CHROMATIC BLOCK MATCHING FOR DENSE STEREO CORRESPONDENCE

ANDREAS KOSCHAN
Institute for Technical Informatics, FR 3-11
Technical University Berlin, Franklinstr. 28-29
10587 Berlin, Germany

ABSTRACT

Only few problems in computer vision have been investigated more vigorously than stereo. Nevertheless, almost all methods use only gray values and most of them are feature-based techniques, i.e., they produce only sparse depth maps. This paper presents an efficient technique to obtain dense stereo correspondence using a new Chromatic Block Matching. Three different color models (RGB, $I_1I_2I_3$, HSI) and five different color measures have been investigated regarding their suitability for stereo matching. Furthermore, the results have been compared to the results of the gray value algorithm. However, the precision of the matching results always improved by 20 to 25% when using color information instead of gray value information. Moreover, we found that the $I_1I_2I_3$ color space provided the best information in this comparison.

1. Introduction and Motivation

Stereo is a well-known technique for obtaining depth information from digital images. The key problem in stereo is how to find the corresponding points in the left and in the right image, referred to as the correspondence problem. Whenever the corresponding points are found, the depth can be computed by triangulation. Worldwide many research activities dealing with stereo vision are known. Nevertheless, almost all authors use only gray value images. In our group we investigated 73 gray value stereo approaches and implemented 10 of them. The results are rather acceptable, but higher precision is required for the reconstruction of visible surfaces. There are two ways to improve the results. One approach is to seek new mathematical techniques; another approach -- as is done here -- is to seek a more complete and efficient use of available image information. That is the analysis of color in stereo images.

There are several motivations for using chromatic information. First, chromatic information is easily obtained with high precision when using a 3-chip CCD camera. Second, color plays an important role in human perception. Livingstone and Hubel showed that humans cannot perceive depth in color stereograms when the colors are at equiluminance. Third, it is obvious that red pixels cannot match with blue pixels although their intensities are equal or similar. Last not least, the existing computational approaches to color stereo correspondence have shown that the matching results can be considerably improved when using color information.

Most color stereo techniques are feature-based, i.e., only scattered control points can be computed for the succeeding surface reconstruction process. Therefore, there is an essential need for algorithms that compute dense disparity maps defined for every pixel in the entire image. Unfortunately, the dense color stereo techniques are very
time consuming because they use simulated annealing\(^6\) or a statistical criterion and an
iterative technique with a priori unknown convergence speed\(^7\).

In this paper a new efficient algorithm for dense color stereo matching is
introduced. The Block Matching technique is chosen for extension to color because of its
efficiency already shown for gray value images\(^8\). So far, all methods mentioned above
only use the RGB color space. In addition, it will be shown that the precision of color
stereo matching is improved when a suitable color coordinate system and/or color
measure is chosen.

2. Color Models and Color Metrics

The selection of a suitable color space or color coordinate system, respectively, is
one of the most sensitive and vigorously discussed problems in color vision. Some of the
color spaces are more related to technical requirements and others are more related to
human perception. In this paper three standard color models (technical and/or perceptual)
will be chosen and investigated.

2.1. Selection of a Color Coordinate System

The RGB space is the most common color space since the C.I.E. developed that
standard system in 1931 with the three monochromatic primaries Red, Green, and Blue.
These three coordinates (R, G, and B) are the reference colors in almost all image
acquisition processes, i.e., when a color picture is digitized we usually get a
representation in RGB-coordinates.

In general, the R, G, and B components are highly correlated. Thus, Ohta, Kanade,
and Sakai\(^9\) suggest to use three color features called I\(_1\)I\(_2\)I\(_3\) being good approximations
to the results of the Karhunen-Loève transform (investigated on 8 images). The three
coordinates are defined by

\[
\begin{align*}
I_1 &= \frac{R + G + B}{3}, \\
I_2 &= \frac{R - B}{2}, \\
I_3 &= \frac{2G - R - B}{4}
\end{align*}
\]  

(1)

Both color spaces mentioned above do not correspond to human color perception.
Thus, an additional color space is needed for the perceptual representation of the colors.
The perceptual color space is usually described by hue, saturation, and intensity (HSI).
The three components H, S, and I are defined by (compare 10):

\[
\begin{align*}
I &= \frac{R + G + B}{3}, \\
S &= 1 - 3 \min \{R, G, B\} \\
H &= \arccos \left[ \frac{1/2 \left( (R - G) + (R - B) \right)}{\sqrt{(R - G)(R - G) + (R - B)(G - B)}} \right], \quad \text{if } B > G \text{ then } H = 2\pi - H.
\end{align*}
\]  

(2)

All three color coordinate systems RGB, I\(_1\)I\(_2\)I\(_3\), and HSI are used and compared for
the evaluation of matching results between color stereo images.

2.2 Selection of a Color Measure

Whenever a color space is selected, the question remains how to describe the
differences of colors in this space. When the tristimulus color solid is considered as a
Riemannian space, some well-known metrics can be adapted to measure the color
difference. For any tristimulus color space ABC (representing RGB, I₁I₂I₃, etc.) the normalized values of the three color channels are defined by

\[ a = \frac{A}{A + B + C}, \quad b = \frac{B}{A + B + C}, \quad c = \frac{C}{A + B + C}. \]  

(3)

Now the color value of a pixel is represented by the normalized values (a, b, c) of the ABC color space and the difference between two colors \( f_1 = (a_1, b_1, c_1) \) and \( f_2 = (a_2, b_2, c_2) \) can be defined as:

\[ D_1(f_1, f_2) = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2}, \]  

(4)

\[ D_2(f_1, f_2) = |a_1 - a_2| + |b_1 - b_2| + |c_1 - c_2|, \]  

(5)

\[ D_3(f_1, f_2) = |a_1 - a_2|^2 + |b_1 - b_2|^2 + |c_1 - c_2|^2, \]  

(6)

\[ D_4(f_1, f_2) = \max \{ |a_1 - a_2|, |b_1 - b_2|, |c_1 - c_2| \}. \]  

(7)

When using the RGB space the distance \( D_1 \) represents the angle between two color vectors \( F_1 \) and \( F_2 \) in the RGB space (see Fig. 1). \( D_2, D_3, \) and \( D_4 \) are approximations of \( D_1 \).

![Fig. 1: Representation of two color vectors \( F_1 \) and \( F_2 \) in the RGB space.](image)

The measures mentioned above are analogously defined for non-normalized values. They are easy to compute and, therefore, are widely used in color vision. Nevertheless, there exists no connection between these measures and human color perception. Moreover, these formulas are not very suitable for the measurement in the HSI space. In that color space we use an additional measure suggested by Tseng and Chang. Assuming \( f_1 = (H_1, S_1, I_1) \) and \( f_2 = (H_2, S_2, I_2) \) are coordinates of two color points in the HSI color space, the color difference between these two points is defined as:

\[ D_5(f_1, f_2) = \sqrt{(d_I)^2 + (d_C)^2}, \]  

(8)

with

\[ d_I = |I_1 - I_2| \quad \text{and} \quad d_C = \sqrt{(S_1)^2 + (S_2)^2 - 2S_1S_2\cos\Theta}, \]

where

\[ \Theta = \begin{cases} |H_1 - H_2| & \text{if } |H_1 - H_2| \leq \pi \\ 2\pi - |H_1 - H_2| & \text{if } |H_1 - H_2| > \pi \end{cases} \]

All measures \( D_1 \) to \( D_5 \) have been selected for the implementation and evaluation process.

3. Chromatic Block Matching

A common technique for image sequence coding is based on motion analysis. Here, the actual image is determined from the temporal preceding image using motion vectors for the pixels. A motion vector is defined by the change of location of a pixel in two
temporally successive images (see Fig. 2). The temporal change between the images corresponds to the different viewpoints of the two cameras in stereo vision. In this way the disparity between both images instead of the motion vector is determined. If the optical axes of the cameras are in parallel (standard stereo geometry), the motion vectors or the disparities, respectively, have to be computed only in the horizontal direction (i.e., in one row. See Fig. 2).

![Diagram showing motion vectors and disparities](image)

Fig. 2: The left figure shows a motion vector of a pixel between the time t and the time \( t + \Delta t \). The right figure shows a disparity vector.

The main idea of Block Matching is a similarity check between two equal sized blocks (nxm-matrices) in the left and the right image (area-based stereo). The same disparity value is assumed for all pixels of one block. Therefore, only one disparity value has to be estimated for every block. This technique can be divided into several processing steps. First, one of the images (e.g. the left) is segmented into a constant number of equal sized blocks. The search for a corresponding block in the right image is only carried out for the segmented blocks in the left image. The mean square error MSE between the pixel values inside the respective blocks is used as measure for the similarity of two blocks. MSE is defined for gray value images as:

\[
\text{MSE} (\Delta) = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| I_L(i,j) - I_R(i+\Delta,j) \right|^2,
\]

(9)

where \( I_L \) and \( I_R \) are the intensity functions of the left and right image and \( \Delta \) is an offset describing the difference \((x_R-x_L)\) between the column positions in the left and in the right image. This definition can be easily extended to color images by using one of the measures defined in (4) to (8). The left image \( I_L \) and the right image \( I_R \) can be represented for any tristimulus color space as \( I_L(i,j) = (A_L(i,j), B_L(i,j), C_L(i,j)) \) and \( I_R(i,j) = (A_R(i,j), B_R(i,j), C_R(i,j)) \). When \( D_3 \) is used (9) changes to

\[
\text{MSE}_{\text{color}} (\Delta) = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} D_3(I_L(i,j), I_R(i+\Delta,j))
\]

(10)

\[
= \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} |a_L(i,j) - a_R(i+\Delta,j)|^2 + |b_L(i,j) - b_R(i+\Delta,j)|^2 + |c_L(i,j) - c_R(i+\Delta,j)|^2.
\]

The disparity \( D \) between the blocks in both images is defined by the distance between the positions (the difference in the columns) of the blocks, showing the minimum mean square error in both images while the block (of size \( n \times m \)) is shifted pixel
by pixel inside the search area. Furthermore, the search area in the right image is limited in the horizontal direction by a predefined maximum disparity $d_{\text{max}}$.

$$D = \min_{|\Delta| \leq d_{\text{max}}} \left\{ \text{MSE}_{\text{color}}(\Delta) \right\}. \quad (11)$$

If there exists more than one minimum, the disparity with the smallest difference to the disparity of the neighboring block is selected (disparity smoothness constraint). An explicit disparity value can be computed for every pixel when extending and applying the pixel selection method introduced by Reuter$^{12}$ for gray value images. This method consists of three processing steps:

- median filtering of the block disparities,
- pixel selection, and
- median filtering of the pixel disparities.

While median filtering is clear, pixel selection has to be detailed below. Pixel selection is the determination of a disparity for every pixel $(x', y')$ inside a block by using the disparity values of the neighboring blocks. The differences $\text{DIFF}(k)$ between the color value at $(x', y')$ in the left image and the color values at $(x'+D(k), y')$ in the right image are determined for all disparities $D(k)$ ($1 < k < 9$) inside the $3 \times 3$-neighboring blocks using one of the color measures (for example $D_3$):

$$\text{DIFF}(k) = D_3( I_L(x', y'), I_R(x'+D(k), y')) \quad \text{with} \quad k = 1, \ldots, 9. \quad (12)$$

The pixel disparity value $\text{DISP}(x', y')$ is defined by the value $D(k)$ minimizing the difference DIFF(k). A dense disparity map is generated when applying this pixel selection technique to every pixel in the image. The Chromatic Block Matching algorithm is implemented as modular program to enable selective testing and evaluation.

4. Experimental Results and Discussion

Six real test images have been generated for the evaluation. The color stereo images represent natural opaque objects of different complexity and different colors. All possible configurations of the algorithm have been applied to these stereo images (a total of 875 tests). The reconstructed right stereo images $I_{\text{Rrec}}$ have been computed using the left stereo image and the disparity matrix. The difference images DFD (displaced frame difference) between the reconstructed and the original images have been computed as well. Moreover, the number $M_{\text{DIFF}}$ and the mean error of the false disparity values have been computed concerning the block sizes, the color models, and the color measures.

As a result, most of the false matches have been detected for pixels representing the image background. This was to be expected because the image background is generally represented by a homogeneous color distribution. Opposed to that, the best results have been detected near the edges of the objects in the images. The matching results have been evaluated regarding the best block size, the best color space and the best color measure.

Block Sizes (BS): The five block sizes $4 \times 4$, $6 \times 6$, $8 \times 8$, $10 \times 10$ and $12 \times 12$ pixels have been investigated for the three color spaces and the corresponding color measures. The selection of the most suitable block size depends on the image function and not on the color space or color measure. In summary, good results have always been obtained with
8x8 blocks, while the mean error of the disparities increases when using a small size (4x4) or a large size (12x12).

**Tab. 1: Mean error of color values in the 6 reconstructed images with different BS.**

<table>
<thead>
<tr>
<th>block size</th>
<th>4 x 4</th>
<th>6 x 6</th>
<th>8 x 8</th>
<th>10 x 10</th>
<th>12 x 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean error</td>
<td>10.2</td>
<td>10.9</td>
<td>10.4</td>
<td>11.1</td>
<td>13.0</td>
</tr>
</tbody>
</table>

**Tab. 2: Mean error of the disparity values in image SHAWL with different BS.**

<table>
<thead>
<tr>
<th>block size</th>
<th>4 x 4</th>
<th>6 x 6</th>
<th>8 x 8</th>
<th>10 x 10</th>
<th>12 x 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean error</td>
<td>6.1</td>
<td>4.9</td>
<td>3.5</td>
<td>3.7</td>
<td>5.8</td>
</tr>
</tbody>
</table>

**Tab. 3: Percentage of right matches in image SHAWL with different BS.**

<table>
<thead>
<tr>
<th>block size</th>
<th>4 x 4</th>
<th>6 x 6</th>
<th>8 x 8</th>
<th>10 x 10</th>
<th>12 x 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>73.8</td>
<td>79.4</td>
<td>83.2</td>
<td>85.2</td>
<td>83.5</td>
</tr>
</tbody>
</table>

**Color Spaces (CS):** The algorithm has been tested with the three color spaces RGB, I₁I₂I₃, HSI, and the gray value image representation. The precision of the matching results always increased when color information was used instead of gray value information. The mean error of the false disparities was reduced by 20 to 25 % with color. Surprisingly, the mean error increased slightly by an average of 0.3 when the HSI presentation was selected. Presently we do not know whether this happened due to the HSI representation or due to the color measure. This will be part of further investigations. In summary, the best results have always been obtained when using the I₁I₂I₃ color space.

**Tab. 4: Mean error of color values in the 6 reconstructed images in different CS.**

<table>
<thead>
<tr>
<th>color space</th>
<th>RGB</th>
<th>I₁I₂I₃</th>
<th>HSI</th>
<th>rgb</th>
<th>Gray</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean error</td>
<td>8.7</td>
<td>8.7</td>
<td>9.8</td>
<td>14.6</td>
<td>11.2</td>
</tr>
</tbody>
</table>

**Tab. 5: Mean error of the disparity values in image SHAWL in different CS.**

<table>
<thead>
<tr>
<th>color space</th>
<th>RGB</th>
<th>I₁I₂I₃</th>
<th>HSI</th>
<th>rgb</th>
<th>Gray</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean error</td>
<td>4.3</td>
<td>4.2</td>
<td>4.6</td>
<td>5.3</td>
<td>5.6</td>
</tr>
</tbody>
</table>

**Tab. 6: Percentage of right matches in image SHAWL in different CS.**

<table>
<thead>
<tr>
<th>color space</th>
<th>RGB</th>
<th>I₁I₂I₃</th>
<th>HSI</th>
<th>rgb</th>
<th>Gray</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>82.5</td>
<td>82.4</td>
<td>80.9</td>
<td>79.8</td>
<td>78.7</td>
</tr>
</tbody>
</table>

**Color Measures (CM):** The color differences have been measured with the formulas (4) to (7) in the color spaces RGB and I₁I₂I₃ and with equation (8) in the HSI space. Additionally, normalized and non-normalized values have been compared in the RGB space for the measures (4) to (7). Generally, the mean error of the false disparities was double as high using normalized values (3) than it was using non-normalized values. However, no significant influence of the color measure on the processing results has been found yet for all non-normalized values in all investigated color spaces. Thus, we believe that any approximation D₂, D₃, or D₄ of the Euclidean metric may be used in measuring color differences if computing time shall be saved. Nevertheless, further investigations are needed to prove this assumption.
Unfortunately, due to space limitation not all the results can be presented in this paper. Thus, the results of Chromatic Block Matching using the I_1I_2I_3 space, the Euclidean measure (4), and a block size of 8x8 pixel are chosen for representation. The original image DIA2 and its reconstructed right image are shown in Fig. 3. The false matches are presented with the color black in the reconstructed image to visualize the loci of the false matches. The original image SHAWL is shown in Fig. 4 with the manually generated depth map in the middle and the depth map obtained when applying Chromatic Block Matching at the right.

5. Conclusion

A new and efficient method for dense stereo correspondence has been presented. It has been shown that the results of Block Matching can be considerably improved by using color information. Previously all other dense color stereo approaches had only used the RGB image representation. In this investigation a further improvement in the results occurred when using the I_1I_2I_3 color space instead of the RGB solid. No significant influence of the color measure on the processing results has been found yet for all non-normalized values in all investigated color spaces. Thus, we believe that any approximation D_2, D_3, or D_4 of the Euclidean metric may be used in measuring color differences if computing time shall be saved. Additional tests and statistical investigations are necessary for a more detailed evaluation of this technique. This is done at present and further results will be presented soon. In summary, we believe that precise results in dense stereo matching can be obtained more easily using this efficient new color method.

Acknowledgment

I should like to thank Alexander Bachem and Rafał Salustowicz for realizing the implementation.

References


Fig. 3: Gray value reproduction of the original stereo image DIA2 (left) and the reconstructed right image (right) with the false matches coded as black pixels.

Fig. 4: Gray value reproduction of the original stereo image SHAWL (left), the ideal synthetic depth map (middle), and the computed depth map (right).