This paper appears as Y. Zhang, L. Ding, and G. Sharma, A local-linear-fitting-based matting approach for accurate depth upsampling, in IEEE Western NY Image and Signal Processing Workshop, Nov 2016.

A LOCAL-LINEAR-FITTING-BASED MATTING APPROACH FOR ACCURATE DEPTH UPSAMPLING

Yanfu Zhang, Li Ding, and Gaurav Sharma

Dept. of Electrical and Computer Engineering, University of Rochester, NY

ABSTRACT

We propose an approach for upsampling depth information for RGB color plus depth (RGB-D) images captured with common acquisition systems, where RGB color information is available at all pixel locations whereas depth information is only available at a subset of the pixels. Depth upsampling is formulated as a minimization of an objective function composed of two additive terms: a data fidelity term that penalizes disagreement with the low-resolution observed data and a regularization term that penalizes weighted depth deviations from a local linear model in spatial coordinates, where the weights are determined to ensure consistency between the RGB color image and the estimated depth image. Analogous to techniques used for optimization formulations of image matting, the upsampled depth image is then obtained by solving a large sparse linear system of equations. Visual evaluation of results obtained with the proposed algorithm demonstrate that the method provides high resolution depth maps that are consistent with the color images. Quantitative comparisons demonstrate that the method offers an improvement in accuracy over current state of the art techniques for depth upsampling.

Index Terms— depth map upsampling, hole filling, Laplacian matrix, RGB-D image

1. INTRODUCTION

RGB-D images are widely used for multiple purposes, for example segmentation, tracking, image dehazing and 3D scene reconstruction. A key challenge in using RGB-D images is that the depth images are limited in resolution compared to RGB images. Time-of-Flight (ToF) and structured light based systems are the two prominent methods for capturing depth data. While ToF based systems provide highly accurate depth information, they are relatively tedious to use and even after sophisticated alignment with images [1], typically offer a lower resolution than typical high resolution color cameras. For structured light based RGB-D images a significant fraction of the pixels (up to 10%) are not assigned depth values due to the challenges of these systems. Thus for both ToF and structured light based RGB-D image capture systems, some forms of depth upsampling (including hole filling) are required to generate a complete RGB-D image.

Traditionally depth upsampling is accomplished by bilinear or bicubic interpolation. These methods have difficulty in preserving the sharp edges in depth maps. Several methods have been developed to overcome these problems, aiming at improving the accuracy of depth upsampling problem. One class of techniques relies on proposing a prior and optimizing an objective function that combines prior and data fidelity terms [2, 3, 4, 5, 6, 7, 8]. Diebel and Thrun [2] proposed an upsampling algorithm based on Markov random field (MRF), which is defined through depth measure potential, depth smooth prior and weighting factors. This MRF framework is further improved by other researchers, such as [9] and [10]. Yang et al. [3] made use of a bilateral filter in an iterative refinement framework. The refinement is constructed on a cost volume defined on the current depth map and the RGB image. This algorithm can also work on two view depth map refinement with a different cost volume definition. In [4], the guided filter was designed for edge preserving filter, which can be viewed as an extension of the bilateral filter. Kopf et al. [5] proposed joint a bilateral filter which is also similar in principle. Both filters can be used to upsample the depth map with a high resolution RGB image. Park et al. [6] gave an algorithm based on a non local mean filter. The low resolution depth map is pre-processed to detect outliers. These points are removed and to obtain the high resolution depth map an objective function consisting of a smooth term, non-local structure term and data term is optimized. This algorithm is also suitable for filling large holes in the depth data. Ferstl et al. [7] gave an algorithm based on total generalization variance (TGV). A TGV regularization weighted according to intensity image texture is used in the objective function and the optimization is solved as a primal-dual problem. Yang et al. [8] built a color-guided adaptive regression model for depth map upsampling. Different edge preserving terms including non-local mean and bilateral filters are tested and an analysis is given on the parameter selection and the system stability.

Another category of depth map upsampling utilizes segmentation techniques to extract depth information. Krishnamurthy and Ramakrishnan [11] and Uruma et al. [12] start from an upsampled depth map using standard interpolation methods and refine the result by image segmentation techniques. The segmentation process serves a similar function in

This paper appears as Y. Zhang, L. Ding, and G. Sharma, A local-linear-fitting-based matting approach for accurate depth upsampling, in IEEE Western NY Image and Signal Processing Workshop, Nov 2016.

preserving edges as the afore-mentioned filters.

A common theme of prior algorithms, also adopted in our work, is to "fix" edges of upsampled depth map for better consistency with the color image. Our work is inspired by Levin et al.'s optimization formulation of matting [13], in which the alpha value for the matting mask is modeled as a linear combination of neighboring color values. Analogous to the matting problem, we formulate depth upsampling as an optimization problem. Specifically, the upsampled image is estimated by minimizing an objective function comprising two additive terms. The first term ensures that the estimated depth map is locally smooth consistent with the color image and the second term ensures consistency of the estimated upsampled data with the low resolution observed data at the corresponding locations. Depth map upsampling is then achieved by solving a large sparse linear system following a similar approach as was done for matting in [13]. A key difference between the matting problem and our approach is that we model the depth as a *linear function of the local spatial coordinates* and not as a linear function of the image intensity values.

The paper is organized as follows: Section 2 describes the scheme of our algorithm. We present both the quantitative and the qualitative results in Section 3, and conclude the paper in Section 4.

2. PROPOSED ALGORITHM

2.1. Problem Formulation

Our proposed method is motivated by the fact that regions of the image that correspond to a smooth 3D surface, can be locally approximated by a plane (for example, via a Taylor series expansion). Thus, over each small patch in the image in regions corresponding to smooth surfaces, a local linear fit (in spatial coordinates) provides a good approximation to the depth. To account for edges, where the assumption breaks down, adaptive nonnegative weights are introduced for the linear fitting. The weighting seeks to effectively concentrate the linear fit at each point on the neighboring pixel locations that are hypothesized, based on their color similarity to the pixel of interest, to be on the same side of the edge. The weights can be obtained from one of several edge preserving techniques, for example, non local mean or bilateral filter. The upsampled depth map is obtained by minimizing an overall objective function that combines a term corresponding to the weighted deviation from the local linear fitting with a data fidelity term that penalizes deviations from observations at the locations where the low resolution depth map is available.

To formally describe our algorithm we use the simplified 1D representation in Fig. 1 that illustrates the contribution of one pixel to the objective function. The axis G represents the relative pixel positions of points in local pixel neighborhood of the target pixel which is located at $G = 0$. The low resolution depth map, denoted by D_L, is available at a subset of the pixel locations in the neighborhood as indicated in the figure

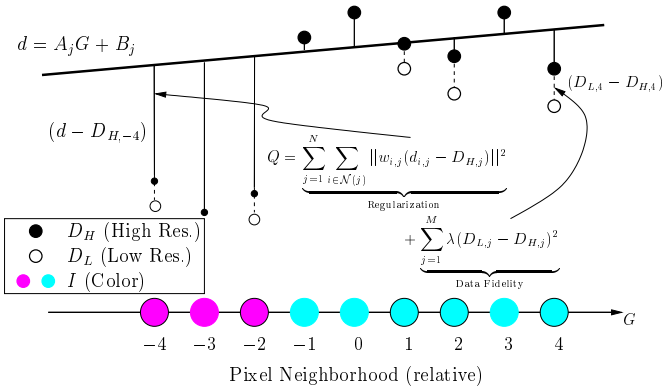

Fig. 1. The illustration of problem formulation in 1D. The magenta and cyan points show different color pixels in the patch of color image, and the circles around data points indicate the available low resolution depth values. The filled and un-filled circles mean the desired upsampled depth map and the input depth map, respectively, and the line is the fitting result on the example area. The weights are illustrated by the size of filled circles.

and color values, denoted by I, form the high resolution RGB image. The goal is to estimate a high resolution depth map D_H. Our objective function is formulated as

$$Q = \sum_{j=1}^{N} \sum_{i \in \mathcal{N}(j)} ||w_{i,j}(d_{i,j} - D_{H,i})||^2 + \sum_{j=1}^{M} \lambda (D_{L,j} - D_{H,j})^2, \tag{1}$$

where j indexes the pixel locations in the upsampled image, N is the number of pixels in the upsampled image, M is the number of pixels in the low resolution depth map, $d_{i,j}$ is the value of linear fitting of pixel i in the neighbor of pixel j, $D_{H,i}$ is the estimated depth at pixel i in the neighbor of pixel j, and $D_{L,j}$ is the depth value at the pixel j of the low resolution depth map, $w_{i,j}$ is the similarity metric of pixel i and j, and λ is the free parameter to control the relation of fidelity and smoothness. The local linear fit is defined as

$$d_{i,j} = A_j G_{i,j} + B_j, \tag{2}$$

where A_j and B_j are the parameters for linear fitting at pixel $D_{H,j}$, A_j is a 1-by-2 vector and B_j is a scalar, and $G_{i,j}$ is a 2-by-1 vector denoting the relative coordinate of pixel i in the neighbor window of pixel j. We define $G_{i,j} \stackrel{def}{=} \{(x,y)| - w_s < x < w_s, -w_s < y < w_s\}$, where w_s is the size of the window. The first term in (1) is the regularization term and the second term is the data fidelity. The formulation is readily extended to the hole filling problem by adding, to the fidelity penalty term, a product with the indicator function of non-missing points and pixel values.

The weights $w_{i,j}$ are defined as

$$w_{i,j} = \exp - \frac{||I_i - I_j||^2}{2\sigma^2}, \tag{3}$$

where I_i and I_j are pixel values of the RGB image I at corresponding position, and σ controls the relative emphasis of

This paper appears as Y. Zhang, L. Ding, and G. Sharma, A local-linear-fitting-based matting approach for accurate depth upsampling, in IEEE Western NY Image and Signal Processing Workshop, Nov 2016.

pixel similarity in the allocation of weights. Alternative, formulations of the weights such as those used in non local mean or bilateral filter can also be used in the proposed framework. Unlike the typical bilateral filter, we do not use the distance decay term in (3) because the window we use is quite small comparing to the high resolution images.

Our problem formulation and the algorithmic approach we use for the solution (described in the next section) are inspired by Levin's formulation of matting as an optimization problem [13], where the alpha channel is formulated as a weighted linear combination of neighboring color values. A key difference in our formulation is that the our weighted local linear fit is formulated in terms of the local relative *spatial* position for the neighborhood, whereas in [13] the weighted linear fit is performed on the *color* values for the neighborhood pixels.

2.2. Optimization Solution

Rewriting (1) in the matrix form, we obtain

$$Q = \sum_{j=1}^{N} (W_j(D_{H,N_j} - GP_j^T))^2 + \lambda F_j, \tag{4}$$

where $G = [G_j, 1]$ and $P_j = [A_j, B_j]^T$. W_j is a diagonal matrix with $w_{i,j}$ being its diagonal entries. D_{H,N_j} is the depth value in the patch[1]. The matrix P_j can be eliminated by replacing it in (4) by its optimal value

$$\begin{aligned} P_j &= \underset{P_j}{\operatorname{argmin}} \left((W_j(D_{H,N_j} - GP_j^T))^2\right) \\ &= (G^T W_{0,j}^T G)^{-1} G^T W_{0,j}^T D_{H,N_j}, \end{aligned} \tag{5}$$

where $W_{0,j}$ is the diagonal matrix $W_{0,j} = W_j^T W_j$.

Replacing P_j in (1) by (5), we obtain,

$$Q = \sum_{j=1}^{N} D_{H,N_j}^T (\overline{G_j}^T W_{0,j} \overline{G_j}) D_{H,N_j} + \sum_{j=1}^{M} \lambda (D_{L,j} - D_{H,j})^2, \tag{6}$$

where $\overline{G_j} = E - G(G^T W_{0,j}^T G)^{-1} G^T W_{0,j}$, with E denoting the identity matrix.

The minimizer for the quadratic objective function Q is readily obtained, specifically, as the solution to the linear equation,

$$LD + \lambda A(D - d) = 0, \tag{7}$$

where $L = \sum_{j=1}^{N} \overline{G_j}^T W_{0,j} \overline{G_j}$ is the Laplacian matrix [14], and A is a diagonal matrix indicating the correspondence of pixels in low resolution map to the upsampled map.

3. EXPERIMENTAL RESULTS

We test our algorithm on the Middlebury (stereo) dataset [15, 16, 17, 18], which provides high resolution RGB images of multiple views and corresponding disparity maps, which are used as the ground truth in our experiment. We use a window size of 7×7 ($\equiv N = 49$), and $\lambda = 10^5$. The RGB-D images are zero-padded for consistent use of (4), and the padded area is cropped out in the final results. The parameter σ^2 in (3) for computation of the weights $w_{i,j}$ is set to one third of the local variance in each window. In each patch, the weight of the center pixel is set to 10^{-5}. We use the built-in Matlab conjugate gradient solver (*cgs*) for solving (7) (a tolerance of 10^{-10} and maximum number of iteration 10^4 were used).

3.1. Qualitative and Quantitative Results

The proposed algorithm is both suitable for hole filling for single disparity map and depth map upsampling, as indicated earlier. In this part, we first visually examine the performance of filling holes in depth map, as shown in Fig. 2. From the images in the last column, we can find that the holes, which correspond to the occluded area in the disparity map, are well filled. Unlike traditional interpolation methods, our algorithm is able to fix the holes in the depth images, so as to keep the consistency of depth map edges with those in the RGB images and avoid smoothing in such areas. For example, see the third row of Fig. 2. The missing points along the wall are well fitted to the two sides, and not blurred as a large patch.

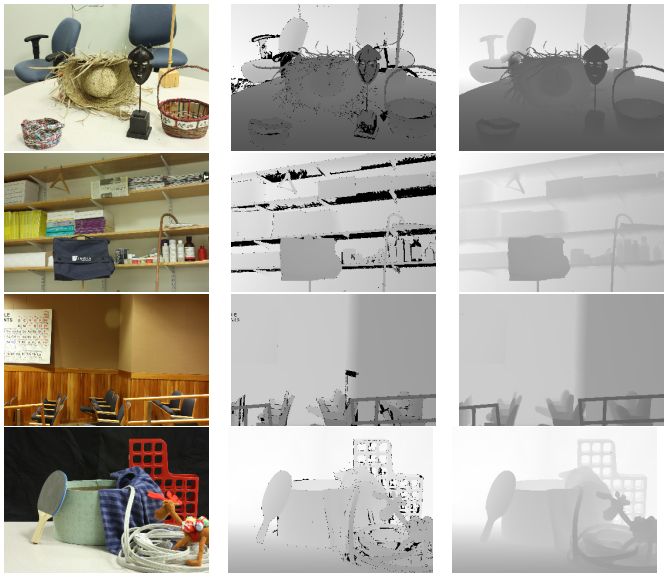

Fig. 2. Qualitative results of the hole filling ability of our algorithm, tested on Middlebury stereo dataset 2014 [18]. The first column shows the input high resolution color images and the second column shows the corresponding depth maps. The results are shown in the last column. The images are processed at a low resolution of approximate 60k to 80k pixels.

Quantitative results comparing the proposed algorithm against prior work are obtained on the Middlebury stereo dataset 2005 [15]. We first downsample the input depth map to obtain the low resolution version, and then run different algorithms on these images to obtain the upsampled versions. We use mean absolute error (MAE) as the metric to evaluate

[1]We pad the image to represent G consistently at all positions.

This paper appears as Y. Zhang, L. Ding, and G. Sharma, A local-linear-fitting-based matting approach for accurate depth upsampling, in IEEE Western NY Image and Signal Processing Workshop, Nov 2016.

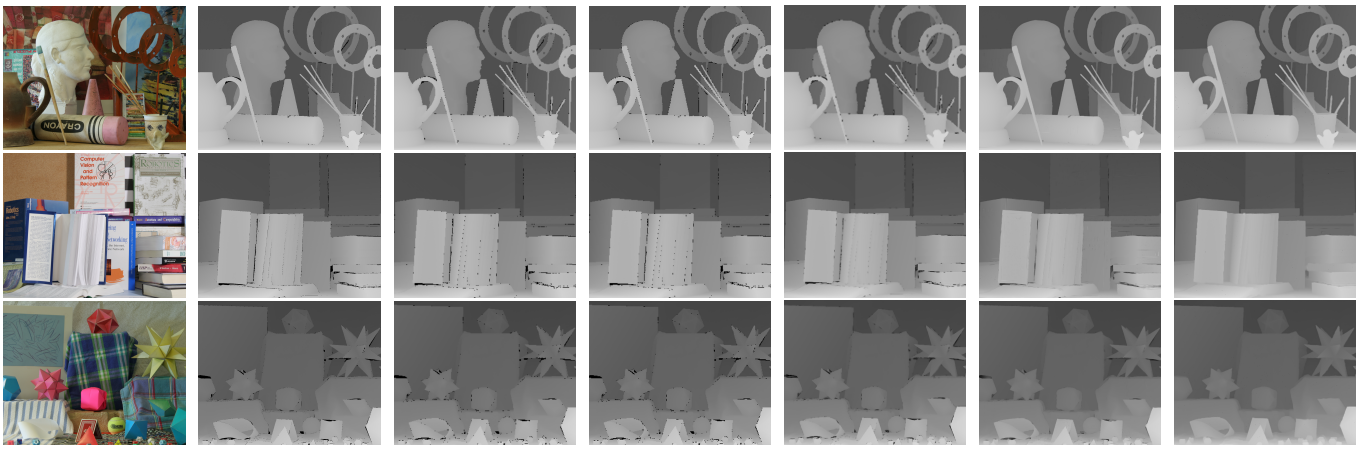

Fig. 3. Visual comparison of different algorithm tested on Middlebury dataset [15] at $4\times$ upsampling rate. Column from left to right: RGB images, ground truth, and the results of: bilinear, bicubic, IBL [8], TGV [7] and proposed; row from top to bottom: Art, Books, and Moebius.

	Images											
	Art				Books				Moebius			
Methods \ Sample Rate	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
bicubic	0.8965	1.4298	2.4363	4.3456	0.7911	1.0842	1.7031	2.5419	0.6855	1.0287	1.5821	2.5527
bilinear	0.7642	1.2300	2.1495	3.9500	0.6620	0.8993	1.4183	2.1174	0.5685	0.8578	1.3347	2.1942
IBL [3]	0.5016	0.8934	**1.7028**	4.2324	0.2790	0.7361	1.4056	2.4561	0.3987	0.7071	1.1289	2.5885
TGV [7]	0.6457	0.8926	3.2633	7.6490	0.5980	0.7507	2.3091	6.324	0.4722	0.5627	2.0375	6.6210
Proposed	**0.4423**	**0.8765**	1.7616	**3.6033**	**0.1986**	**0.3594**	**0.6655**	**1.1888**	**0.1864**	**0.3426**	**0.6478**	**1.2393**

Table 1. Quantitative comparison of different algorithms tested on Middlebury dataset [15]. The results is evaluated as MAE (the smaller the better) for four different sample rates, as listed in the second row. All values are computed based on the disparity maps. The best result in each situation is in bold font.

the performance of different algorithms. Table 1 summarizes the results[2]. Fig. 3 shows the corresponding visual results at the upsampling rate of 4. The results show that our algorithm is particularly suitable for depth map upsampling, and our algorithm provides a better result compared with the other algorithms.

3.2. Discussion

Table 1 illustrates that the depth map upsampling obtained with the proposed algorithm is accurate and achieves the state of the art results on the common benchmarking dataset, providing a better performance compared with other algorithms. The proposed algorithm, however, still suffers from two limitations. First, there are a few outlier points where the method yields a large error. Second, the edges are not sharply defined, especially under high upsampling rate, which is typical in most depth upsampling algorithms. The computational requirements are an additional challenge: to process a 1088×1296 pixel image, our algorithm takes about 40min. While the time requirement is analogous for several other upsampling algorithms, a speed-up is desirable for many applications. In our future work, we aim at alleviating these limitations by implementing some post processing using segmentation techniques, and by adopting computational speed-up techniques that have already been successfully applied in other very similarly structured optimization problems [19]. Although desirable, quantitative assessment of the impact of the improved depth map on subsequent processing is not considered in this paper because of the challenges of application and content dependence.

[2]Code for IBL implementation is provided by Chunhua Shen: https://bitbucket.org/chhshen/depth-enhancement.

4. CONCLUSION

The algorithm proposed in this paper provides an effective method for depth map upsampling and hole filling. Quantitative results on test data indicates that the method offers an improvement over current state of the art methods and visual assessment shows that the depth map estimated by the proposed technique is consistent with the color images.

5. ACKNOWLEDGMENT

We thank the Center for Integrated Research Computing, University of Rochester, for providing access to computational resources.

This paper appears as Y. Zhang, L. Ding, and G. Sharma, A local-linear-fitting-based matting approach for accurate depth upsampling, in IEEE Western NY Image and Signal Processing Workshop, Nov 2016.

6. REFERENCES

[1] L. Ding and G. Sharma, "Fusing structure from motion and lidar for accurate dense depth map estimation," 2017, submitted to IEEE ICASSP 2017.

[2] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Adv. in Neural Info. Proc. Sys.*, vol. 5, 2005, pp. 291–298.

[3] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2007, pp. 1–8.

[4] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. European Conf. Computer Vision*, 2010, pp. 1–14.

[5] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM Trans. on Graphics*, vol. 26, no. 3, 2007, p. 96.

[6] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *IEEE Intl. Conf. Comp. Vision.*, 2011, pp. 1623–1630.

[7] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *IEEE Intl. Conf. Comp. Vision.*, 2013, pp. 993–1000.

[8] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," in *IEEE Trans. Image Proc.*, vol. 23, no. 8, 2014, pp. 3443–3458.

[9] J. Lu, D. Min, R. S. Pahwa, and M. N. Do, "A revisit to MRF-based depth map super-resolution and enhancement," in *IEEE Intl. Conf. Acoust., Speech, and Signal Proc.*, 2011, pp. 985–988.

[10] A. Harrison and P. Newman, "Image and sparse laser fusion for dense scene reconstruction," in *Field and Service Robotics*, 2010, pp. 219–228.

[11] S. Krishnamurthy and K. R. Ramakrishnan, "Image-guided depth map upsampling using normalized cuts-based segmentation and smoothness priors," in *IEEE Intl. Conf. Image Proc.*, 2016, pp. 554–558.

[12] K. Uruma, K. Konishi, T. Takahashi, and T. Furukawa, "High resolution depth image recovery algorithm based on the modeling of the sum of an average distance image and a surface image," in *IEEE Intl. Conf. Image Proc.*, 2016, pp. 2836–2840.

[13] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," in *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 30, no. 2, 2008, pp. 228–242.

[14] R. Merris, "Laplacian matrices of graphs: a survey," *Linear algebra and its applications*, vol. 197, pp. 143–176, 1994.

[15] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, vol. 1, 2003, pp. I–195.

[16] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2007, pp. 1–8.

[17] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2007, pp. 1–8.

[18] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *German Conf. on Pattern Recog.*, 2014, pp. 31–42.

[19] C. Yu, G. Sharma, and H. Aly, "Computational efficiency improvements for image colorization," in *IS&T/SPIE Electronic Imaging*, 2014, pp. 902 004–902 004.