

# Aggregation functions to combine *RGB* color channels in stereo matching

Mikel Galar,\* Aranzazu Jurio, Carlos Lopez-Molina, Daniel Paternain,  
Jose Sanz and Humberto Bustince

*Departamento de Automatica y Computacion, Universidad Publica de Navarra,  
Campus Arrosadia sn, 31006 Pamplona, Spain*

[\\*mikel.galar@unavarra.es](mailto:mikel.galar@unavarra.es)

**Abstract:** In this paper we present a comparison study between different aggregation functions for the combination of *RGB* color channels in stereo matching problem. We introduce color information from images to the stereo matching algorithm by aggregating the similarities of the *RGB* channels which are calculated independently. We compare the accuracy of different stereo matching algorithms and aggregation functions. We show experimentally that the best function depends on the stereo matching algorithm considered, but the dual of the geometric mean excels as the most robust aggregation.

© 2012 Optical Society of America

**OCIS codes:** 100.0100 Image processing, 150.0150 Machine vision, 330.0330 Vision, color, and visual optics.

---

## References and links

1. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *Int. J. Comput. Vision* **47**, 7–42 (2002).
2. B. Cyganek and J.P. Siebert, *An Introduction to 3D computer vision techniques and algorithms* (Wiley, 2009).
3. R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence", presented at the Third European Conference on Computer Vision, Stockholm, Sweden, 2-6 May 1994.
4. E. Trucco, V. Roberto, S. Tinonin and M. Corbato, "SSD disparity estimation for dynamic stereo", presented at The British Machine Vision Conference, Edinburgh, England, 1996.
5. O. Faugeras, T. Vieville, E. Theron, J. Vuillemin, B. Hotz, Z. Zhang, L. Moll, P. Bertin, H. Mathieu, P. Fua and G. Berry, C. Proy, "Real-time correlation-based stereo: algorithm, implementations and applications", INRIA Technical Report **2013**, 1993.
6. X. Hu and P. Mordohai, "Quantitative evaluation of confidence measures for stereo vision", *IEEE T. Pattern Anal.* **34**, 2121–2133 (2012).
7. X. Xiang, M. Zhang, G. Li, H. Yuyong and Z. Pan, "Real-time stereo matching based on fast belief propagation", *Mach. Vision Appl.* **23**, 1219–1227 (2012).
8. M. Bleyer and M. Gelautz, "Graph-based surface reconstruction from stereo pairs using image segmentation", *Proc. SPIE* **5665**, 288–299 (2005).
9. L. Hong and G. Chen, "Segment-based stereo matching using graph cuts", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2004), pp. 74–81.
10. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts", *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2001), pp. 508–515.
11. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2004), pp. 261–268.
12. G. Tolt and I. Kalaykov, "Measures based on fuzzy similarity for stereo matching of colour images", *Soft Comput.* **10**, 1117–1126 (2006).
13. A. Klaus, M. Sormann and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", *Proceedings of IEEE International Conference on Pattern Recognition* (IEEE, 2006), pp. 15–18.

14. M. Galar, J. Fernandez, G. Beliakov and H. Bustince, "Interval-valued fuzzy sets applied to stereo matching of color images", *IEEE T. Image Process.* **20**, 1949–1961 (2011).
  15. D. Van der Weken, M. Nachttegael, E. E. Kerre, "Using similarity measures and homogeneity for the comparison of images", *Image Vision Comput.* **22**, 695–702 (2004).
  16. W.J. Wang, "New similarity measure on fuzzy sets and on elements", *Fuzzy Set. Syst.* **85**, 305–309 (1997).
  17. S.M. Chen, M.S. Yeh and P.Y. Hsiao, "A comparison of similarity measures of fuzzy values", *Fuzzy Set. Syst.* **72**, 78–89 (1995).
  18. S. Kullback, *Information theory and statistics* (Wiley, 1959).
  19. C.P. Pappis and N.I. Karacipilidis, "A comparative assessment of measures of similarity of fuzzy values", *Fuzzy Set. Syst.* **56**, 171-174 (1993).
  20. E. Deza and M.M. Deza, "Image and audio distances", in *Dictionary of distances*, E. Deza and M.M. Deza (Elsevier, 2006), pp. 262–278.
- 

## 1. Introduction

The stereo matching problem consists in obtaining the three-dimensional information from two bi-dimensional images of the same scene taken from different viewpoints. When an image is taken, the depth of each point in the scene is lost. Therefore, the objective of a stereo matching algorithm is to retrieve this information.

The basis of stereo vision is that a single physical point in the scene is uniquely projected to a pair of image locations. Hence, to obtain the depth from both images, first we have to estimate the correspondence between the pixels in each image. This step consists in identifying the same physical point in both projections to determine the difference between the position in each image. This difference is called *disparity*. The disparity, together with the parameters of the camera allows us to obtain the depth.

Thereby, the main problem of the stereo matching is the difficulty to find the correspondence correctly. The images are taken from different cameras with different viewing angles. These facts sometimes produce occlusions, perspective distortion, different lighting intensities, reflections, shadows, repetitive patterns, sensory noise, etc. All these facts convert a simple correspondence task in a very difficult one.

For this reason, an ideal configuration of the cameras is usually supposed. That is, they are only horizontally displaced and the focus lines are parallel. This ideal configuration and epipolar geometry allow to restrict the search of one point in the first image to the epipolar line in the second image, reducing the search space and hence, greatly decreasing the computational cost.

An exhaustive overview on stereo matching can be found in [1], while a complete introduction to stereo vision can be found in [2]. Stereo matching algorithms can be classified into local and global methods. The local approaches compare the intensity levels of a finite window to determine the disparity for each pixel. These methods use different metrics or similarities to compare intensity levels such as SAD [3], SSD [4] or NCC [5], which are widely applied despite its simplicity due to their low computational complexity [6, 7]. Global approaches apply some global assumptions about smoothness and try to determine all the disparities at the same time by using different optimization techniques such as, graph cuts [8, 9, 10], belief propagation [11], etc. These methods usually start from a local disparity estimation.

The utilization of color information, specifically the use of *RGB* color space, improves the stereo matching results achieved with gray scale images [12, 13, 14]. The extra information provided by color channels removes some ambiguities produced when the information is reduced to gray scale. Therefore, matching can be improved avoiding false correspondence matches, but we will show that this improvement is directly related with the aggregation function considered to add color similarities. Having more information might lead to worse results if it is not handled properly.

In general, there exist several techniques to calculate the disparity map using color informa-

tion, but there is no agreement about which is the best color space to work with. In this work *RGB* representation is used following previous approaches [12, 13].

Local search matching algorithms assign to every pixel of the right image a correspondence with another pixel of the left image. Using a local search algorithm with color images we will obtain for each pixel in the right image three correspondences (one per channel) in the left image. Usually these degrees will not match, so we will have different matching scores. A simple and efficient solution is to aggregate the similarity information of channels and then choose as corresponding pixel the one having the largest aggregated similarity value.

In this work we study the performance and influence of different aggregation operators, such as the arithmetic mean, the median, the minimum, etc. To do so, we use several test images from [1] which have been taken using the ideal configuration of the cameras. Our aim is to study their different behaviors among different measures used in stereo matching and image comparison [14, 6]. We empirically show that using the proper aggregation functions can produce significant better results, whereas using inappropriate ones could decrease the performance. Our objective is to find an aggregation for color similarities that works well whichever method (similarity measure) is used. In this sense, we want to study which aggregation is more robust. While some aggregations excels in some methods or images, our interest reside in finding an appropriate set of aggregations that can be safely used among different metrics and images.

This work is organized as follows: In Section 2 we remain the classical stereo matching algorithm and we present the metrics that we have considered. In Section 3 we present the aggregation operators that we are going to compare in Section 4, where the experimental study is carried out. Section 5 concludes this work.

## 2. Stereo matching for color images

In this section, we recall the typical steps of the classical stereo matching algorithm. Afterward, we present the different metrics and similarity measures used in the comparison.

### 2.1. Stereo matching algorithm

Minor changes have to be applied to transform the original stereo matching algorithm for gray scale images to color ones. In the first case, the algorithm computes the similarity of the window surrounding each pixel in the right image with several windows in the left image (considering the epipolar constraint and the maximum disparity). The pixel which surrounding window reaches the largest similarity degree is chosen and used to compute the disparity.

Regarding color images, we simply compute the correspondence between color channels independently, and then we aggregate these correspondence scores (similarity degrees). We can summarize the algorithm for color images as follows (in Fig. 1 we depict an overall view of the method):

```
Algorithm Stereo Matching
  const
    Window size := n x m
  begin
    For each pixel right image
      For each pixel in the epipolar line left image
        For each color channel
          Calculate the similarity between the window centered
            at the pixel of the right image and the window
              centered at the pixel of the left image
        end For
        Aggregate similarities
      end For
      Set correspondence:= arg max{value of aggregation of similarities}
      Disparity:= difference between the x-position of two pixels
    end For
    Create a disparity map from all the disparities obtained
```

end.

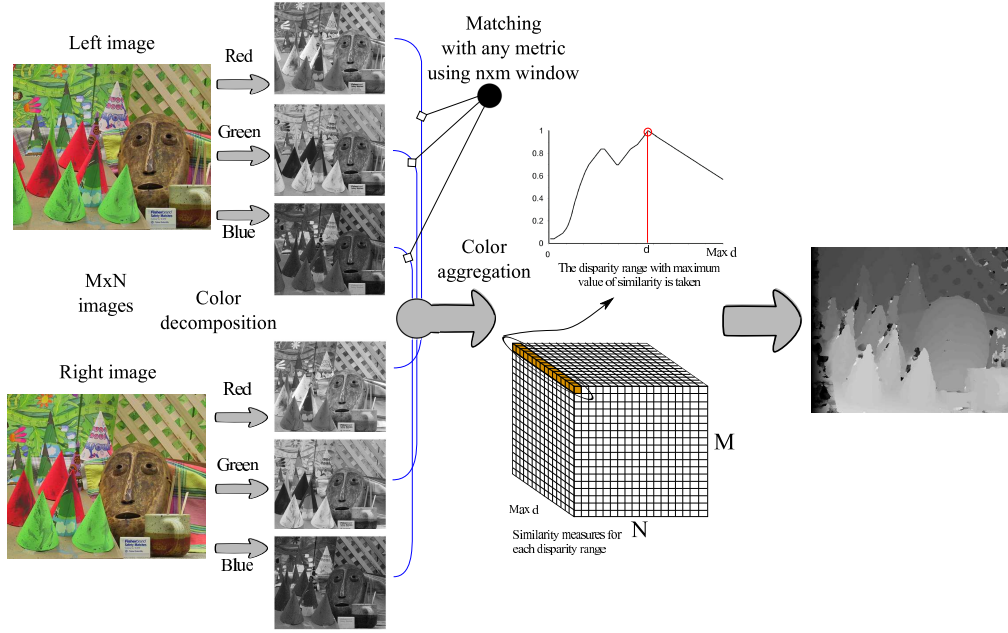


Fig. 1. Stereo matching algorithm scheme using color information from *RGB* channels.

There are many different versions of the classical stereo matching algorithm. However, most of them use the scheme we have presented. The metric or the similarity measure used is usually the biggest difference between algorithms. Another key-factor of the algorithm is the aggregation function used to aggregate the color similarities. In the following subsection we present seven common similarity measures for stereo matching problem. Then, in Section 3 we present the aggregation functions considered for the empirical study.

## 2.2. Correspondence and similarity measures between windows

There exist several methods to compute the similarity between windows. The results (given by the obtained disparity maps) directly depends on these measures. In this paper, we study several metrics to show the behavior of different aggregations within each method.

### 2.2.1. Sum of Square Differences (SSD)

SSD [4] computes the matching score as the sum of the square differences between all pixels intensities from left window with respect to right window. Then, the disparity is computed with the one with the lowest value (largest correspondence), which indicates the most similar window. SSD can be expressed as follows:

$$SSD(I_r(x, y), I_l(x + k, y)) = \sum_{m, n \in W} (I_r(x + m, y + n) - I_l(x + m + k, y + n))^2 \quad (1)$$

being  $x, y$  the position of the pixel,  $k$  the displacement of the left window respect to the right window,  $W$  the window (size  $n \times m$ ) considered and  $I_r, I_l$  right and left images respectively.

### 2.2.2. Sum of Absolute Differences (SAD)

SAD [3], computes the disparity in the same way as SSD, but using the absolute differences between pixel intensities instead of the square differences:

$$SAD(I_r(x, y), I_l(x + k, y)) = \sum_{m, n \in W} |I_r(x + m, y + n) - I_l(x + m + k, y + n)| \quad (2)$$

### 2.2.3. Normalized Cross-Correlation (NCC)

NCC [5] is expressed by the following formula:

$$NCC(I_r(x, y), I_l(x + k, y)) = \frac{\sum_{m, n \in W} (I_r(x + m, y + n) \cdot I_l(x + m + k, y + n))}{\sqrt{\sum_{m, n \in W} (I_r(x + m, y + n))^2 \cdot \sum_{m, n \in W} (I_l(x + m + k, y + n))^2}} \quad (3)$$

The disparity is obtained from the  $k$  reaching the maximum value.

### 2.2.4. Fuzzy similarity ( $SM_{FS}$ )

The fuzzy similarity introduce the fuzzy set theory to compute the correspondence between two windows. It is computed with the following expression [12]:

$$SM_{FS}(I_r(x, y), I_l(x + k, y)) = \frac{1}{m \times n} \sum_{m, n \in W} s(I_r(x + m, y + n), I_l(x + m + k, y + n)) \quad (4)$$

$$\text{where } s(a, b) = \begin{cases} 1 - \frac{|a - b|}{\alpha}, & \text{if } |x - y| < \alpha \\ 0, & \text{otherwise,} \end{cases}$$

the parameter  $\alpha = 16$  is used generally [12]. Disparity is computed with the pixel which similarity measure attains its maximum value.

### 2.2.5. Distance-based similarities ( $SM_M$ and $SM_K$ )

Distance-based similarities are widely used in image processing for image comparison techniques [15]. Hence, they are appropriate to compare the correspondence between windows. The smaller the distance is, the greater similarity is obtained. In our experiments we consider two different cases. The first one is based on Minkowski distance  $d_r$  with  $r = 1$ , that is equivalent to Manhattan distance, but in this case the measure is normalized by the sum of the intensities within windows (note that different w.r.t. Eq. (2)). We denote this measure as  $SM_M$  [16, 17]:

$$SM_M(I_r(x, y), I_l(x + k, y)) = 1 - \frac{\sum_{m, n \in W} |I_r(x + m, y + n) - I_l(x + m + k, y + n)|}{\sum_{m, n \in W} (I_r(x + m, y + n) + I_l(x + m + k, y + n))} \quad (5)$$

The second one is based on the Kullback distance [18] between fuzzy sets. We denote this similarity as  $SM_K$ :

$$\begin{aligned}
SM_K(I_r(x, y), I_l(x + k, y)) &= 1 - \frac{1}{MN2 \ln 2} \cdot \\
&\sum_{m, n \in W} \left[ (I_r(x + m, y + n) - I_l(x + m + k, y + n)) \ln \left( \frac{1 + I_r(x + m, y + n)}{1 + I_l(x + m + k, y + n)} \right) + \right. \\
&\left. (I_l(x + m + k, y + n) - I_r(x + m, y + n)) \ln \left( \frac{2 - I_r(x + m, y + n)}{2 - I_l(x + m + k, y + n)} \right) \right] \quad (6)
\end{aligned}$$

In both cases, we normalize the intensities of the pixels to the unit interval in such way that we can apply these similarities and then compute the disparity from the largest one.

#### 2.2.6. Similarity Measure based on Union and Intersection ( $SM_{UI}$ )

The concept of similarity from union and intersection operations also comes from fuzzy set theory [19], where the similarity between two fuzzy sets can be computed as the division of intersection's cardinality and the union's cardinality. In the same way as in distance-based methods, we normalize the intensities before applying this similarity. Then, the disparity is obtained from the largest output. The expression is as follows:

$$SM_{UI}(I_r(x, y), I_l(x + k, y)) = \frac{\sum_{m, n \in W} \min(I_r(x + m, y + n), I_l(x + m + k, y + n))}{\sum_{m, n \in W} \max(I_r(x + m, y + n), I_l(x + m + k, y + n))} \quad (7)$$

### 3. Aggregation functions

In our experiments we use *RGB* representation following previous works in this field [12, 13]. Similarly to [12] we treat each color channel separately until we aggregate their similarity values. By using color information in the stereo matching algorithm we can avoid some false correspondence produced by color ambiguities.

In [12] they propose to use the minimum as aggregation function for this task, but some inconsistencies can be produced. For example, a pixel with low similarity values in all channels will have a larger matching score than another pixel with a great value of similarity in two channels and the other value near (but under) the similarities of the first pixel. Hence, the minimum would cause some undesirable matches (mismatches). Therefore, it is necessary to study of several aggregation functions to find which one is the most suitable (robust) to aggregate color similarities in the stereo matching problem. Next we present different aggregation functions that we will analyze in the experimental study presented in Section 4.

**Note:** We denote a vector of  $n$  elements with  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ .

- **Minimum**

$$M(\mathbf{x}) = \min(x_1, x_2, \dots, x_n) \quad (8)$$

- **Product**

$$M(\mathbf{x}) = \prod_{i=1}^n x_i \quad (9)$$

- **Arithmetic Mean (A-Mean)**

$$M(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i \quad (10)$$

- **Weighted Mean (W-Mean)**

$$M(\mathbf{x}) = \sum_{i=1}^n x_i \cdot w_i \quad (11)$$

where  $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$  is the weight vector that satisfies  $\sum_{i=1}^n w_i = 1$ .

In our comparison we consider different weight vectors to compute the final similarity:

$$\mu(x) = w_R \cdot \mu_R(x) + w_G \cdot \mu_G(x) + w_B \cdot \mu_B(x) \quad (12)$$

For example if  $w_R = 0.1$ ,  $w_G = 0.8$  and  $w_B = 0.1$ , we obtain

$$\mu(x) = 0.1 \cdot \mu_R(x) + 0.8 \cdot \mu_G(x) + 0.1 \cdot \mu_B(x) \quad (13)$$

If  $w_R = 0,299$ ,  $w_G = 0,5870$  and  $w_B = 0,1140$ , we obtain

$$\mu(x) = 0,299 \cdot \mu_R(x) + 0,5870 \cdot \mu_G(x) + 0,1140 \cdot \mu_B(x) \quad (14)$$

The weights values of Eq. (14) belong to the computation of the luminance of a *RGB* image [20]. The expression of luminance is used to transform *RGB* color images into gray scale. The purpose of luminance is to represent the brightness of colors just as human perceive them. In this manner, it represents that humans consider the color green brighter than the color blue.

- **Harmonic Mean (H-Mean)**

$$M(\mathbf{x}) = n \left( \sum_{i=1}^n \frac{1}{x_i} \right)^{-1} \quad (15)$$

- **Median**

$$M(\mathbf{x}) = \begin{cases} \frac{1}{2}(x_{(k)} + x_{(k+1)}), & \text{if } n = 2k \text{ is even} \\ x_{(k)}, & \text{if } n = 2k \text{ is odd,} \end{cases} \quad (16)$$

where  $x_{(k)}$  is the  $k$ -th largest (or smallest) component of  $\mathbf{x}$ .

- **Geometric Mean (G-Mean)**

$$M(\mathbf{x}) = \left( \prod_{i=1}^n x_i \right)^{(1/n)} \quad (17)$$

- **Mode**

The mode is the most frequent value in  $\mathbf{x}$ .

We also consider in our experiments the dual aggregation function of the Geometric and Harmonic means constructed as:  $M(\mathbf{x}) = 1 - M(1 - \mathbf{x})$ .

#### 4. Experimental results

In our experimental study, we compare the behavior of different aggregation functions to aggregate color similarities. We have three main objectives:

- To check whether using color information by aggregating the similarities improves the results of using gray scale images.
- To study which aggregation fits each correlation method or similarity measure in order to carry out a global analysis.
- To study which is the best (most robust) aggregation function in order to combine color similarities among all methods.

In order to evaluate the performance we use the Middlebury test bed proposed by Scharstein and Szeliski [1] (<http://cat.middlebury.edu/stereo>), which is established as the common benchmark for stereo matching methods and allows one to easily reproduce the results obtained. In this test, the disparity maps obtained by each algorithm are compared with the ideal disparity maps. The test images are shown in Fig. 2 with their corresponding ideal disparity maps. We refer to each image pair by the name given in [1]: “Tsukuba”, “Teddy” and “Cones”. We have to recall that stereo matching algorithms do not use any type of preprocessing or post processing steps (such as, optimization techniques, occlusion detection or image filtering).

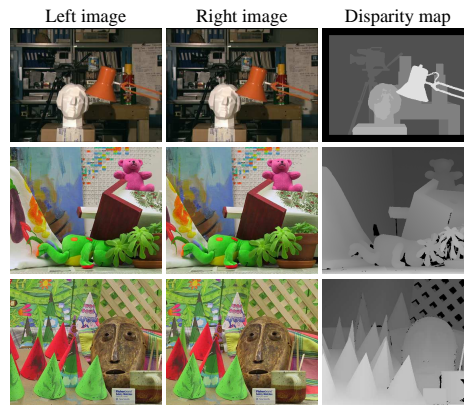


Fig. 2. Test images from Middlebury test bed

As a particular case, we start presenting the quantitative results for the stereo matching algorithm using the fuzzy similarity measure ( $SM_{FS}$ , Eq. (4)) and different aggregation functions to merge color channels similarities in Table 1. The leftmost column of Table 2 indicates the aggregation function used. Then, we present three columns for each image pair. These columns represent the percentage of absolute disparity error greater than one for three different regions in the image:

- **no-oc.**: only non-occluded pixels are considered.
- **all**: whole image is considered.
- **disc.**: only pixels near discontinuities are considered.

The rightmost column is the overall performance of the algorithm computed by the arithmetic mean of all other columns. The results are listed following the total error in descending order, and the row corresponding to the stereo algorithm applied to gray scale images is shaded to ease the comparison with respect to the performance using color images.



Table 1. Quantitative evaluation results for different aggregation functions to add color similarities using fuzzy similarity measure where \*Weighted Mean 262 means that  $w_R = 0.2$ ,  $w_G = 0.6$  and  $w_B = 0.2$

Aggregation	Tsukuba			Teddy			Cones			%E <sub>t</sub>
	%no-oc	%tot	%disc	%no-oc	%tot	%disc	%no-oc	%tot	%disc	
A-Mean	7.44	9.27	18.21	18.35	26.67	31.56	12.47	21.96	23.07	<b>18.78</b>
Product	7.46	9.31	18.11	18.59	26.87	31.72	12.49	21.96	22.83	18.81
G-Mean	7.46	9.31	18.11	18.59	26.87	31.72	12.49	21.96	22.83	18.81
H-Mean	7.45	9.33	18.10	18.87	27.11	31.91	12.56	22.02	<b>22.72</b>	18.90
G-Mean Dual	7.52	9.33	18.34	18.81	27.09	31.55	12.69	22.17	23.54	19.00
W-Mean Luminance	<b>7.37</b>	<b>9.23</b>	<b>17.85</b>	<b>18.01</b>	<b>26.37</b>	<b>30.63</b>	14.04	23.32	24.43	19.03
W-Mean 262*	7.45	9.26	18.10	18.37	26.68	30.79	13.46	22.83	24.36	19.03
W-Mean 622*	7.48	9.34	17.99	19.07	27.34	32.33	13.61	22.93	23.55	19.29
H-Mean Dual	7.65	9.41	18.71	19.96	28.12	31.82	13.23	22.67	24.02	19.51
W-Mean 181*	7.63	9.43	18.39	19.41	27.60	31.04	14.87	24.10	25.83	19.81
W-Mean 226*	8.26	9.97	19.38	21.38	29.36	33.76	<b>12.29</b>	<b>21.90</b>	23.94	20.02
Gray Scale	7.84	9.71	18.66	20.07	28.18	30.85	15.57	24.61	26.35	20.21
W-Mean 811*	7.82	9.71	18.00	20.71	28.81	33.69	15.11	24.25	24.98	20.34
Median	8.12	9.91	19.11	21.93	29.88	33.92	15.56	24.76	25.73	20.99
Min	8.18	10.10	18.32	26.02	33.49	36.18	16.02	25.07	25.03	22.04
Mode	8.40	10.32	18.52	26.73	34.12	36.56	16.40	25.42	25.31	22.42
W-Mean 118*	9.29	10.95	21.15	25.82	33.35	37.03	14.40	23.84	26.72	22.51

Following Table 1, we can conclude that using color information is beneficial. Notice that the aggregations using color information performing worse than the usage of gray scale images mainly consider one of the coefficients in the aggregation (in the case of the weighted means, the weights give most of the importance to a unique channel). However, these statements are based on a unique similarity measure; hence, we should study whether these conclusions are maintained across all metrics.

Table 2 summarizes the total error obtained for the considered metrics and aggregations. The last two columns present the mean error for each aggregation and the average rank (and rank position) respectively. The average rank is obtained by assigning a position to each aggregation depending on its performance on each metric. The aggregation which achieves the best accuracy in a specific method will have the first ranking (value 1); then, the aggregation with the second best accuracy is assigned rank 2, and so forth. This task is carried out for all metrics and finally an average ranking is computed as the mean value of all rankings. It provides a quick view of the overall behavior of each aggregation with respect to the others. Bold numbers indicate the best aggregation (rank 1) among the method. In addition, Fig. 3 shows the comparison of the best disparity maps obtained using color aggregation (those stressed in bold-face in Table 2) and the ones obtained with gray scale images.

This experiment shows that the optimal aggregation function to be used depends on the stereo matching algorithm considered. However, the previously stated facts are confirmed, the usage of color information can improve the matching as long as all color channels are taken into account. The usage of color information may need a greater computational effort in order to compute the disparity map, but it can be really low since the process can be easily carried out in parallel, whereas the obtained improvements can make a difference. It is remarkable that the total error obtained using the weighted arithmetic mean based on the luminance formula is always better than using gray scale images. Gray scale images are obtained by the *RGB* luminance formula, and hence, it is advisable to use the three color channels to compute the matching independently, aggregating the matching scores instead of aggregating the information first and then applying the algorithm for a unique color channel.

Table 2. Total error obtained for each color aggregation and similarity measure where \*Weighted Mean 262 means that  $w_R = 0.2$ ,  $w_G = 0.6$  and  $w_B = 0.2$ .

Aggregation	SAD	SSD	NCC	$SM_{FS}$	$SM_M$	$SM_K$	$SM_{UI}$	$Av.E_i$	Avg. Rank
G-Mean Dual	22.16	24.31	23.34	19.00	21.82	<b>23.93</b>	<b>21.78</b>	<b>22.33</b>	<b>1 (3.14)</b>
W-Mean Luminance	22.41	24.24	24.94	19.03	22.50	24.23	22.43	22.82	2 (4.43)
H-Mean Dual	22.17	24.31	<b>22.88</b>	19.51	<b>21.81</b>	24.53	21.81	22.43	3 (4.71)
W-Mean 622*	22.43	24.30	25.39	19.29	22.47	24.31	22.36	22.94	4 (5.43)
Product	<b>22.06</b>	<b>23.83</b>	26.71	18.81	22.83	24.33	22.73	23.04	5 (5.50)
G-Mean	<b>22.06</b>	<b>23.83</b>	26.71	18.81	22.83	24.33	22.73	23.04	5 (5.50)
A-Mean	22.13	24.32	26.64	<b>18.78</b>	22.70	24.33	22.48	23.05	7 (5.79)
W-Mean 262*	22.47	24.32	25.88	19.03	22.65	24.32	22.55	23.03	8 (6.86)
H-Mean	22.32	24.24	26.77	18.90	22.97	24.33	22.97	23.21	9 (7.64)
W-Mean 811*	23.10	24.69	24.56	20.34	22.96	24.69	22.93	23.32	10 (9.43)
Gray Scale	23.20	24.87	23.49	20.21	23.06	24.86	23.06	23.25	11 (10.57)
W-Mean 181*	23.14	24.72	25.32	19.81	23.14	24.71	23.11	23.42	12 (10.71)
W-Mean 226*	23.07	25.32	28.61	20.02	24.26	25.34	24.07	24.38	13 (12.57)
Median	24.00	25.57	24.97	20.99	24.13	25.53	24.13	24.19	14 (12.71)
W-Mean 118*	24.93	26.64	30.14	22.51	26.04	26.68	25.92	26.12	15 (15.43)
Min	24.83	26.70	31.33	22.04	27.97	26.78	27.97	26.80	16 (15.93)
Mode	25.11	26.74	31.33	22.42	27.97	26.78	27.97	26.90	17 (16.64)

Moreover, among the tested aggregations, the usage of the dual of the geometric mean (G-Mean Dual), the weighted mean with the weights from luminance formula (W-Mean Luminance) and the dual of the harmonic mean (H-Mean Dual) can be recommended, since they have stand out as the most robust ones. They behave well within different images and metrics, which address for their appropriateness on different frameworks. We should also note the good behavior of the product and the geometric mean in SAD and SSD, which are commonly used [6, 7]. Both aggregations obtain equivalent results despite of the similarity values are different, since the order between these scores is the same (the results of the geometric mean are equal to the product ones but applying the root).

## 5. Conclusions

We carried out a comparison study of the performance of different aggregation functions in the stereo matching algorithm to aggregate the similarities from different color channels in *RGB* color space. We can conclude that it is better to make the color aggregation after the similarities are computed in order to avoid ambiguities (produced by color) than to aggregate the color to obtain gray scale images and then compute the similarities. That is, color information is useful for the matching process and must not be overlooked.

The experiment has shown that despite the optimum aggregation function depends on the metric used, there are robust aggregations such as the dual of the geometric and harmonic mean, the weighted arithmetic mean based on the luminance formula, the geometric mean or the product which performs properly in all metrics among all images and hence, whose usage can be recommended.

## Acknowledgements

This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2010-15055, TIN2011-29520 and the Research Services of the Universidad Publica de Navarra.

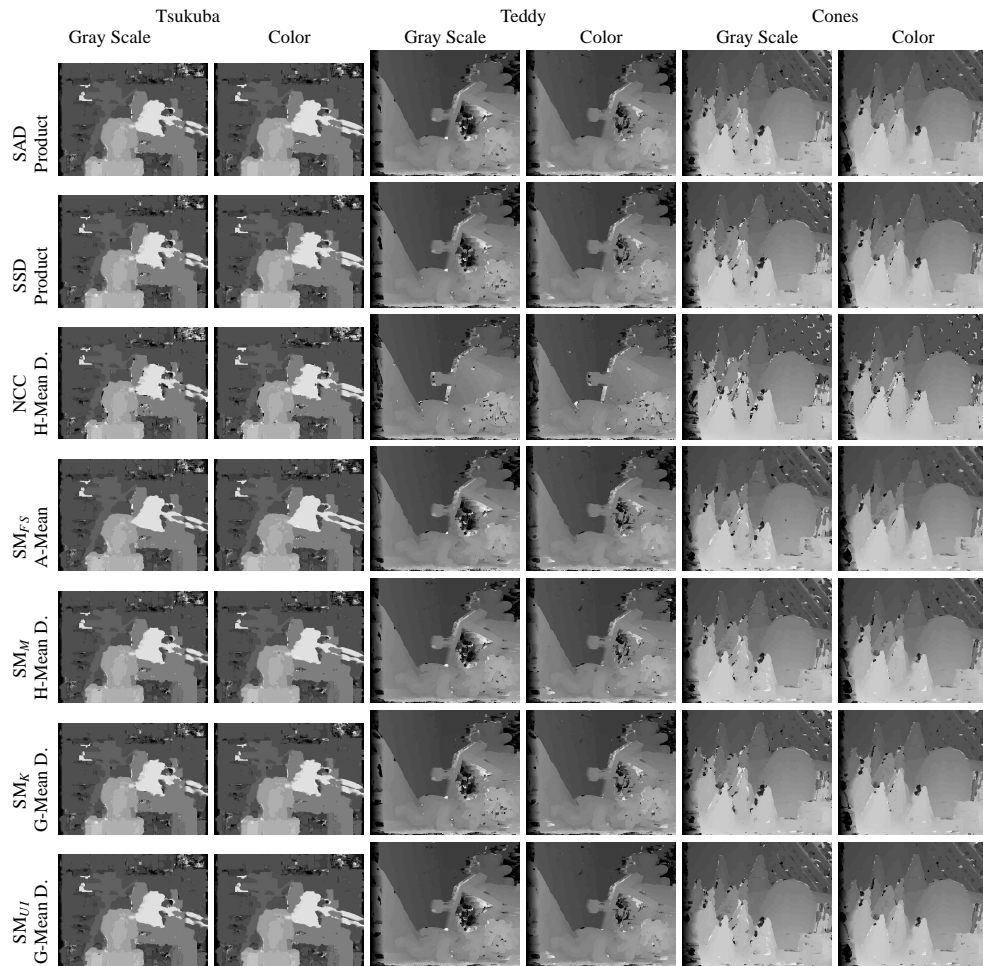


Fig. 3. Best disparity maps for each method compared with the disparity map extracted from gray scale images