CORRECTING SHARPNESS VARIATIONS IN STEREO IMAGE PAIRS

C. Doutre and P. Nasiopoulos

Departement of Electrical and Computer Engineering
University of British Columbia, Vancouver, Canada

Abstract

When stereo images are captured under less than ideal conditions, there may be inconsistencies between the two images in brightness, contrast, blurring, etc. When stereo matching is performed between the images, these variations can greatly reduce the quality of the resulting depth map. In this paper, we propose a method for correcting sharpness variations in stereo image pairs which is performed as a pre-processing step to stereo matching. We modify the more blurred of the two images so that it will match the less blurred image in sharpness as closely as possible. The 2D discrete cosine transform (DCT) of both images is taken, and the DCT coefficients are divided into a number of frequency bands. The signal and noise energy in each band are estimated, and the coefficients are scaled so that the two images have equal signal energy in each band. Experiments show that applying the proposed correction method can greatly improve the disparity map quality when one image in a stereo pair is more blurred than the other.

Keywords: stereo images, sharpness, blur, depth estimation, discrete cosine transform (DCT).

1 Introduction

Stereo matching is a classical problem in computer vision that has been extensively studied [1]. It has many applications such as 3D scene reconstruction, image based rendering, and robot navigation. Most stereo matching research uses high quality datasets that have been captured under very carefully controlled conditions. However, capturing well-calibrated high quality images is not always possible, for example when cameras are mounted on a robot [2], or simply due to a low cost camera setup being used. Inconsistencies between cameras can cause the images to differ in brightness, colour, sharpness, contrast, etc. These differences reduce the correlation between the images and make stereo matching more challenging, resulting in lower quality depth maps.

A number of techniques have been proposed to make stereo matching robust to radiometric differences between images (i.e., variations in brightness, contrast, vignetting). An evaluation of several matching techniques that are robust to radiometric differences is presented in [3]. Much less work has addressed stereo matching when there are variations in sharpness/blurring between images. Variations in image sharpness may result from a number of causes, such as the cameras being focused at different depths, variations in shutter speed, and camera shake causing motion blur.

In [4], a stereo method is proposed using a matching cost based on phase quantization. Their phase-based matching is invariant to convolution with a centrally symmetric point spread function (psf), and hence is robust to blurring. In [5] a method is proposed for performing stereo matching on images where a small portion of the images suffers from motion blur. A probabilistic framework is used, where each region can be classified as affected by motion blur or not. Different smoothness parameters in an energy minimization step are used for the pixels estimated as affected by motion blur. Note that neither [4] nor [5] attempts to correct the blurring in the images (they do not modify the input images); instead, they attempt to make the matching process robust to blurring.

In this paper, we propose a fast method for correcting sharpness variations in stereo images. Unlike previous works [4], [5] the method is applied as pre-processing before depth estimation. Therefore, it can be used together with any stereo method. Our method takes a stereo image pair as input, and modifies the more blurry image so that it matches the sharper image. Experimental results show that applying the proposed method can greatly improve the quality of the depth map when there are variable amounts of blur between the two images. The rest of this paper is organized as follows. The proposed method is described in section 2, experimental results are given in section 3 and conclusions are drawn in section 4.

2 Proposed Sharpness Correction Method

When an image is captured by a camera, it may be degraded by a number of factors, including optical blur, motion blur, and sensor noise. Hence in a stereo image pair, the captured left and right images can be modelled as:

$$\begin{aligned} \tilde{i}_L(x,y) &= h_L(x,y) * i_L(x,y) + n_L(x,y) \\ \tilde{i}_R(x,y) &= h_R(x,y) * i_R(x,y) + n_R(x,y) \end{aligned} \quad (1)$$

where *i(x,y)* is the "true" or "ideal" image and *h(x,y)* is the point spread function (psf) of the capturing process. The subscripts *L* and *R* denote the left and right images. The '*' operator represents two dimensional convolution. The *n(x,y)* term is additive noise, which is usually assumed to be independent of the signal and normally distributed with some variance σ_n^2. Throughout the paper, we will use the tilde '~' to denote an observed (and hence degraded) image. The psf,

978-0-7695-3893-8/09 $26.00 © 2009 IEEE
DOI 10.1109/CVMP.2009.7

h(x,y), is usually a low-pass filter, which makes the observed image blurred (high frequency details are attenuated).

If the same amount of blurring occurs in both images, i.e., h_L and h_R are the same, the images may lack detail but they will still be consistent. Therefore, stereo matching will still work reasonably well. We are interested in the case where different amounts of blurring occur in the images, so h_L and h_R are different. Our method attempts to make the more blurred image match the less blurred image in sharpness, by scaling the DCT coefficients of the more blurred image. The basis for our method is that un-blurred stereo images typically have very similar frequency content, so that the signal energy in a frequency band should closely match between the two images. Therefore, we scale the DCT coefficients in each frequency band so that after scaling the image that originally had less energy in the band will have the same amount of energy as the other image. The resulting corrected images will match closely in sharpness, making stereo matching between the images more accurate. The steps of our method are described in detail in the following subsections.

2.1 Removing non-overlapping edge regions

Typical stereo image pairs have a large overlapping area between the two images. However, there is usually also a region on the left side of the left image and a region on the right side of the right image that are not visible in the other image (Fig. 1a). If these non-overlapping areas are removed, the assumption that the two images will have similar frequency content will be stronger.

In order to identify the overlapping region between the two images, we consider two strips; one along the right edge of i_L and one along the left edge of i_R (see Figure 1a, regions highlighted in red). Strips five pixels wide are used in our experiments. We find matching strips in the other image using simple block based stereo matching. Using the sum of absolute differences (SAD) as a matching cost, two SAD values are calculated for each possible disparity d, one for the edge of i_L and one for the edge of i_R:

$$SAD_L(d) = \sum_{(x,y)\in edge_L} \left| \tilde{i}_L(x,y) - \tilde{i}_R(x-d,y) \right|$$
$$SAD_R(d) = \sum_{(x,y)\in edge_R} \left| \tilde{i}_R(x,y) - \tilde{i}_L(x+d,y) \right| \quad (2)$$

The disparity value d that minimizes the sum $SAD_L(d) + SAD_R(d)$ is chosen as the edge disparity D. Cropped versions of i_L and i_R are created by removing D pixels from the left of i_L and D pixels from the right of i_R (Figure 1b). These cropped images, which we will denote i_{Lc} and i_{Rc}, contain only the overlapping region of i_L and i_R.

In equation (2), we have used the standard sum of absolute differences as the matching cost. If there are variations in brightness between the images, a more robust cost should be

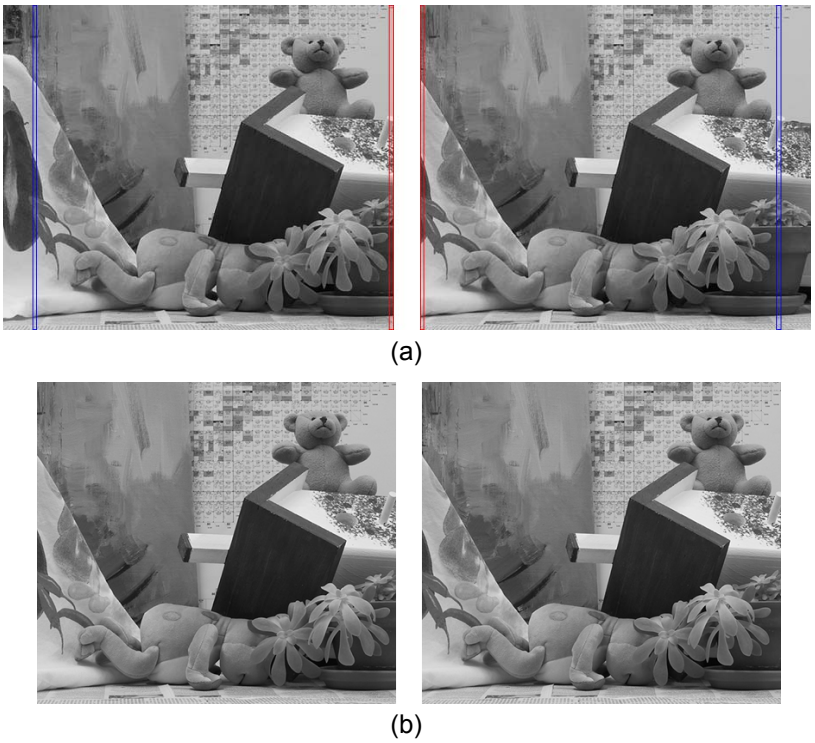

Figure 1: Removing non-overlapping edge regions. (a) Left and right original images with edge strips used in search shown in red and matching regions found through SAD search (equation (2)) shown in blue. (b) Images cropped with non-overlapping regions removed.

used, such as normalized cross correlation or mean-removed absolute differences.

2.2 Noise variance estimation

Noise can have a significant effect on blurred images, particularly in the frequency ranges where the signal energy is low due to blurring. We wish to remove the effect of noise when estimating the signal energy, which requires estimating the noise variance of each image.

We take the two dimensional discrete cosine transform (DCT) of the cropped images, which we will denote as $\widetilde{I}_{Lc}(u,v)$ and $\widetilde{I}_{Rc}(u,v)$. The indices u and v represent horizontal and vertical frequencies, respectively. These DCT coefficients are affected by the additive noise. We can obtain an estimate for the noise standard deviation from the median absolute value of the high frequency DCT coefficients [6]:

$$\sigma_N = \frac{\underset{u>u_T, v>v_T}{\text{median}} \left(\left| \widetilde{I}\ (u,v) \right| \right)}{0.6745} \tag{3}$$

Values u_T and v_T are the thresholds for deciding which DCT coefficients are classified as high frequency. We have used 20 less than the maximum values for u and v as the thresholds in our tests, resulting in 400 coefficients being used when calculating the median. Using equation (3), we obtain an estimate for the noise level in each image, $\sigma_{N,L}$ and $\sigma_{N,R}$.

2.3 Division into frequency bands

We wish to correct the full left and right images, without the cropping described in section 2.1. Therefore, we also need to take the DCT of the original images, $\widetilde{I}_L(u,v)$ and $\widetilde{I}_R(u,v)$, so that those coefficients can be scaled. This is in addition to taking the DCT of the cropped images $\widetilde{I}_{Lc}(u,v)$ and $\widetilde{I}_{Rc}(u,v)$, from which we will calculate the scaling factors. The DCT coefficients of each image have the same dimensions as the image in the spatial domain. If the width and height of the original images are W and H, the cropped images will have dimensions (W-D)xH. In the DCT domain, the dimensions of the coefficients will also be WxH for the original images and (W-D)xH for the cropped images.

We divide the DCT coefficients of both the original and cropped images into a number of equally sized frequency bands. Each frequency band consists of a set of (u,v) values such that $u_i \le u < u_{i+1}$ and $v_j \le v < v_{j+1}$, where u_i is the starting index of band i in the horizontal direction and v_j is the starting index of band j in the vertical direction. If we use M bands in both the horizontal and vertical directions, then the starting frequency index of each band in the original and cropped images can be calculated as:

$$\begin{aligned} u_i &= \text{round}\left(\frac{i \cdot W}{M}\right), \quad v_j = \text{round}\left(\frac{j \cdot H}{M}\right) \\ u_{i,c} &= \text{round}\left(\frac{i \cdot (W-D)}{M}\right), \quad v_{j,c} = \text{round}\left(\frac{j \cdot H}{M}\right) \end{aligned} \tag{4}$$

where u_i and v_j are the indexes for the original images, and $u_{i,c}$ and $v_{j,c}$ are the indexes for the cropped images. Although u_i and $u_{i,c}$ are different numbers, they correspond to the same spatial frequencies.

The number of frequency bands to use in each direction, M, is a parameter that must be decided. If more bands are used, the correction can potentially be more accurate. However, if too many bands are used, each band will contain little energy and therefore the estimate for the scaling factor will be less reliable. Experimentally, we have determined that using M=20 frequency bands is works well over a wide set of stereo images and levels of blurring.

2.4 DCT coefficient scaling

We wish to scale the coefficients in each frequency band so that the left and right images have the same amount of signal energy in the band. The energy of each observed (degraded) image in band ij can be computed as:

$$\text{En}_{ij}\left(\widetilde{I}\ \right) = \sum_{u=u_i}^{u_{i+1}-1} \sum_{v=v_i}^{v_{i+1}-1} \left(\widetilde{I}\,(u,v)\right)^2 \tag{5}$$

We wish to remove the effect of noise from the energy calculated with (5). Let us define $HI(u,v)$ as the DCT of the blurred signal $h(x,y) * i(x,y)$, and define $N(u,v)$ as the DCT of the noise. Since we are using an orthogonal DCT, $N(u,v)$ is also normally distributed with zero mean and variance σ_N^2. Given that the noise is independent of the signal and zero mean, we can calculate the expected value of the energy of the observed signal:

$$\begin{aligned} E\left[\left(\widetilde{I}\ (u,v)\right)^2\right] &= E\left[\left(HI(u,v) + N(u,v)\right)^2\right] \\ &= E\left[\left(HI(u,v)\right)^2\right] + E\left[\left(N(u,v)\right)^2\right] \\ &= E\left[\left(HI(u,v)\right)^2\right] + \sigma_N^2 \end{aligned} \tag{6}$$

Summing the above relation over all the DCT coefficients in a frequency band gives:

$$\begin{aligned} \sum_{u=u_i}^{u_{i+1}-1} \sum_{v=v_i}^{v_{i+1}-1} E\left[\left(\widetilde{I}\,(u,v)\right)^2\right] &= \sum_{u=u_i}^{u_{i+1}-1} \sum_{v=v_i}^{v_{i+1}-1} E\left[\left(HI(u,v)\right)^2\right] + \sum_{u=u_i}^{u_{i+1}-1} \sum_{v=v_i}^{v_{i+1}-1} \sigma_N^2 \\ &= \text{En}_{ij}(HI) + C_{ij}\sigma_N^2 \end{aligned} \tag{7}$$

where $\text{En}_{ij}(HI)$ is the energy of the blurred signal in frequency band ij and C_{ij} is the number of coefficients in the band. The left hand side of equation (7) can be estimated with the observed signal energy calculated with equation (6).

Figure 2: Test images used in experiments (only left image of each pair is shown). In reading order: Tsukuba, teddy, cones, art, laundry, moebius, reindeer, aloe, baby1, and rocks

Therefore, we can estimate the energy of the blurred signal in the band as:

$$\mathrm{En}_{ij}(HI) = \max\left(0, \mathrm{En}_{ij}\left(\tilde{I}\right) - C_{ij}\sigma_N^2\right) \tag{8}$$

In (8) we have clipped the estimated energy to be zero if the subtraction gives a negative result since the energy must be positive (by definition). Using (8) we estimate the signal energies $\mathrm{En}_{ij}(HI_L)$ and $\mathrm{En}_{ij}(HI_R)$. We wish to multiply the coefficients in the image with less energy by a gain factor (G_{ij}) so that it has the same amount of signal energy as the other image. The scale factors to apply to each image in this band can be found as:

$$\mathrm{En}_{ij,\max} = \max\left(\mathrm{En}_{ij}(HI_L), \mathrm{En}_{ij}(HI_R)\right) \tag{9}$$

$$G_{ij,L} = \sqrt{\frac{\mathrm{En}_{ij,\max}}{\mathrm{En}_{ij}(HI_L)}} \qquad G_{ij,R} = \sqrt{\frac{\mathrm{En}_{ij,\max}}{\mathrm{En}_{ij}(HI_R)}} \tag{10}$$

Note that either $G_{ij,L}$ or $G_{ij,R}$ will always be one, because no gain needs to be applied to the sharper image. The gain factors calculated with (10) do not consider the effect of noise. If the signal energy is very low, the gain will be very high, and noise may be amplified excessively (this is a common issue in de-blurring methods [7]). To prevent noise amplification from corrupting the recovered image, we clip the gain factors to have a maximum value of 4.

The scaling factors, $G_{ij,L}$ and $G_{ij,R}$ are calculated based on the DCT coefficients of the cropped images (because the assumption of equal signal energy in each frequency band will be stronger for the cropped images). So equations (6) through (10) are all applied only to the DCT coefficients of the cropped images. Once the gains are calculated for a frequency band *ij*, we scale the DCT coefficients of the original images to get the DCT coefficients of the corrected images:

$$I_{L,cor}(u,v) = G_{ij,L} \cdot \tilde{I}_L(u,v) \qquad I_{R,cor}(u,v) = G_{ij,R} \cdot \tilde{I}_R(u,v) \tag{11}$$

After calculating all of the scaling factors, and applying equation (11) for every frequency band, we will have the complete DCT coefficients of the corrected images, $I_{L,cor}(u,v)$ and $I_{R,cor}(u,v)$. Then we simply take the inverse DCT to get the final corrected images in the spatial domain.

3 Results

We test our proposed method on ten stereo image pairs from the Middlebury stereo page [8] that have ground truth disparities obtained through a structured light technique [9]. Thumbnails of the test images are shown in Figure 2. The 2005 and 2006 data sets from the Middlebury page (art, laundry, moebius, reindeer, aloe, baby1, rocks) all have seven views; we used the one-third size versions of views 1 and 3 in our tests.

Our sharpness correction method is a pre-processing step done before stereo matching, and therefore it can be used together with any stereo method. We test our method together with a representative fast global stereo method that solves a 2D energy minimization problem using Belief Propagation (BP) [10]. We refer readers to [10] for the details of the BP stereo algorithm.

Two kinds of blurring filters are tested. Out-of-focus blur, which is modelled with a disk filter of a given radius [11], and linear motion blur, which is modelled as an average of samples along a straight line of a given length. We use the MATLAB commands `fspecial('disk',…)` and `fspecial('motion',…)` to generate the blurring filters.

The performance metric we use is the percentage of 'bad' pixels in the disparity map in the un-occluded regions of the image. A bad pixel is defined as one whose estimated disparity differs from the true disparity by more than one

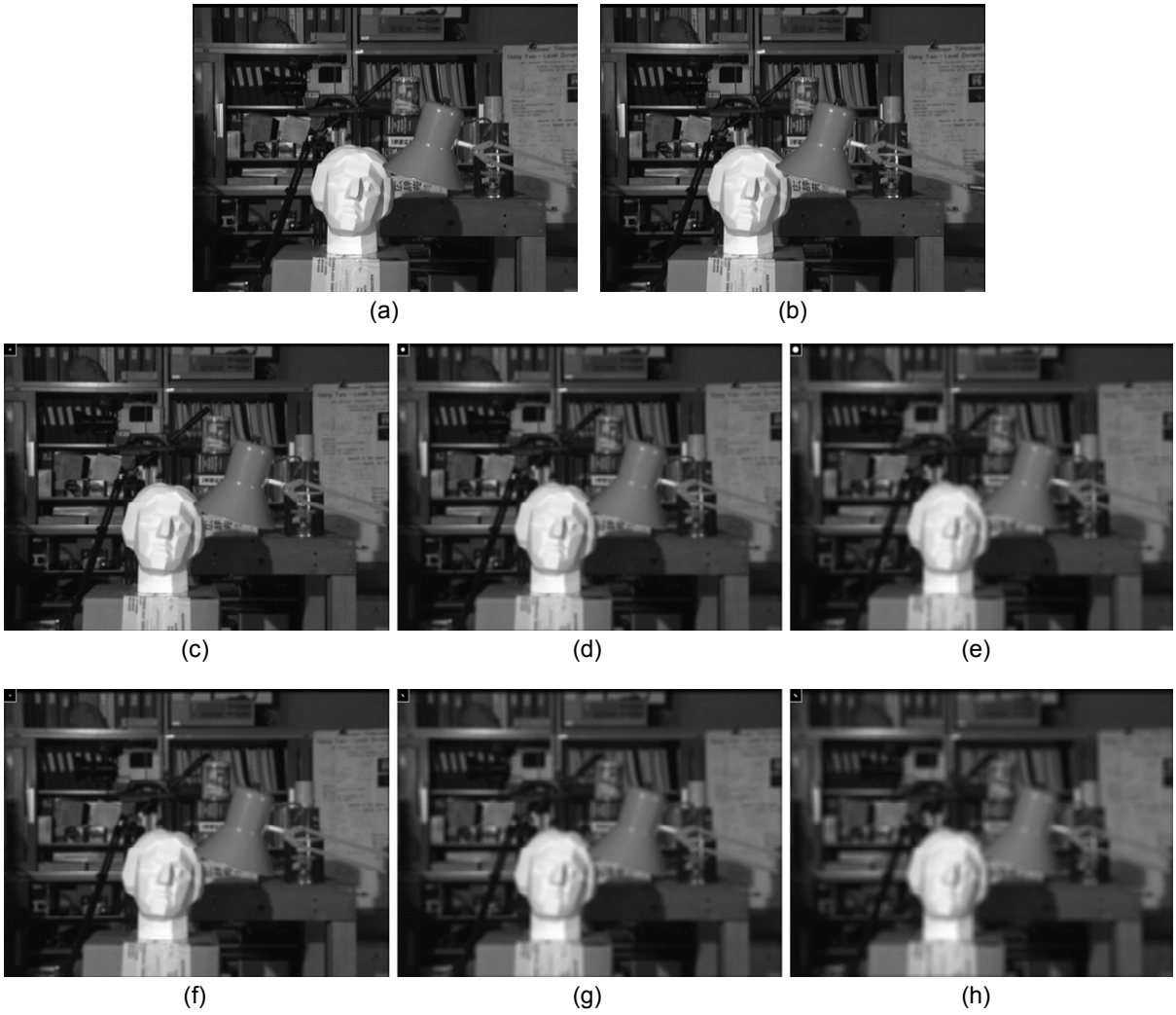

Figure 3: Demonstration of blurring filters used in our tests on the Tsukuba images (a) Original left image (b) Original right image, (c)-(h) Left image blurred with: (c) Out-of-focus blur, radius 1 (d) Out-of-focus blur, radius 2 (e) Out-of-focus blur, radius 3 (f) motion blur, length 2 (g) motion blur, length 3 (h) motion blur, length 4. The blurring filter is illustrated in the top left corner of each image.

pixel. This is the most commonly used quality measure for disparity maps, and it has been used in major studies such as [1] and [3].

To show the effectiveness of our proposed method, we compare the quality of depth maps obtained using our proposed method relative to performing stereo matching directly of the blurred images. In each test, the right image was left unfiltered, while the left image was blurred with either a disk filter (simulating out-of –focus blur) or a linear motion blur filter at 45 degrees to the x axis. We tested of-focus-blur with radii of 0, 1, 2 and 3 pixels and motion blur with lengths of 2, 3 and 4 pixels. A larger radius or length means the image is blurred more. A blur radius of zero means the image is not blurred at all, i.e., the filter is an impulse response and convolving it with the image leaves the image unaltered. White Gaussian noise with a variance of 2 was added to all of the blurred images (which is typical of the amount of noise found in the original images). Figure 3 shows the Tsukuba image blurred with all of the filters tested, to give the reader an idea of how severe the blurring is in different tests.

Tables 1 and 2 show the percentage of errors in the disparity maps obtained when different levels of blurring are applied, with and without the proposed correction. The second column of each table (images) shows whether stereo matching was performed on either the blurred left image and original right image (the "blurred" case), or on the left-right pair obtained by applying our proposed method (the "corrected" case). Table 1 gives results for of-out-focus blur and Table 2 gives results for motion blur.

Radius	Images	Teddy	Cones	Tsukuba	Art	Laundry	Moebius	Reindeer	Aloe	Baby1	Rocks	**Average**
0	Blurred	14.8	9.7	2.0	6.8	13.2	7.9	5.9	2.8	1.6	3.8	**6.8**
	Corrected	12.3	5.0	2.4	7.2	12.4	6.1	3.6	2.8	1.5	3.3	**5.7**
1	Blurred	16.8	13.5	3.1	8.7	13.2	8.2	7.9	3.8	1.7	4.2	**8.1**
	Corrected	11.6	5.4	3.5	7.8	11.1	6.3	5.6	3.6	1.5	2.9	**5.9**
2	Blurred	28.2	28.6	5.1	12.9	18.9	13.0	14.6	8.6	6.4	6.6	**14.3**
	Corrected	18.3	8.1	5.3	11.1	13.1	8.9	11.8	6.4	2.1	4.9	**9.0**
3	Blurred	40.7	56.0	10.7	26.4	35.0	26.3	24.2	32.3	26.1	16.4	**29.4**
	Corrected	31.6	24.3	8.8	19.4	21.5	17.4	22.2	20.2	4.3	10.7	**18.0**

Table 1: Percentage of errors in disparity maps with out-of focus blurring

Length	Images	Teddy	Cones	Tsukuba	Art	Laundry	Moebius	Reindeer	Aloe	Baby1	Rocks	**Average**
2	Blurred	15.6	10.7	2.3	7.7	12.3	7.8	7.2	3.2	1.6	3.9	**7.2**
	Corrected	11.8	5.4	3.1	7.7	11.4	5.8	5.0	3.2	1.6	3.3	**5.8**
3	Blurred	15.9	12.2	2.6	8.0	11.1	8.3	8.2	3.5	1.8	3.7	**7.5**
	Corrected	11.6	6.1	2.5	8.0	10.4	6.7	5.5	3.5	1.6	3.1	**5.9**
4	Blurred	21.4	20.5	3.4	9.5	14.3	10.4	10.9	4.5	3.0	4.6	**10.2**
	Corrected	16.7	7.3	3.6	9.1	13.1	8.3	6.9	3.9	1.8	2.8	**7.3**

Table 2: Percentage of errors in disparity maps with linear motion blur

From Tables 1 and 2, we can see there is a substantial reduction in the number of errors in the disparity maps when the proposed correction is used, particularly when the amount of blurring in the left image is high. For the case of out-of-focus blur with a radius of 3, the average number of errors is reduced from 29.4% to 18% using our proposed method. For all images, there is some improvement when the proposed correction is used if one image is blurred.

Even when neither image is blurred at all (the out-of-focus, zero radius case in Tables 1), there is some improvement on average when using the proposed method. The average amount of errors is reduced from 6.8% to 5.7% on the original image pairs. This suggests that some the original images may have slightly different levels of sharpness that the proposed method can correct.

An example demonstrating the subjective visual quality of the corrected images and resulting disparity maps is shown in Figure 4. The blurred left image and original right image of the cones stereo pair is shown, along with the result of correcting the image pair with our method. The disparity maps obtained based on the blurred pair and corrected pair are also shown. Comparing (c) and (f) in Figure 4, we can see that the proposed method greatly reduces the errors in the disparity map, and produces more accurate edges. We can also see that the corrected images, (d) and (e), are perceptually much closer in sharpness than the blurred images, (a) and (b). Therefore, the proposed method may also be useful in applications such as Free Viewpoint TV 0 and other multiview imaging scenarios, for making the subjective quality of different viewpoints uniform.

We have implemented our proposed method in C code, and the running time for the Tsukuba images is 62 ms on an Intel Core 2 E4400 2 GHz processer under Windows XP.

4 Conclusions

In this paper we have proposed a pre-processing method for correcting sharpness variations in stereo image pairs. We modify the more blurred image to match the sharpness of the less blurred image as closely as possible, taking noise into account. The DCT coefficients of the images are divided into a number of frequency bands, and the coefficients in each band are scaled so that the images have the same amount of signal energy in each band. Experimental results show that applying the proposed method before estimating depth on a stereo image pair can significantly improve the accuracy of the disparity map compared to performing stereo matching directly on the blurred images, particularly when one image is blurred much more than the other.

References

[1] D. Scharstein and R. Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, 47(1): 7-42, 2002.

[2] D. Murray and J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots*, 8(2): 161–171, 2000.

[3] H. Hirschmuller, and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," *Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007.

[4] M. Pedone and J. Heikkila, "Blur and Contrast Invariant Fast Stereo Matching," *Proc. Advanced Concepts for Intelligent Vision Systems*, pp. 883-890, Oct. 2008.

[5] W. Wang, Y. Wang; L. Huo, Q. Huang, and W. Gao, "Symmetric segment-based stereo matching of motion blurred images with illumination variations," *Proc.*

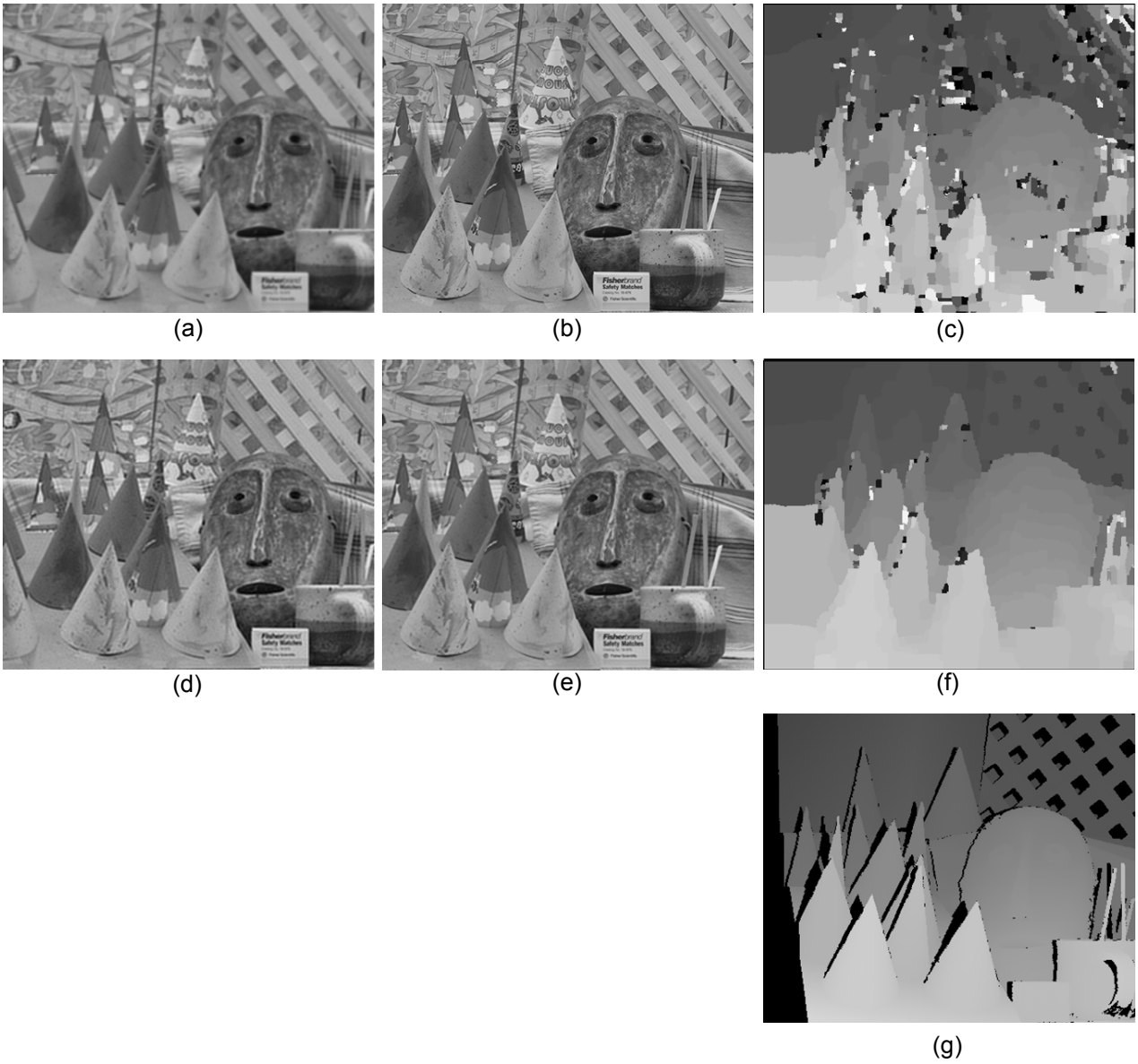

Figure 4: Example of cones image pair before and after correction. (a) Blurred left image (b) Original right image (c) Disparity map obtained from images a and b (d) Corrected left image (e) Corrected right image (f) Disparity map obtained from corrected images with Belief Propagation (g) Ground truth disparities, with black indicating occluded regions

International Conference on Pattern Recognition (ICPR), Dec. 2008.

[6] M. K. Hasan, S. Salahuddin, M. R. Khan, "Reducing signal-bias from MAD estimated noise level for DCT speech enhancement," *Signal Processing*, 84(1): 151-162, 2004.

[7] R. C. Gonzalez and R. E. Woods, Digital Image Processing (2nd Edition). Prentice Hall, Jan. 2002.

[8] D. Scharstein and R. Szeliski, Middlebury stereo vision page, http://vision.middlebury.edu/stereo.

[9] D. Scharstein, R. Szeliski, "High-accuracy stereo depth maps using structured light". *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 195-202, June 2003.

[10] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *International Journal of Computer Vision*, 70(1), 2006.

[11] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(2): 97–108, 1993.

[12] M. Tanimoto, "Free Viewpoint Television – FTV," *Proc. Picture Coding Symposium (PCS 2004)*, Dec. 2004.