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@lcpc.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 information. 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

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

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

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

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]

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

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 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

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

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

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

where $\sigma_i(p)$ returns the value of the ith color channel at pixel coordinates p. Note that wemdel:

[1] This 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.

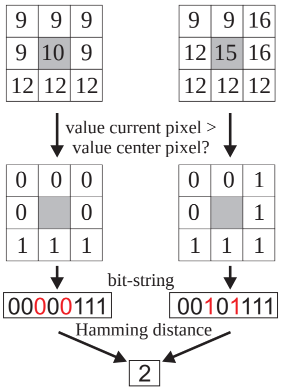


Figure 1. The Census measure. A window is centered on pixels of left and right images. The color values inside a window are converted to a binary representation where 1 means that the pixel's color value is larger than that of the center pixel. 0 means that the opposite is true. The resulting bit-strings 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,

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	$L = \begin{cases} 116\,(Y/Y_w)^{\frac{1}{3}} - 16 & \text{if } Y/Y_w > 0.01 \\ 903.3\,Y/Y_w & \text{otherwise} \end{cases}$ $U = 13i\ f$

[3] Obviously, 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.

[4] We 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.

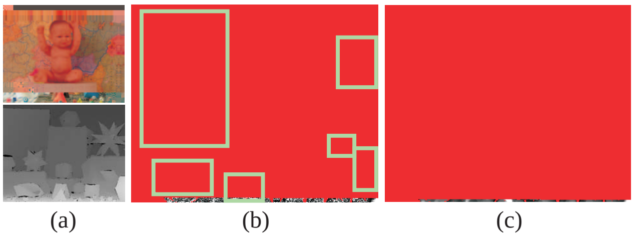


Figure 2. Extraction of radiometric distorted regions. (a) Left image of the Moebius set (top) and corresponding ground truth disparities (bottom). (b) Data costs of the ground truth solution. Bright pixels denote high costs and red pixels are occlusions. (c) Map of radiometric distorted regions. This 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 borders, 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 extend 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.

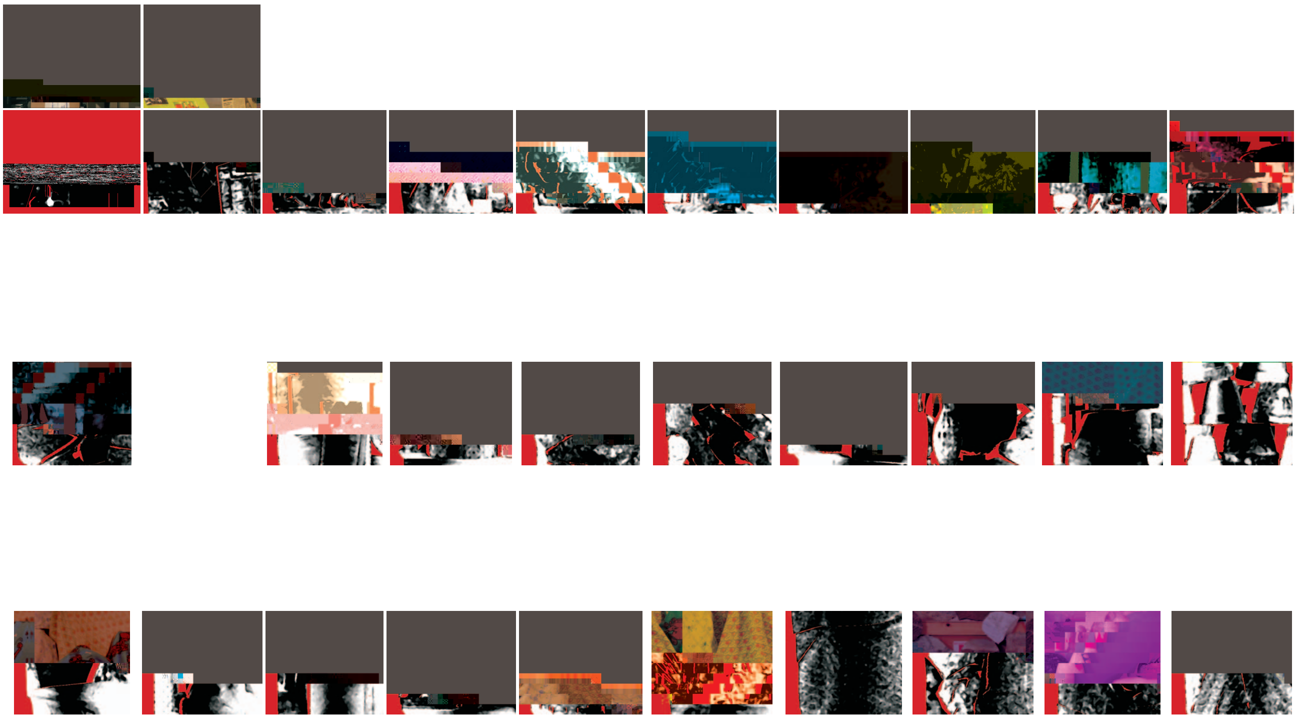
Our other two error measures operate on the maps of figure 3. The first measure $E_{\mathcal{D}}$ computes the percentage of wrong pixels in radiometric distorted regions and is defined as

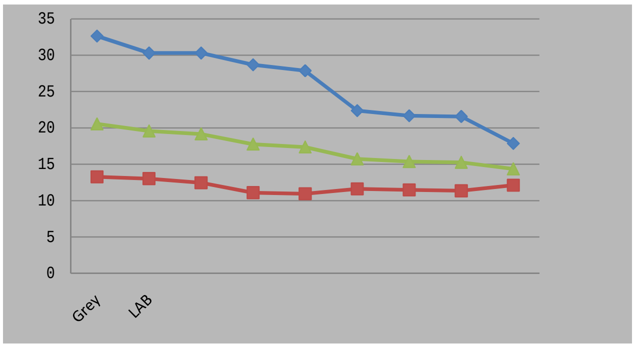
$$E_{\mathcal{D}}(D) = \frac{\sum_{p \in \mathcal{V}} T\left[|d_p - d'_p| > 1\right] \cdot w_p}{\sum_{p \in \mathcal{V}} w_p}. \tag{7}$$

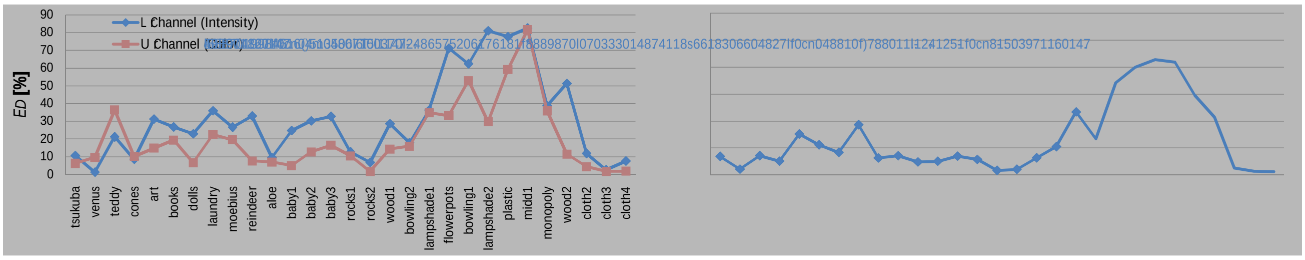
Here, $\mathcal{V}$ is the set of all non-occluded pixels. T is the indicator function that returns 1 if its argument is true and 0, otherwise. d'_p returns the disparity of p according to the ground truth solution. Finally, w_p returns a weight that lies between 0 and 1. This weight represents the confidence 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.

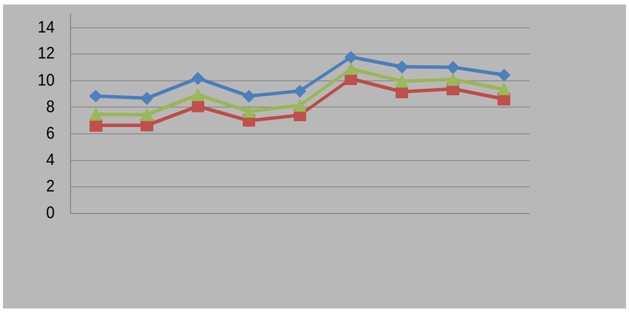
Our final error measure $E_{\mathcal{C}}$ calculates the percentage of

[5] For 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].




35
30
25
20
15
10
5
0
Grey
LAB


L Channel (Intensity)
U Channel (Color)
ED [%]
90
80
70
60
50
40
30
20
10
0
tsukuba
venus
teddy
cones
art
books
dolls
laundry
moebius
reindeer
aloe
baby1
baby2
baby3
rocks1
rocks2
wood1
bowling2
lampshade1
flowerpots
bowling1
lampshade2
plastic
midd1
monopoly
wood2
cloth2
cloth3
cloth4


14
12
10
8
6
4
2
0