matching as in [10], but for example one could use the automatic warping/morphing technique of Shinagawa and Kunii [14], or perform warping based on an estimation of the optical flow (see Papenberg et al. [15] and references therein).

C. Total energy

We propose minimizing the following energy, which combines a gradient direction term and a local histogram matching term:

$$E(I) = E_{g,I_u}(I) + \lambda E_{h,\omega,\tilde{I}_o}(I), \quad (5)$$

where E_{g,I_u} and $E_{h,\omega,\tilde{I}_o}$ are defined above in (1) and (4) respectively, $\lambda > 0$ is a weighting parameter, and the energy E is minimized over all possible fused images I.

Remark: Since $E_{g,I_u}(I) \geq 0$, it is straightforward to show the existence of a minimizer to (5) using a proof analogous to *Proposition 3* in Bertalmío et al. [16].

III. Implementation

Computing the Euler-Lagrange equation of (5) and using a gradient-descent approach we obtain the following update equation for I:

$$I_t(x,t) = \kappa_I(x,t) - \kappa_{I_u}(x) + 2\lambda\int \omega(x,y)s(I(x)-I(y))dy - 2\lambda\int \omega(x,y)s(I(x)-I_o(y))dy, \quad (6)$$

where $\kappa_I(x,t)$ stands for the Euclidean curvature of the level line of I at location x and evolution time t, $\kappa_{I_u}(x)$ is the curvature of the level line of I_u at location x (no evolution time here since I_u is given at the beginning, as a constant image), and s is the sign function. The initial condition for I is I_u: $I(x,0) = I_u(x),\ \forall x \in \Omega$.

For the computation of the curvature terms we use the classical approach of Rudin et al. [17], i.e. an explicit, forward-time scheme with alternating forward-backward spatial differences. For the integral terms we notice that they involve a convolution with the Gaussian kernel ω, and therefore computing these terms explicitly would have a computational complexity of order $O(N^2)$ where N is the number of pixels. Following the approach introduced in [16], we approximate the sign function with a polynomial and compute the convolutions using the Fast Fourier Transform, thus reducing the complexity to $O(NlogN)$ (see [16] for details).

IV. Experiments and comparisons

We have run our algorithm on several pairs of exposure bracketing images taken with a consumer camera. In all the results shown we have used the same set of values for the parameters: $\lambda = 20$, $\sigma = 35$ pixels. Using non-optimized code on a dual processor 1.87GHz, 3Gb PC, the computational cost for each 770×430 image is of about two minutes (we're not taking into account the warping of I_o to obtain $\tilde{I}_o$ as it is not part of our algorithm, but a warping time of a few seconds for this size of images is reasonable). The reason why our algorithm takes this much time lies in the fact that we are using an explicit scheme for the curvature, which imposes a very small time-step in our iterations. We are currently investigating ways to speed this up: one possibility we want to investigate is splitting the evolution equation and iterating between curvature and histogram terms.

Some example results can be found in figure 2 for a variety of real-life scenarios. In general we notice that the colors are correctly transferred from I_o, without any sort of color leakage artifacts, and that noise is reduced. Note that the original images, taken without a tripod, have some challenging characteristics that our algorithm is able to handle. For example, they feature people whose mouths are open in one image and closed in the other. Moreover, although the subjects move very little, this small motion causes noticeable motion blur due to the exposure time required to properly capture the colors. Again, these are very common situations, in which methods that require static scenes and still subjects wouldn't be practical. We must also point out that in the case of the bottom row images the use of flash was not allowed (another rather common scenario), which prevents the use of fusion methods involving flash/no-flash image pairs.

One observation to be made is the following: global histogram modification does not alter the level lines, and while local histogram modification does modify the level line structure of the image it does so in a way that might be unnoticeable, depending on the width of the kernel and the amplitude of the discontinuities present in the image (see [16] for a rigorous discussion of this). Given that the first term of our functional has as only purpose the preservation of the level lines, and that we start with the initial condition $I = I_u$, we could consider dropping this first term and perform only local histogram matching. But figure 3 shows that the gradient direction term is essential in order to suppress the noise and also to prevent the appearance of spurious colors, this latter phenomenon associated to color modification due to histogram matching regardless of spatial considerations. This is of particular relevance given that underexposed images normally suffer from abundant noise, and simply enhancing the colors and increasing brightness would only worsen this problem by making noise more apparent.

It may seem surprising to find that the gradient direction term has a regularizing effect on the fusion result, because if we only considered this term then the minimum $E_{g,I_u}(I) = 0$ would be attained at $I = I_u$, i.e. the fusion result would be as noisy as the original underexposed image. The first energy term preserves the shape of *all* level lines, so it also

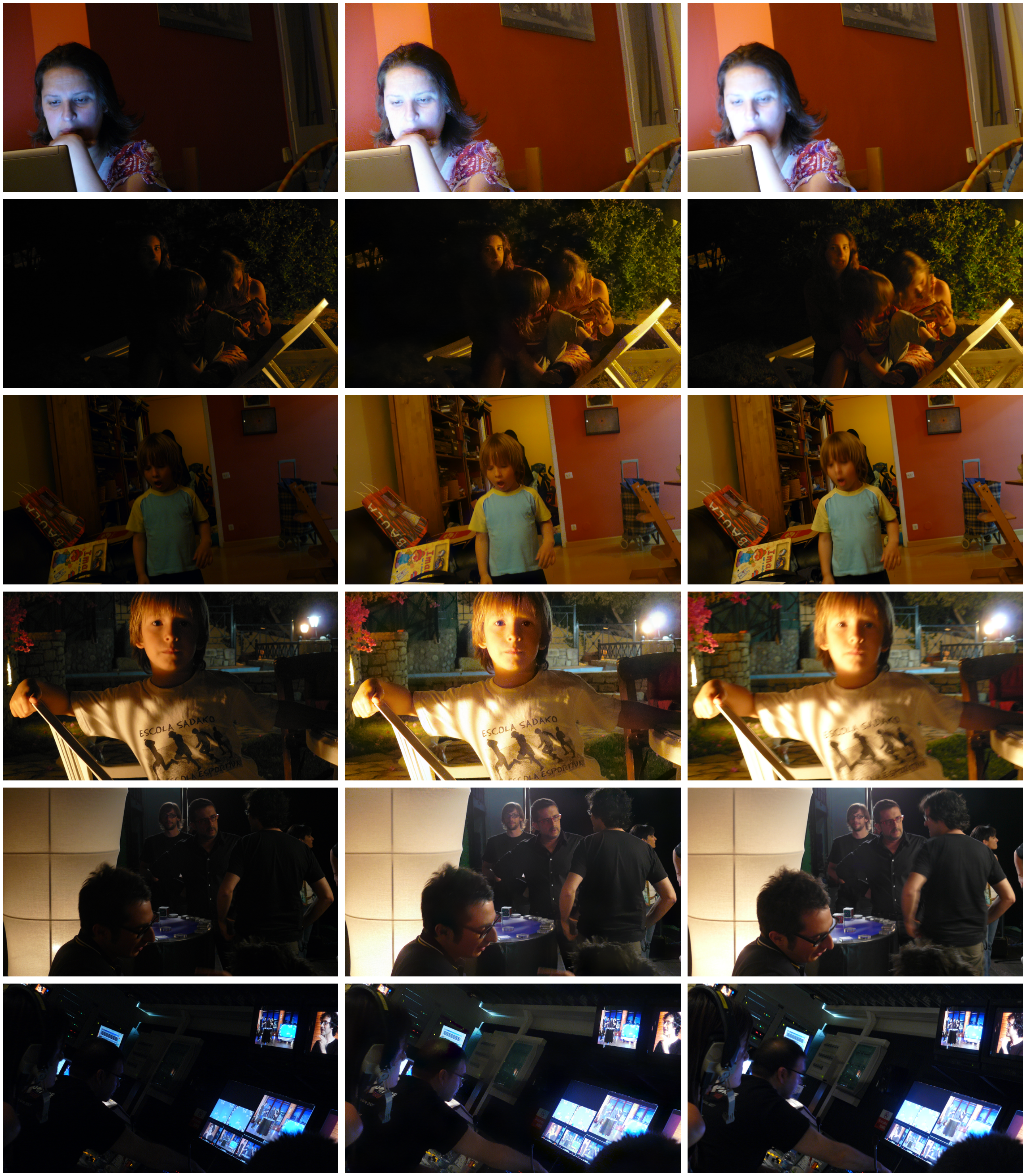

Fig. 2. Left: underexposed images. Right: overexposed images. Middle: our fusion result.

preserves noise, which corresponds to level lines of tiny length. The second energy term matches colors with the overexposed image I_o, also tending to respect *the shape* of the level lines: they are just moved upward in intensity level, because we start the minimization with $I(t = 0) = I_u$ and I_u is darker than I_o. So we can see that in the combined effect of both terms, noise is actually preserved (because of the first term) but its amplitude is kept constant while the amplitude of the clean signal increases (because of the second term), with the net result that noise, in terms of Signal to Noise Ratio, is reduced because the SNR goes up.

Too much motion and/or occlusion may cause problems with the warping which then turn into color artifacts after the processing, see figure 4. Since most of the guitar is occluded in one view, the fusion result presents wrong colors on the guitar and the singer's jacket. One possible solution would be

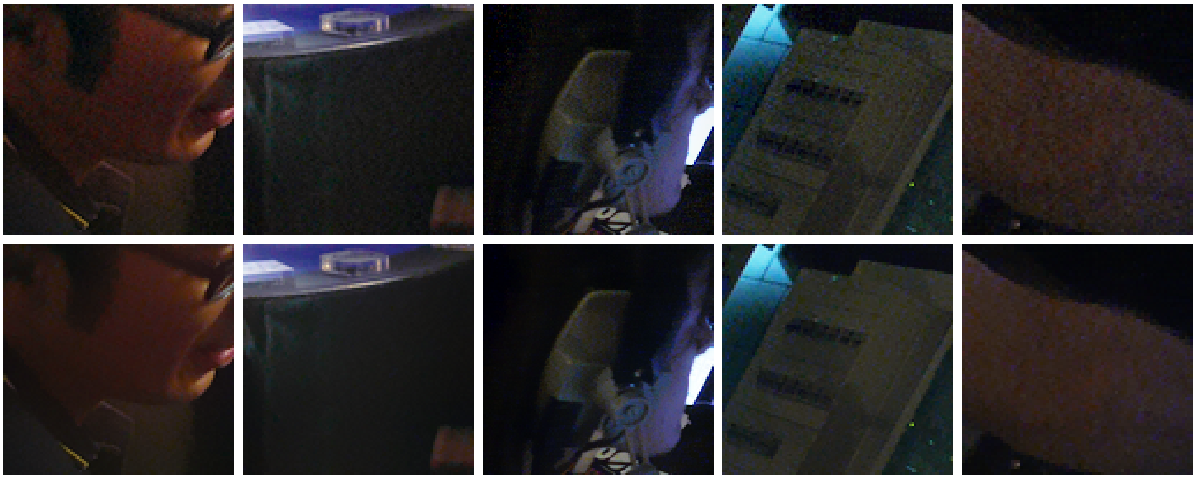

Fig. 3. Some image details showing the importance of the gradient direction term. Top: fusion results without this term. Bottom: fusion results with full model, notice reduction of noise and spurious colors.

to compute a sort of fidelity mask for the warping, so that our color transfer procedure is performed only over the regions where the fidelity is high (i.e. where the warping is good). These would leave color gaps in the regions where the fidelity is low, but given that we have the gradient orientation inside these regions we could apply the Poisson Editing technique of Pérez et al. [18] to propagate valid colors inwards into these gaps. We have left this as future work since a masked local histogram matching approach is not compatible with the speed-up implementation described in section III, where we combine polynomial expansion of the sign function with a FFT computation for the convolutions.

For comparisons, we must recall from the introduction that most works in the literature on image fusion of pairs of photographs have limitations that make them impractical for our particular problem: they may require no camera and/or subject motion, no noise in the underexposed image, or no blur. In fact, to the best of our knowledge the only works with no restrictions on camera/object motion and presence of blur are those by Jia et al. [6], Huang and Chen in [9] and the authors in [10]. The method in [6] requires the underexposed image to be noise-free, a requisite which does not usually hold in practice, and the spatial color matching of small regions is problematic. In [9] the results may suffer from color leakage artifacts. Therefore we have decided to compare our results with those obtained with the method presented in [10], and these comparisons can be seen in figure 5. In the top row images notice how the method in [10] produces severe artifacts on and around the head of the subject in the foreground. In the bottom row images we can see that [10] is not as capable as our method of dealing with the noise, which is quite noticeable on the flat surfaces, and also that some contrast has been lost on the face of the subject.

V. Conclusions and future work

In this work we propose a variational formulation for automatically combining a bracketed set of images (one underexposed, one overexposed) in such a way that yields an image retaining the optimal qualities in each one. The method can handle camera and subject motion and reduces the noise present in the underexposed image. The results don't show the usual color leakage artifacts and they compare favorably with the state of the art. The main shortcomings of the proposed technique are its computational cost and its reliance on a good quality warping of the overexposed image: if the warping is poor, its problems translate into visual artifacts in the final result. We are currently investigating how to overcome these limitations.

This work has an interesting potential extension to video processing. In particular, if the first frame F_0 of a video is obtained with a long exposure time and the remaining frames, F_i for $i = 1, ..., T$, with short exposure, we can apply the proposed method in the following way. First apply the method described in section II to the pair (F_0, F_1) obtaining $\tilde{F}_1$, then successively to each pair $(\tilde{F}_{n-1}, F_n)$ obtaining $\tilde{F}_n$. In theory, the color from F_0 should be transferred to all remaining frames, while they still retain a good level of detail. But occlusion can be a problem as soon as the frames are too far apart, and also as the geometric configuration of the scene changes over time so will the global illumination configuration. This will be the subject of further work.

Acknowledgments

We would like to thank Prof. Vicent Caselles for his suggestions and discussions, and the Institute for Mathematics and Applications (IMA) in which the authors were visitors in 2011. The first author acknowledges partial support by PNPGC project, reference MTM2009-08171. The second author is funded in part by NSF-DMS $\#0915219$.

References

[1] G. Petschnigg, M. Agrawala, H. H., R. Szeliski, M. Cohen, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 664–672, 2004.

[2] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 673–678, 2004.

Fig. 4. Left: underexposed image. Right: overexposed image. Middle: our fusion result.

Fig. 5. Comparisons with method in [10]. Left: fusion result from [10]. Right: our result.

[3] R. Fattal, M. Agrawala, and S. Rusinkiewicz, "Multiscale shape and detail enhancement from multi-light image collections," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 51, 2007.

[4] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," *IACM Transactions on Graphics*, vol. 26, no. 3, pp. 1–10, 2007.

[5] ——, "Progressive inter-scale and intra-scale non-blind image deconvolution," in *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*. New York, NY, USA: ACM, 2008, pp. 1–10.

[6] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, "Bayesian correction of image intensity with spatial consideration," in *Computer Vision - ECCV 2004*, ser. Lecture Notes in Computer Science, T. Pajdla and J. Matas, Eds. Springer Berlin / Heidelberg, 2004, vol. 3023, pp. 342–354.

[7] A. Buades, Y. Lou, J. M. Morel, and Z. Tang, "A note on multi-image denoising," *Proceedings of the International Workshop on Local and Non-Local Approximation (LNLA) in Image Processing*, August 2009.

[8] J. Rabin, J. Delon, and Y. Gousseau, "Regularization of transportation maps for color and contrast transfer," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, pp. 1933–1936.

[9] T. Huang and H. Chen, "Landmark-based sparse color representations for color transfer," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 199–204.

[10] M. Bertalmío and S. Levine, "Fusion of Bracketing Pictures," in *Visual Media Production, 2009. CVMP'09. Conference for*. IEEE, 2009, pp. 25–34.

[11] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[12] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1200–1211, 2001.

[13] G. Sapiro and V. Caselles, "Histogram modification via differential equations," *J. Differential Equations*, vol. 135, no. 2, pp. 238–268, 1997.

[14] Y. Shinagawa and T. Kunii, "Unconstrained automatic image matching using multiresolutional critical-point filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 994–1010, 1998.

[15] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 141–158, 2006.

[16] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1058–1072, 2007.

[17] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[18] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3. ACM, 2003, pp. 313–318.