

Does Color Really Help in Dense Stereo Matching?

Michael Bleyer*

Institute of Software Technology
Vienna University of Technology
Favoritenstrasse 9-11/188/2, A1040 Vienna, Austria
bleyer@ims.tuwien.ac.at

Sylvie Chambon

Laboratoire Central des Ponts et Chaussées
Route de Pornic, BP 4129
44341 Bouguenais Cedex, Nantes, France
chambon@lcpcc.fr

Abstract

This paper investigates the role of color in global stereo matching approaches. In our evaluation study, we build various energy functions by combining nine color spaces with four dissimilarity functions and test their performance on 30 ground truth stereo pairs. Our experiments start by computing the matching scores via the absolute difference of color values. As is consistent with previous studies, we observe that color-based matching clearly outperforms grey-scale matching. However, our key observation is that this improvement largely stems from considerably improved performance in radiometric distorted regions, i.e. regions where corresponding pixels have different intensities/colors in the two input images, which is e.g. caused by illumination variations. Hence, we claim that color basically serves the same purpose as radiometric insensitive measures, namely to reduce matching errors in radiometric distorted image areas. However, the important difference is that radiometric insensitive measures are considerably superior in this respect, which we demonstrate by using Mutual Information, ZNCC and Census as dissimilarity functions in our experiments. Interestingly, we observe that for these dissimilarity functions color even has a negative effect. Therefore, our suggestion is to not use color at all, but radiometric insensitive measures on grey-scale images, also on images where radiometric distortions seem to be very small.

1. Introduction

Using color intuitively represents a good idea in binocular dense stereo. A color image provides more information than a grey-scale image, and this additional information should help in reducing the ambiguity in stereo matching. For example, consider the case where we have a green pixel in the left image. Let us suppose that we have two match candidates in the right image, i.e. a green and a red pixel. In this case, it is easy to compute the green pixel as being the correct match. Let us now discard the color in-

formation. Red and green might then project to the same intensity value. Hence, matching now becomes ambiguous, and it can be expected that matching performance becomes worse. Despite this obvious advantage of color, there have recently been contradicting statements about whether color should be used in stereo algorithms or not.

There are several previous studies that assess the performance of color-based stereo, either in the context of local methods [5, 8, 9, 10] or, more recently, in the context of global algorithms [2]. These papers consistently claim that color improves the performance in comparison to grey-scale matching. For example, the authors of [2] report a relatively high performance gain, i.e. matching errors are reduced by up to 25% in their experiments. In this work, we will confirm the results of these studies, but we argue that previous papers do not tell the “whole story” about color in stereo matching, i.e. they do not give enough details about the conditions of this improvement.

A recent study [7] has evaluated the performance of different match measures that are insensitive to radiometric distortions. By radiometric distortion we mean that corresponding pixels have different intensity/color values in the two input views, which violates the commonly applied photo-consistency assumption. There are various sources for radiometric distortions such as different camera settings (e.g. in exposure times), vignetting or slightly different illumination conditions under which the images have been acquired. Although the authors of [7] almost exclusively operate on grey-scale images, they also present a preliminary experiment addressing the role of color. Surprisingly, the authors report very little improvement when using color in conjunction with their radiometric insensitive match measures. Consequently, they state that color does not help in stereo matching, which seems to contradict the evaluation studies cited above.

In this paper, we aim at shedding light onto these contradicting results. We perform an evaluation study that compares competing color as well as grey-scale energy functions against each other. An important concept of this study is to separately analyze matching errors that occur in (1) radiometric distorted and (2) radiometric clean image regions. This separation allows us to show by experiment

*Michael Bleyer received financial support from the Austrian Science Fund (FWF) under project P19797 and the Vienna Science and Technology Fund (WWTF) under project ICT08-019.

that the major argument for using color is an increased robustness against radiometric problems. In particular, we demonstrate that the overall improved performance due to using color reported in [2] can largely be explained by a significant improvement in the handling of radiometric distorted regions. This finding also establishes a link to [7], namely: If radiometric distortions are already accounted for by the match measure, the major argument for color (*i.e.* robustness against radiometric problems) seems to become useless. It is therefore not surprising that color does not improve performance when using radiometric insensitive match measures.

The remainder of this paper is organized in two parts. The first part (section 2), describes our evaluation methodology including the energy functions to be evaluated, the stereo algorithm in which the energy functions are embedded and our evaluation metrics. The second part (section 3) then presents our experiments on which the findings of the paragraph above are based on.

2. Testbed

2.1. Energy Functions

We model the stereo problem using a standard energy function. The energy measures the quality of a disparity map D that assigns pixels to discrete disparity values and is defined as

$$E(D) = E_{data}(D) + E_{smooth}(D). \quad (1)$$

Let us first discuss the smoothness term E_{smooth} whose evaluation is beyond the scope of this paper. We define it as

$$E_{smooth}(D) = \sum_{(p,p') \in \mathcal{N}} s(d_p, d_{p'}) \quad (2)$$

where \mathcal{N} is the set of all spatial neighbors in 4-connectivity and d_p denotes the disparity of p according to disparity map D . To define the function $s()$, we use a simplified truncated linear model:¹

$$s(d_p, d_{p'}) = \begin{cases} 0 & \text{if } d_p = d_{p'} \\ P_1 & \text{if } |d_p - d_{p'}| = 1 \\ P_2 & \text{otherwise.} \end{cases} \quad (3)$$

Here, P_1 and P_2 are user-defined constants where P_1 represents a penalty for small variations in disparity and P_2 penalizes disparity discontinuities. In our experiments, we set $P_1 := \frac{P_2}{2}$. We tune the value of P_2 individually for each energy function of our benchmark. The goal of this tuning is to achieve good-quality results on our test set. Note that

¹This is not a perfect choice, because the truncated linear model has a bias towards fronto-parallel surfaces. However, more advanced smoothness terms such as the second-order term of [14] are considerably more difficult to optimize. We have therefore decided against using them.

this individual tuning is required, since the application of different data terms changes the balance between data and smoothness terms.

Let us now define the data term whose evaluation is the topic of this paper. It is defined as

$$E_{data}(D) = \sum_{p \in \mathcal{I}} \rho(p, p - d_p) \quad (4)$$

where \mathcal{I} is the set of all pixels in the left view. The function $\rho(p, q)$ computes the color dissimilarity between a pixel p of the left and a pixel q of the right image. In our study, we will use different implementations of $\rho()$ and will compare their performance against each other.

2.1.1 Dissimilarity Functions

Absolute Color Difference The first dissimilarity function investigated in our study is the absolute difference of color values. This simple measure cannot handle radiometric distortions, but is important, since it represents the standard way for computing the data costs in global methods. We define ρ_{AD} as

$$\rho_{AD}(p, p - d) = \sum_{1 \leq i \leq 3} |\sigma_i(p) - \sigma_i(p - d)| \quad (5)$$

where $\sigma_i(p)$ returns the value of the i th color channel at pixel coordinates p . Note that we use nine different color representations in our study and hence nine different implementations of $\sigma()$. This is discussed in section 2.1.2.²

Window-based Measures Let us now define two window-based measures that can handle radiometric distortions. It is not common to use window-based measures in global stereo. This is most likely due to performance degradations that can be expected at disparity borders where the window contains pixels of different disparities. To weaken this effect, we apply small windows of size 5×5 pixels.

Our first measure is the Zero mean Normalized Cross-Correlation (ZNCC), which is defined as

$$\rho_{ZNCC}(p, p - d) = \frac{\sum_{q \in W_p} [\sigma_i(q) - \bar{\sigma}_i(p)][\sigma_i(q - d) - \bar{\sigma}_i(p - d)]}{\sum_{1 \leq i \leq 3} \sqrt{\sum_{q \in W_p} [\sigma_i(q) - \bar{\sigma}_i(p)]^2 \sum_{q \in W_p} [\sigma_i(q - d) - \bar{\sigma}_i(p - d)]^2}} \quad (6)$$

where W_p denotes a square window centered at pixel p . The function $\bar{\sigma}_i(p)$ returns the mean value of the i th color channel computed over all pixels inside W_p .

The second window-based measure is Census [17]. Figure 1 explains this measure by using a simple example. To

²We will also evaluate the performance of grey-scale matching where there is just one channel. In this case, the dissimilarity function is defined as $\rho_{AD}(p, p - d) = |\sigma(p) - \sigma(p - d)|$. The adaptations for the other dissimilarity functions work analogously.

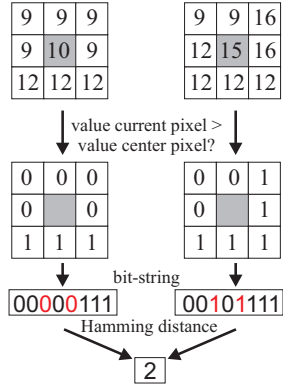


Figure 1. The Census measure. A window is centered on pixels of left and right images. The color values inside a window are converted to a binary representation where 1 means that the pixel’s color value is larger than that of the center pixel. 0 means that the opposite is true. The resulting bit-strings are compared against each other by computing the Hamming distance, which represents the dissimilarity of the two windows according to Census.

compute Census for color pixels, we first compute the Census value individually for each color channel. We then sum up the results over all three color channels.³

Mutual Information Mutual Information (MI) is attractive for global methods, since it is a radiometric insensitive measure defined on a pixel basis (and not on windows). Hence one has the advantage of being insensitive to radiometric distortions without introducing artifacts at disparity boundaries. The disadvantage is that MI requires a disparity map for computing the matching scores. This dilemma is typically solved by an iterative computation scheme, *i.e.* a disparity map is determined given initial matching scores, then the matching scores are computed using the disparity map and so on. Our implementation follows the hierarchical MI approach of [6] to speed up this procedure. The reader is referred to the same paper [6] for a more detailed description of MI. To incorporate color, we proceed as for the measures above, *i.e.* MI is computed individually for each color channel and the resulting values are summed up.

2.1.2 Color Systems

As stated above, $\sigma()$ implements nine color spaces. These are basically identical to the color systems evaluated in previous studies [2, 5].⁴ We use three different categories of color spaces: (1) primary systems (RGB and XYZ), (2) luminance-chrominance systems (LUV , LAB , AC_1C_2 ,

³Obviously, there are alternative ways for fusing the three color channels to obtain the matching costs. For example, one can compute the minimum or the median value over the three channels [5]. We have decided for the sum of channel values, since this represents the simplest and probably most commonly used method.

⁴We have excluded HSI , since its performance has been reported to be very poor in [2, 5]. Moreover, one needs to modify the dissimilarity function $\rho()$ to correctly handle HSI . This modification is not trivial.

Name	Definition
$Grey$	$I = 0.299R + 0.587G + 0.114B$
XYZ	$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
LUV	$L = \begin{cases} 116(Y/Y_w)^{\frac{1}{3}} - 16 & \text{if } Y/Y_w > 0.01 \\ 903.3 Y/Y_w & \text{otherwise} \end{cases}$ $U = 13L(u' - u'_w) \text{ with } u' = \frac{4X}{X+15Y+3Z}$ $V = 13L(v' - v'_w) \text{ with } v' = \frac{9Y}{X+15Y+3Z} X_w,$ Y_w, Z_w : white reference components
LAB	$A = 500(f(X/X_w) - f(Y/Y_w))$ $B = 200(f(Y/Y_w) - f(Z/Z_w))$ $f(x) = \begin{cases} x^{1/3} & \text{if } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{otherwise} \end{cases}$
AC_1C_2	$\begin{pmatrix} A \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
YC_1C_2	$\begin{pmatrix} Y \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{-\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
$I_1I_2I_3$	$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{-1}{2} \\ \frac{-1}{4} & \frac{-1}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$
$H_1H_2H_3$	$\begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ \frac{-1}{2} & 0 & \frac{-1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$

Table 1. Conversions from RGB to other color spaces.

YC_1C_2) and (3) statistical independent component systems ($I_1I_2I_3$, $H_1H_2H_3$). In addition, we also provide a grey-level version of $\sigma()$ ($Grey$) to implement intensity-based matching. Table 1 provides formulas for the conversion from RGB to the other color spaces.

At this point it is important to understand that we build an energy function by combining a dissimilarity function with one of our color spaces. Having four dissimilarity functions and nine color systems, this leads to $4 \cdot 9 = 36$ potential energy functions.

2.2. Energy Optimization

There is a large amount of work dealing with effective optimization of energies in the form of equation (1). The most popular optimizers are Belief Propagation and graph cut-based α -expansions. However, since computational speed is vital for handling our large number of energy functions and test sets, we have decided to incorporate our energy functions into the fast stereo matcher of [3].

Roughly spoken, the method of [3] constructs an individual tree-based approximation of the standard four-connected grid at each pixel. The global energy optimum on these trees is then computed via dynamic programming.

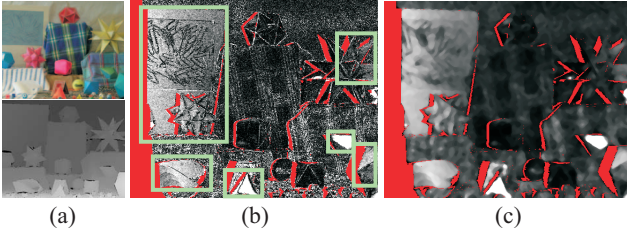


Figure 2. Extraction of radiometric distorted regions. (a) Left image of the Moebius set (top) and corresponding ground truth disparities (bottom). (b) Data costs of the ground truth solution. Bright pixels denote high costs and red pixels are occlusions. (c) Map of radiometric distorted regions. This image is generated by applying a median filter on the image of (b).

The method also incorporates a simple form of occlusion handling. It first determines a disparity map for the right input image to obtain the occluded pixels of the left image. This occlusion information is then exploited in the computation of the disparity map for the left view. From a practical point of view, the method delivers high-quality results at low processing time, which is typically less than a second.

2.3. Test Set and Quality Metrics

2.3.1 Extracting Radiometric Distorted Regions

A key concept of this paper is to separately analyze matching performance in (1) regions that are affected by radiometric distortions and in (2) regions that are not. We can extract these regions from ground truth stereo pairs.

Let us look at the data costs of the ground truth disparity solution for this purpose. For each non-occluded pixel of the left image, we look up its matching point in the right image using the ground truth disparity map. We then compute the absolute intensity difference between the two corresponding points and store the calculated value in an image. An example result of this procedure is shown in figure 2b where bright pixels represent high intensity differences.

The interesting point in figure 2b are the large regions of homogeneously high dissimilarity that we have marked with green boxes. These high-dissimilarity areas are the result of radiometric differences that exist between left and right input images (*e.g.* caused by slight variations in illumination and/or exposure times). We extract these regions by applying denoising on the ground truth cost image of figure 2b. For simplicity, we use a median filter, which generates the desired map of radiometric distorted pixels (figure 2c).

Note that apart from radiometric problems, high dissimilarities in the ground truth cost image can also be attributed to other problems such as sensor noise, sampling artefacts [1] and the stereo matting problem [4, 15]. In the ground truth cost image, noise typically leads to isolated high-cost pixels, while sampling and matting artefacts lead to thin high-cost lines in the proximity of texture and disparity bor-

ders, respectively (also see figure 2b). We do not investigate these problems in our study and have therefore treated these high-cost pixels as noise. In fact, we believe that the smoothness term of our energy can successfully suppress these high-dissimilarity pixels to a large extent due to their number being small. This is in contrast to radiometric distorted regions, where large areas of high-cost pixels lead the energy optimum away from the correct disparity.

2.3.2 Test Set

To accomplish our experiments, we select a large number of 30 ground truth stereo pairs [7, 11] that can be obtained from the Middlebury website. For each pair, we compute a map of radiometric distorted pixels using our procedure described in section 2.3.1. The left images of the stereo pairs and corresponding distortion maps are shown in figure 3. It is interesting to note that although the images from the Middlebury set have been acquired under laboratory conditions and using controlled illumination, they contain a considerable amount of radiometric distortion.⁵

2.3.3 Quality Metrics

To evaluate the quality of a disparity map, we compare it against the ground truth image. Our first error measure $E_{\mathcal{V}}$ follows [11] and computes the percentage of visible (unoccluded) pixels having a disparity error larger than one pixel.

Our other two error measures operate on the maps of figure 3. The first measure $E_{\mathcal{D}}$ computes the percentage of wrong pixels in radiometric distorted regions and is defined as

$$E_{\mathcal{D}}(D) = \frac{\sum_{p \in \mathcal{V}} T[|d_p - d'_p| > 1] \cdot w_p}{\sum_{p \in \mathcal{V}} w_p}. \quad (7)$$

Here, \mathcal{V} is the set of all non-occluded pixels. T is the indicator function that returns 1 if its argument is true and 0, otherwise. d'_p returns the disparity of p according to the ground truth solution. Finally, w_p returns a weight that lies between 0 and 1. This weight represents the confidence to which we believe that p is part of a radiometric distorted region and is directly inferred from the distortion maps of figure 3. The weight is thereby 0 if p is black in the distortion map and 1 if p is white. For grey-scale values different from black or white, the weight takes a fractional value. Note that our distortion maps are not binary (*i.e.* distorted / not distorted), since we want to avoid the problem of finding an appropriate threshold for binarization.

Our final error measure $E_{\mathcal{C}}$ calculates the percentage of

⁵For each of the new Middlebury pairs, there exist various versions, which differ in that they have been recorded using varying illumination conditions and exposure times. In this study, we only use image pairs for which left and right images have been acquired using constant illumination conditions and exposure times. It is not our goal to study extremely strong radiometric distortions as has, for example, been a topic in [7].

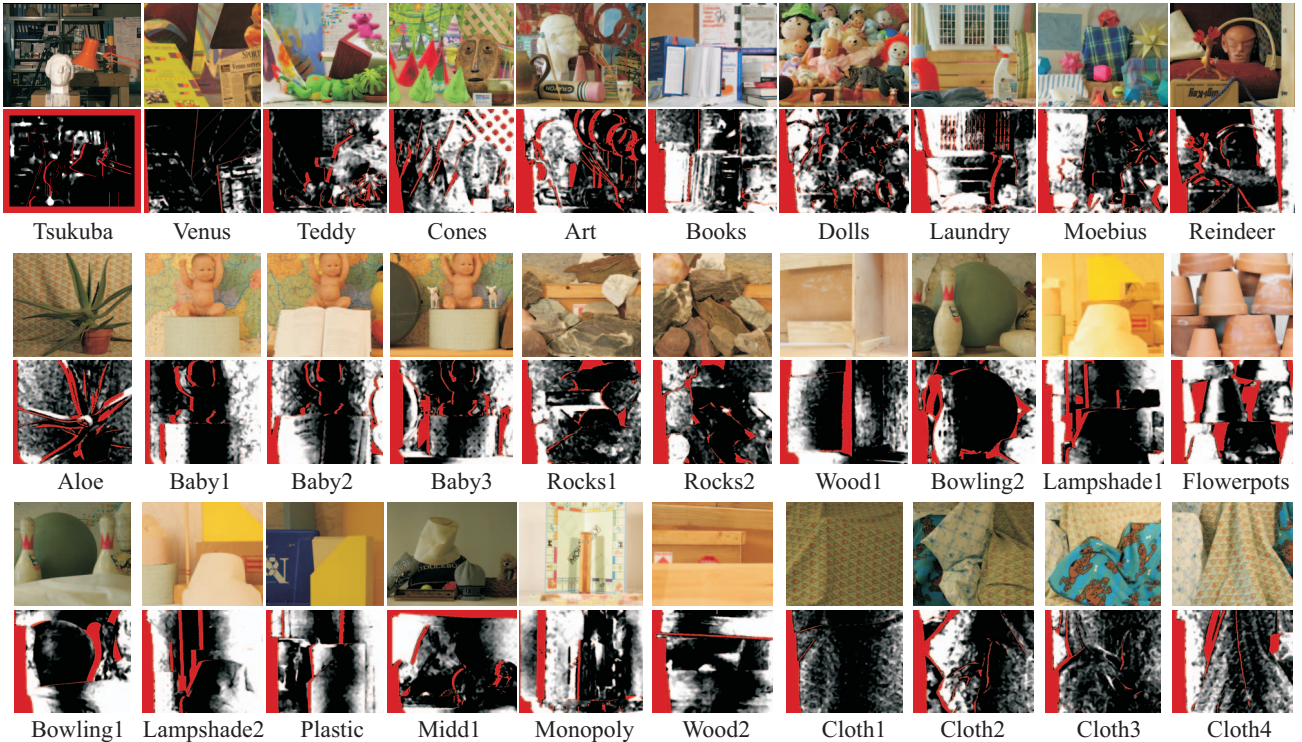


Figure 3. The 30 test sets used in our test runs and computed maps representing radiometric distortions. High intensity values represent pixels of high radiometric distortion. Red pixels denote occlusions.

wrong pixels in radiometric “clean” regions. It is defined as

$$E_C(D) = \frac{\sum_{p \in \mathcal{V}} T[|d_p - d'_p| > 1] \cdot (1 - w_p)}{\sum_{p \in \mathcal{V}} (1 - w_p)}. \quad (8)$$

For later use in section 3, we also define \bar{E}_V , \bar{E}_D and \bar{E}_C as the average of these error measures computed over all 30 test stereo pairs.

3. Experiments

3.1. Standard Color-Based Matching

We start by using absolute color differences (see equation (5)) as a dissimilarity function and compare the performance of our nine color systems. This experiment is equivalent to the one performed in [2], but [2] has not given an explanation for the results, as we will do in the following.

In our experiments, we have recognized that when comparing disparity maps obtained by using different color systems, disparity errors are not randomly distributed all over the images, but occur in approximately the same image regions. For example, let us consider the Moebius image pair. Figure 4 shows error maps that we have produced using grey-scale matching and using color-based matching with *RGB* and *LUV* color systems. In the error maps, we plot pixels that have a disparity error larger than one pixel in comparison to the ground truth image. We use the same

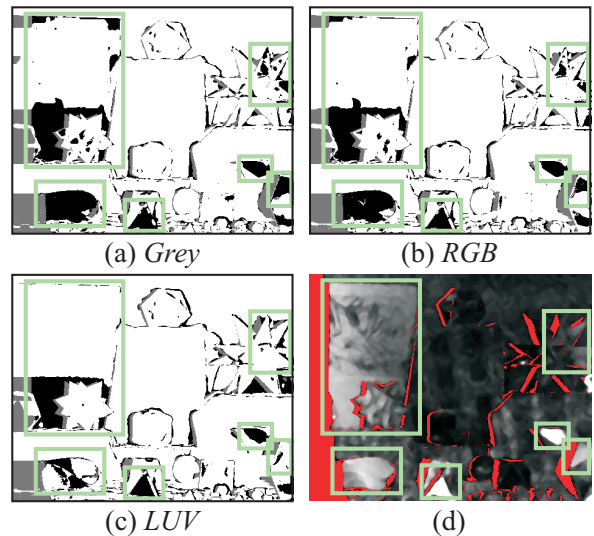


Figure 4. Disparity errors in the Moebius test set using (a) intensity-based, (b) *RGB*-based and (c) *LUV*-based matching. We use absolute differences in color values as dissimilarity function. Disparity errors (black and grey pixels) always occur in approximately the same image areas. These areas seem to correspond to regions of high radiometric distortions (bright pixels in (d)). In comparison to grey-scale matching, *RGB* and *LUV* seem to improve matching performance in exactly those regions.

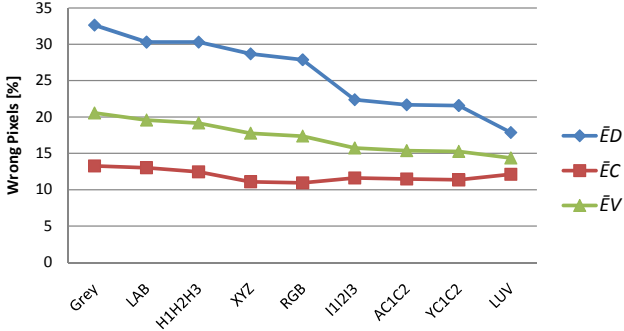


Figure 5. Our quality measures computed for nine different color spaces using absolute color differences as a dissimilarity function. Analogously to [2], we report that color matching outperforms intensity-based matching (see values of \overline{E}_V). The important point stressed in this paper is that this “overall” improvement is mainly due to a large improvement in radiometric distorted areas (see values of \overline{E}_D). In radiometric clean areas, the benefit of color seems to be low (see values of \overline{E}_C).

error coding as on the Middlebury webpage [11], *i.e.* black pixels show errors in visible regions, while grey pixels show errors in occluded image areas. It can be seen from the error maps that the majority of disparity errors occur in approximately the same image regions for all three color representations (green rectangles in figure 4). These areas seem to have large overlap with areas affected by radiometric distortions (see figure 4d). The important point is that color (especially *LUV*) effectively reduces the errors in exactly those areas. This motivates the conclusion that color is of specific importance in regions of radiometric distortions.

Obviously, it is not valid to draw this conclusion from a single image pair. We have therefore computed our quality measures \overline{E}_V , \overline{E}_D and \overline{E}_C that are defined over a large image set of 30 stereo pairs. Figure 5 plots the corresponding results for our nine color representations. Here, the values of \overline{E}_V confirm the result of [2], *i.e.* color outperforms grey-scale matching and the luminance-chrominance color systems are the best performing color representations.

Figure 5 shows only small variations in the values of \overline{E}_C , which measures the error in radiometric “clean” regions. Therefore, it seems that color does not have significant influence if there are no radiometric distortions. Consequently, this cannot represent an explanation for the “overall” improvement measured by \overline{E}_V . However, the important point is that color leads to a large performance gain in radiometric distorted regions (see values of \overline{E}_D). For example, the value \overline{E}_D drops from 32,6% for grey-scale matching to 17,8% when the *LUV* color system is used, *i.e.* *LUV* eliminates almost half of the errors in distorted regions. Therefore, a key observation of this study is that an “overall” improvement by using color is largely caused by considerably improved performance in radiometric distorted regions.

We have identified radiometric distorted pixels as being

the major source of matching errors in our test set. In the case of grey-scale matching, 57% of the “overall” wrong matches stem from these distorted pixels. (This is in contrast to their relatively small number, *i.e.* 33% of the “overall” pixels are affected by radiometric problems.) Therefore, improving the value of \overline{E}_D has strong influence on the overall matching performance \overline{E}_V . Note that the amount of radiometric problems might obviously be different on a different set of test images. However, one has to consider that the Middlebury images have been captured under controlled illumination and exposure conditions in laboratory conditions. Hence, we strongly believe that the influence of radiometric distortions will even be higher on “more practical” image pairs.

The conclusion that color is more robust in radiometric distorted regions is not surprising. For example, consider two images that have been recorded under different illumination conditions. In this case, it is likely that a green pixel of the left image corresponds to a dark green pixel of the right view, *i.e.* that the pixel changes its intensity. In contrast to this, we believe that it is quite unlikely that these illumination differences lead to the green pixel turning into a red one, *i.e.* that the pixel changes its color. Despite this robustness against radiometric distortions, it is not a good idea to solely rely on color information in the matching process as is demonstrated next.

Let us now take a closer look at the *LUV*-color system, which is the top performer according to figure 5. *LUV* strictly separates intensity from color, *i.e.* the *L*-channel contains the intensity information, while the *U*- and *V*-channels hold the color information. Let us use this property of *LUV* to compare “pure” intensity against “pure” color matching. We therefore construct a first energy function that only operates on the *L*-channel and a second one that solely operates on the *U*-channel.⁶ We use the absolute difference of channel values as a dissimilarity function. The values of E_D and E_C are plotted for each test pair in figure 6. As expected, color is clearly superior to intensity in regions of radiometric distortions (see left plot in figure 6). When looking at radiometric clean areas (right plot), the opposite is true. Here, intensity outperforms color matching. The problem of “pure” color matching is that one loses a lot of texture when discarding intensity information from the image. Since stereo matching relies on the presence of texture, this leads to performance degradations. To handle both cases, *i.e.* radiometric distorted and undistorted pixels, it is a good idea to combine intensity and color information.

We believe that the performance differences between color systems in figure 5 can be explained by the “recipe” for mixing color and intensity. The color systems *H1H2H3*, *XYZ*, *RGB* and *I1I2I3* interleave color with intensity in

⁶We omit the results for the *V*-channel for the sake of legibility in figure 6. These results are almost identical to that of the *U*-channel.

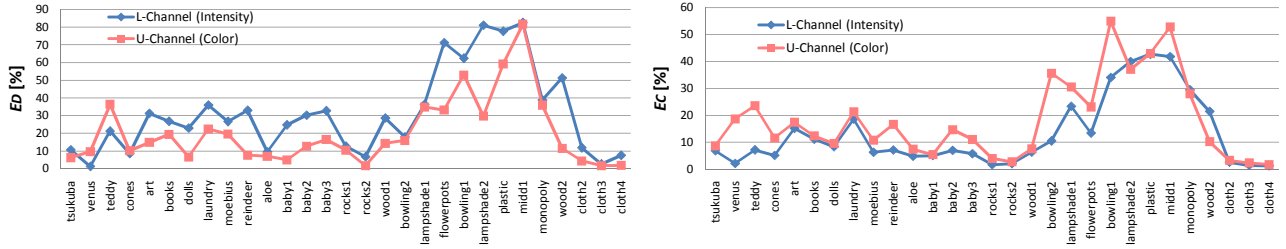


Figure 6. “Pure” intensity-based versus “pure” color-based matching. Our first energy function computes data costs as the absolute differences in L -channel values of the LUV color system. The L -channel holds the intensity information. The second energy function operates on the U -channel that holds color information only. The left diagram shows that pure color-based matching outperforms intensity-based matching in radiometric distorted regions. If there are no radiometric distortions, intensity-based matching is clearly superior (right plot).

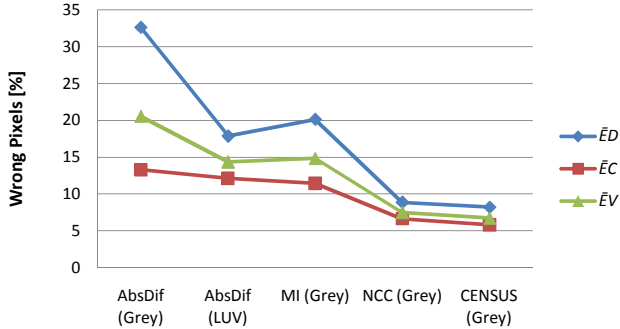


Figure 7. Standard color matching versus radiometric insensitive match measures. The radiometric insensitive measures ZNCC and Census (used on grey-scale images) outperform other strategies.

each channel, *i.e.* the mixture ratio is 1:1. In contrast to this, the better-performing luminance-chrominance systems AC_1C_2 , YC_1C_2 and LUV use one dedicated channel for intensity, but two dedicated channels for color, *i.e.* the ratio is 1:2.⁷ Hence, we can say that the luminance-chrominance systems contain “more color”, and this is important to handle the radiometric distorted regions that represent the major source of matching errors on our test sets.

3.2. Radiometric Insensitive Measures

We have argued that the main benefit of color is an improved robustness against radiometric problems. However, as an alternative, one can directly use radiometric insensitive measures on a grey-level basis to obtain this robustness. We now compare these two options against each other.

We have the following competing energy functions in our experiment: (1) We use the absolute difference of grey values as a matching function. (2) As a representative of standard color-based matching, we determine the absolute differences of LUV color values (which has been the best-performing strategy in figure 5). (3) For evaluating the performance of radiometric insensitive measures, we use MI, ZNCC and Census on grey-scale images to calculate the

⁷We currently do not have a good explanation for the bad performance of the luminance-chrominance system LAB .

data term. Our error measures representing the matching performance of these strategies are plotted in figure 7.

The results speak a clear language. ZNCC and Census have the same effect as color matching, *i.e.* they reduce the error in radiometric distorted regions. However, ZNCC and Census are considerably more effective in this respect. In addition, both measures are even capable of reducing the errors in radiometric clean regions, which leads to the overall best performance. Note that the performance improvements are extreme. For example, the value of \bar{E}_V drops from 20.5% for *AbsDif (Grey)* to 6.7% for *Census (Grey)*. This demonstrates the potential that lies in the data term of common energy functions. MI delivers approximately the same performance as standard color-based matching, which is surprisingly poor in comparison to ZNCC and Census. Nevertheless, from these results, it is clear that radiometric insensitive measures should be preferred over standard dissimilarity functions such as the absolute difference. We strongly suggest to follow this advice also in cases where radiometric differences do not seem to be present to the human observer. For example, radiometric problems are hardly visible on our test set with bare eyes.

Let us now see if we can get additional performance improvement for ZNCC and Census when incorporating color information. In figure 8, we build the energy functions by combining our nine color spaces with ZNCC and Census. The results show that color does not help and even worsens the results. We believe that this does not represent a contradiction to figure 5 where a clear improvement due to color has been observed. The main benefit of color (*i.e.* robustness against radiometric distortions) becomes unimportant, since such distortions are already handled by the dissimilarity function. A similar result has been reported in [7] where the authors observed only very little improvement due to color. The authors of [7] have argued that color is not important, since by deleting it from the images, one practically does not lose texture. Moreover, they state that intensity is, in general, more robustly captured by the camera than color information. We believe that these two arguments become dominant when the major argument for using color, *i.e.* handling of radiometric distorted regions, falls away.

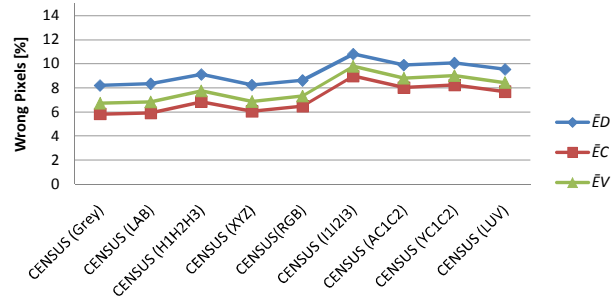
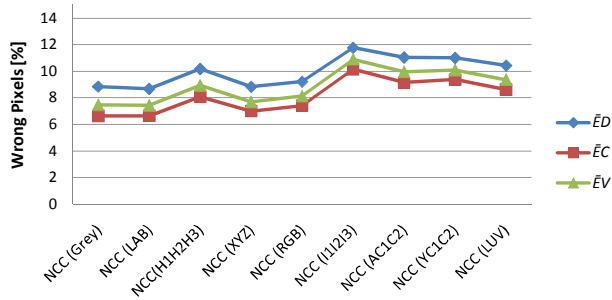


Figure 8. ZNCC and Census measures used with intensity and color. Color does not represent an improvement for both measures.

4. Conclusions

This paper has investigated color stereo matching with a specific focus on radiometric distortions. We have claimed that the major argument for color is an improved treatment of radiometric distorted regions. However, we believe that this benefit is low considering that the alternative of directly using radiometric insensitive match measures on grey-scale images works considerably better. We answer the question whether color should currently be used in stereo matching by no: In our experiments with radiometric insensitive measures, color has consistently led to performance degradation. In accordance with [7], we believe that in contrast to intensity information, color is simply less robustly captured by nowadays cameras (or at least by those cameras used to record the applied test images).

In future work, there are several ways to extend our study. One obvious extension is the application of additional optimization algorithms for minimizing our energy functions (e.g. α -expansions or Belief Propagation). One could also extend the set of evaluated energy functions, e.g. by incorporating more advanced smoothness terms [14] or dissimilarity functions [16]. However, we believe that none of these modifications would lead to substantially different results. However, a more crucial point is that due to the absence of ground truth data different from the Middlebury set, our findings are based on images that have all been obtained under similar conditions using the same cameras. In future work, one should generate additional ground truth images using a set of different cameras to check the validity of our conclusions on other data. An interesting aspect also lies in the upcome of novel “segmentation-based” cost aggregation strategies that have been evaluated in [12, 13]. Fusing our results with [12, 13] would represent another step on the way to an “optimal” data term for stereo.

References

- [1] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *TPAMI*, 20(4):401–406, 1998.
- [2] M. Bleyer, S. Chambon, U. Poppe, and M. Gelautz. Evaluation of different methods for using colour information in global stereo matching. In *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XXXVII, pages 415–422, 2008.
- [3] M. Bleyer and M. Gelautz. Simple but effective tree structures for dynamic programming-based stereo matching. In *VISAPP*, volume 2, pages 415–422, 2008.
- [4] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann. A stereo approach that handles the matting problem via image warping. In *CVPR*, pages 501–508, 2009.
- [5] S. Chambon and A. Crouzil. Colour correlation-based matching. *Int J Robot Autom*, 20(2):78–85, 2005.
- [6] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 30(2):328–341, 2008.
- [7] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *PAMI*, 31:1582–1599, 2009.
- [8] A. Koschan. Dense stereo correspondence using polychromatic block matching. In *CAIP*, volume 719 of LNCS, pages 538–542, 1993.
- [9] K. Mühlmann, D. Maier, J. Hesser, and R. Männer. Calculating dense disparity maps from color stereo images, an efficient implementation. *IJCV*, 47(1):79–88, 2002.
- [10] M. Okutomi and G. Tomita. Color stereo matching and its application to 3-d measurement of optic nerve head. In *ICPR*, volume 1, pages 509–513, 1992.
- [11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1/2/3):7–42, 2002.
- [12] F. Tombari, S. Mattoccia, L. D. Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. In *CVPR*, 2008.
- [13] L. Wang, M. Gong, M. Gong, and R. Yang. How far can we go with local optimization in real-time stereo matching. In *3DPVT*, pages 129–136, 2006.
- [14] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. In *CVPR*, pages 1–8, 2008.
- [15] W. Xiong and J. Jia. Stereo matching on objects with fractional boundary. In *CVPR*, pages 1–8, 2007.
- [16] K. Yoon and I. Kweon. Stereo matching with the distinctive similarity measure. In *ICCV*, 2007.
- [17] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. In *ECCV*, pages 151–158, 1994.