SCHOOL OF ENGINEERING - STI
ELECTRICAL ENGINEERING INSTITUTE
SIGNAL PROCESSING LABORATORIES

Pablo Revuelta Sanz

EPFL - STI - IEL - LTS5
Station 11
CH-1015 LAUSANNE

Phone: +41 21 693 78 91
E-mail address: prevuelt@ing.uc3m.es

Reference: EPFL-REPORT-150511

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Stereo Vision Matching using Characteristics Vectors

Pablo Revuelta Sanz[1], Belén Ruiz Mezcua[1], José M. Sánchez Pena[1], Jean-Phillippe Thiran[2]

[1]*Carlos III University of Madrid, Spanish Center for Captioning and Audiodescription.*
Av. Peces Barba, 1. 28918 Leganés, Madrid. Spain.
[2]*Signal Processing Laboratory (LTS5), École Polytechnique Fédéral de Lausanne. (EPFL),*
Station 11, CH1015 Lausanne, Switzerland.

Abstract

Stereo vision is a usual method to obtain depth information from images. The problems encountered when applying the majority of well established algorithms to provide this information are due to the high computational load required. This occurs in both the block matching and graphical cues (such as edges) matching. In this article we address this issue by performing an image analysis which considers each pixel only once, thus enhancing the efficiency of the image processing. Additionally, when matching is carried out over statistical descriptors of the image regions, commonly referred to as characteristic vectors, whose number of these vectors is, by definition, lower than the possible block matching possibilities, the algorithm achieves an improved level of performance. In this paper we present a new algorithm which has been specifically designed to solve the commonly observed problems which arise from other well know techniques. This algorithm was designed using a previous work carried out by the authors in this area to determine the descriptors extraction processes. The complete analysis has been carried out over gray scale images. The results obtained from both real and synthetic images are presented in terms

of matching quality and time consumption and compared to other published results. Finally, a discussion is provided on additional features related to the matching process.

Keywords: stereo vision; region growing; matching; depth map; characteristics;

1. Introduction

Stereo vision is a process that provides 3D perception by means of two different images of the same scene. This process is of great importance in computer vision, as it allows to obtain the distance from the cameras to any specific object of the scene.

The calculations required for depth mapping of images has been studied in detail. This procedure can be implemented using both passive and active methods. In the former group several approaches exist which are based on monocular vision, using structures within the image [1], the relative movement of independent points located in the image [2], using visual cues from the image [3], the focus properties of the image [4], stereovision [5] and multiview systems [6,7]. Within the field of active methods, several systems are present which make use of technologies based on laser [3], ultrasound [8], pattern projection [9] or X rays [10].

The main drawbacks associated with active methods are due to sophisticated hardware requirements and the energy required to obtain an accurate depth map of the image. In this paper we propose a passive technique, thus reducing the amount of energy required allowing it to be implemented within portable hardware.

Different approaches exist when considering passive methods, depending on the number of cameras used and the position of these cameras. It is important to indicate that point tracking methods, as described in [2], only retrieve a relative measurement of the distance. Other passive methods based on stereovision and multivision allow real distance maps to be computed with the aid of a calibration process. Multivision systems attempt to compute the depth map by recomposing the scenario from different points of view, obtaining results from a broad range of angles [11]. Data collected from these systems present a high amount of geometric distortions that must be corrected. This effect also appears when using widely separated stereovision systems [12].

In both cases, when different viewpoints from the same scene are compared, a further problem arises that is associated with the mutual identification of images. The solution to this problem is commonly referred to as matching. The matching process consists of identifying each physical points within different images [6]. The difference observed between images is referred to as disparity and allows depth information to be obtained from either a sequence of images or from a group of static images from different viewpoints.

There are three main methods which have been used to solve matching problems. The first, where a detailed description can be found in [13], is based on identifying points from specific characteristics of their background within the image. The second technique consists of identifying curves, shapes and edges [14-16]. The third most commonly used method, known as Area-based matching, attempts to match windowed areas which present similar intensities in both images [17,18].

Matching techniques are not only used in stereo or multivision procedures but also widely used for image retrieval [14] or fingerprint identification [19] where it is important to allow rotational and scalar distortions [20]. The results presented in this paper have been obtained using these processing techniques but taking especially into account scale and orientation information to accurately perform image matching.

Each of the three previously described approaches to matching problem presents several computing problems. In the case of edges, curves and shapes, a differential operator must be used (typically Laplacian or Laplacian of Gaussian, as in [21,5]). This task requires a convolution of 3x3, 5x5 or even bigger windows; as a result, the computing load increases exponentially. This problem also arises when using area-based matching algorithms. The use of a window to analyze and compare different regions is seen to perform satisfactorily [22] however this technique requires many computational resources. A further group of algorithms are based on wavelets, as described in [16]. These algorithms present similar problems.

In this paper we propose a novel matching algorithm based on characteristic vectors for gray scale images.

Considering the stereo matching approach, there are several geometrical and camera response assumptions that have been made to compare two images that, essentially, are different. These assumptions are described in detail in [6,23]. There are also various constraints that are generally satisfied by true matches thus simplifying the depth estimation algorithm, these are: similarity, smoothness, ordering and uniqueness [21]. All of these constraints are taken into account with both the so called fronto parallel and brightness constancy hypotheses [6].

Taking this into account, a region matching algorithm is proposed, that reduce the number of operations needed for stereo matching, obtaining at the same time results that are relevant compared to those found in the bibliography.

The presented algorithm works as follows:

1. Image preprocessing. First of all, a Gaussian low pass filter is explained, to reduce outlier pixels that are not representative. This task is crucial to perform the region growing algorithm, that is described in the next section:
2. Characteristics extraction by region growing. Secondly, the region growing algorithm is applied and regions descriptors are obtained.
3. Vectors Matching. Once the vectors with the extracted descriptors are created, the matching process over the pair of vectors (one vector for each region, one array of vectors for each image) is implemented.
4. Depth Estimation. The depth estimation is computed from the horizontal distance of the centroid of every matched pair of region descriptors.

This paper follows the structure of the algorithm. Additionally, a results section is presented with depth estimations of both real and synthetic standard images. An in depth discussion is done with the obtained results and a final conclusion section closes this work.

2. Image preprocessing

The application of region growing algorithms is generally preceded by image processing techniques to adjust the visual data to the algorithm affordable range. The proposed algorithm has been designed to operate on grey scale images. Color images are first converted to 256 gray levels by considering their brightness.

Additionally, a smoothing filter is applied to reduce the influence of impulsive noise on the processing. This is carried out using a 3x3 Gaussian filter. It is envisaged that this filter will be replaced by an optical filter in a final hardware implementation to reduce computational load.

The scope of the work presented in this part of the paper is restricted to the fast segmentation of different regions, and not coherent image segmentation, i.e. the fact that different segmented parts belong or not to the same physical object is not of interest. An efficient method to set regions is done by truncating the image. In this implementation, this is carried out by using the 3 most significant bits and, then the gray scale reduced to 8 levels. The truncation process has been implemented on-the-flight inside the region growing and characteristics extraction algorithm.

3. Characteristics Vectors for Image Identification

Descriptors or characteristics vectors are extracted on-the-flight during region-growing segmentation where each pixel is only examined once [24]. The most relevant characteristics obtained from each region are the area, gray value, centroid, length, width, boundary length and orientation, where these have been based on the most relevant visual cues used in human vision [25]. All of these are retrievable from the region moments [26] with some other relevant data. In the following figure, it is shown an example extracted from [24], where region is extracted from a 300x200 synthetic image and characterized.

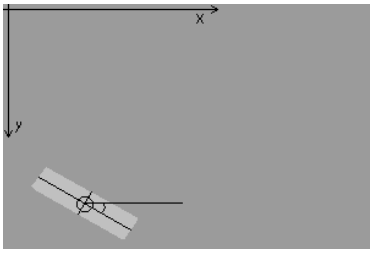


Fig. 1. Segmented region

Area [pixels]	1841
xc (of centroid) [pixels]	66.616516
yc (of centroid) [pixels]	162.355789
Angle [º]	-28.872931
Length	25.090574
Width	6.107242
Boundarylength [pixels]	292

Table 1. Characteristics descriptor of the region shown in figure 1

When the segmentation step is performed a set of characteristic vectors describing each region of the image is provided. In addition to these vectors, an image is required that maintains the reference between each pixel and the region identifier to which it belongs. This image is referred to as the “footprint image” and is composed of one byte per pixel which represents the index of the characteristics region identifier in the vector. This limits, in this implementation, the number of segmented regions to 255 (the value ‘0’ is reserved for unlabeled pixels).

Another advantage of this method over the windows and edges matching algorithm is that the interpolation process required to create the disparity surfaces [23] is avoided. This is due to the fact that in our approach virtually all pixels are linked to a region (several pixels remain unlabeled in the segmentation process as they belong to regions with insufficient area to be considered).

4. Matching based on extracted features

The matching process requires a specific characteristic which belongs to each of the different images and obtained from the different viewpoints.

Using this novel algorithm, a chain of conditions has been proposed to verify the compliance between regions. With this structure, increased efficiency is achieved as every test is not always performed. The majority of the region characteristics are compared according to expression 1:

$$val = \frac{abs\left(Ch_{left}^{i} - Ch_{right}^{i}\right)}{\max\left(Ch_{left}^{i}, Ch_{right}^{i}\right)} \qquad (1)$$

Where *i* represents the i-th characteristic (Ch) of those presented in table 1 of the left or right image.

For this case, the possible range of differences is between [0, 1]. We refer to this particular operation as Relative Difference.

Table 2 shows the order of conditions, the compared characteristic and the acceptance threshold. Thresholds and comparison tests will be explained after the table.

Item	Characteristic	Comparison test	Acceptance thresholds
1	Centroid ordinates	Absolute difference	<(Image Height)/4
2	Centroid abscissa	Absolute difference and non-negative	[0, (Image Width)/4]
3	Value	Equality	= 0
4	Area	Relative difference	<20%
5	Frontier length	Relative difference	<40%
6	Length	Relative difference	<30%
7	Width	Relative difference	<30%
8	Angle	Weighed difference	<30
9	Vote of characteristics	Absolute addition	Total Threshold

Table 2. Comparison tests and their corresponding thresholds.

The centroid coordinate matching in tests 1 and 2 is only searched in ¼ of the image in each axis, assuming all the potentially matched objects are located far enough from the cameras and in the same scan-line. Moreover, the difference of left and right centroid abscissas cannot be negative (which should represent objects far away from the infinite, or a myopic orientation of cameras. Both cases are not taken into account since the specified geometrical assumptions are applied).

The preprocessed images, as said before, have been truncated, so pixel values must have the same values to be included within the same region, as done in the third test.

The angle comparison (item number 8) needs a deeper explanation. Because of the ambiguity of the angle measurement when length and width are similar, a comparative function described by equation 2 has been implemented:

$$\Delta\alpha = 100 \cdot abs\left(\frac{2}{\pi} \cdot a\tan\left(\min\left(L_l - W_l, L_r - W_r\right)\right) \cdot \sin\left(\alpha_l - \alpha_r\right)\right) \qquad (2)$$

In equation (2) L_x and W_x are the length and the width of the left and right image regions, respectively, and α_x, denoted by the subscript *l* and *r* are the angles of the left and right image region, respectively. By using this equation the magnitude of the angle is maximized when there is a large difference between the length and width, and vice versa. This is done because the angle measurement of a compact object (similar length and width) is highly noise sensitive and is not suitable as a representative descriptor for the region.

Finally, if the result from all of the previous tests is positive, all of the differences obtained are added and compared to the sum of the applied thresholds. "Total_Threshold" is then computed in the first step by means of equation 3.

$$\text{Total\_Threshold} = \sum \text{partial\_thresholds} \qquad (3)$$

After this operation (which is always positive in the first voting test), the result is stored and used as the new Total_Threshold value for further comparisons. Using this procedure, if a further region is observed to fit more effectively to the reference region (i.e. the result from the addition is smaller), uniqueness of the matching function is enforced and only this new region will be matched (the previous region will be left unmatched).

This matching methodology has been implemented as consecutive steps in a partial matching chain:

```
For region_i^left
  Threshold = Σ partial thresholds
   For region_j^right
    If 0 < (imageLeft (centroid (X) – imageRight (centroid (X)) < width/4 Then
      If (imageLeft (centroid (Y) – imageRight (centroid (Y)) < height/4 Then
        If imageLeft(value) == imageRight(value) Then
          […]
            If Σ differences < Threshold
              Match Found
              Threshold = Σ differences
              Depth (region_i^left, region_j^right) = imageLeft(centroid(X) – imageRight(centroid(X)
```

If some of the comparisons do not comply with the partial threshold, the inner loop is broken and reinitialized, saving computational load.

As stated in the introduction of this paper, several geometrical, sensor and segmentation assumptions have been considered, resulting in the following consequences:

- No geometrical correction is implemented.
- It has been assumed that every well-matched left centroid abscissa is higher than the right one (and they are equal when the region is located at infinity) but lower than 25% of the image range. This signifies that the matching regions are assumed to be close to each other and, thus, located far enough from the cameras.
- Both images are taken from the same camera height, so the scan-line to find matches can be assumed to be horizontal where only a range of +/- 12.5% will be tolerated when searching for matches.
- No region with an area below 0.1% of the image size will be catalogued as a significant region and as a result will not be matched.

As the images projected from each camera are different, several of the areas within the image might be projected in one of them, producing what is commonly referred to as the occlusion effect. These areas cannot be matched, as it has widely been discussed in stereovision literature [27]. It will be demonstrated that the method proposed in this paper leaves several regions where no matching occurs, and they will be indiscernible from occluded regions.

5. Depth Estimation

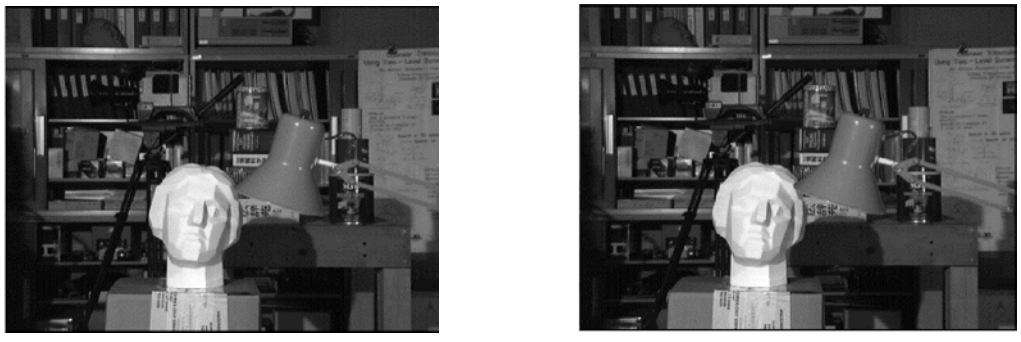
The main advantage of stereovision is the correspondence between differences in the X axis and the distance between the object and the cameras either once the cameras have been calibrated or the required assumptions have been made. The absolute distance of the centroid abscissas (in pixels) is measured for every matched pair of regions. This gives a non-calibrated measure of the depth map.

In this work the left image is used as the background and used to compute the depth map. This is an arbitrary choice, without any loss in generality, based on the fact that the left image is the last one to be segmented and processed and, then, the footprint image (that is unique in the algorithm) corresponds to this at the end of the segmentation process.

6. Results

The previously described algorithm has been implemented using the OpenCV library and tested over different standard stereo pairs of images. These tests have been implemented using a 1.6 GHz microprocessor. All the tested images encounter geometrical constrictions assumed by our algorithm.

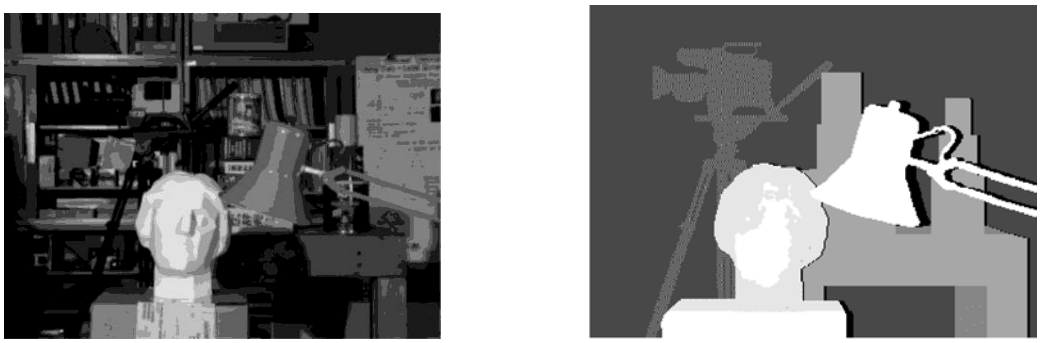
First of all, it was run over the Tsukuba pair of images which are presented in figure 2, the resolution of the image is 346x364.



(a) (b)

Fig. 2. Tsukuba pair of images: (a) left image, (b) right image

In Figure 3, the left version of the filtered and truncated image and the ground truth depth image of the Tsukuba pair are presented.



(a) (b)

Fig. 3. (a) Tsukuba filtered and truncated image and (b) ground truth depth map

In figure 3.b, the occluded areas are drawn in black. In figure 4, the results of depth estimation using the proposed algorithm are presented.

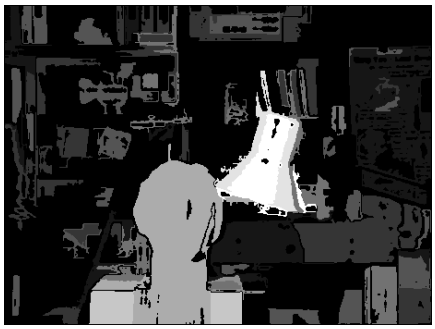
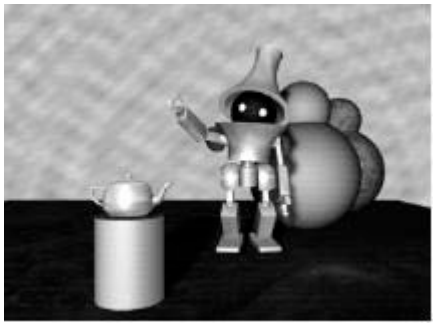


Fig. 4. Depth estimation, computed over the left image of the Tsukuba pair of images. The image is normalized to its maximum value, which is achieved in the lamp

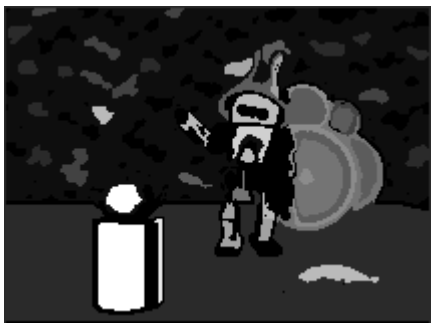
These images have been segmented into 103 and 106 regions where the algorithm takes close to 35.7 ms for the segmentation process of each image. As shown in [24], the segmentation time is directly related to the number of regions of the segmented image and to the image size.

The characteristics vectors must be prepared, normalized and several of the descriptors are computed after the image segmentation from the extracted data. The time taken in this case is close to 90.5 μs for each one. This is not significant regarding the segmentation and characteristic extraction processes. Finally, the matching has been carried out over the computed vector and not over the original images. In this case, the proposed algorithm takes up to 700 μs to compare and match both vectors. We can see the total time consumption is around 36.5 ms for this stereo pair.

In another example, this algorithm has also been tested on the synthetic image presented in figure 5.a, where the resolution was, in this case, 213x158. The computed depth map of this figure is presented in figure 5.b.



(a)



(b)

Fig. 5. (a) Synthetic left gray scale image and (b) computed depth map

The time analysis in this second example has provided the following results. The segmentation process (86 and 89 regions) takes 7.19 ms for each image. The characteristics computation carried out on the extracted data takes 84.9 μs and the matching process 808 μs. These results will be discussed in the following section.

Final tests have been carried out using the Teddy and Venus images, which are shown here along with their true depth map (figure 6).

(a)

(b)

(c)

(d)

Fig. 6. (a) Teddy left image, (b) Teddy true depth map, (c) Venus left image and (d) Venus true depth map

The corresponding results are presented in the following figure 7.

(a) (b)

Fig. 7. (a) Teddy and (b) Venus computed depth maps

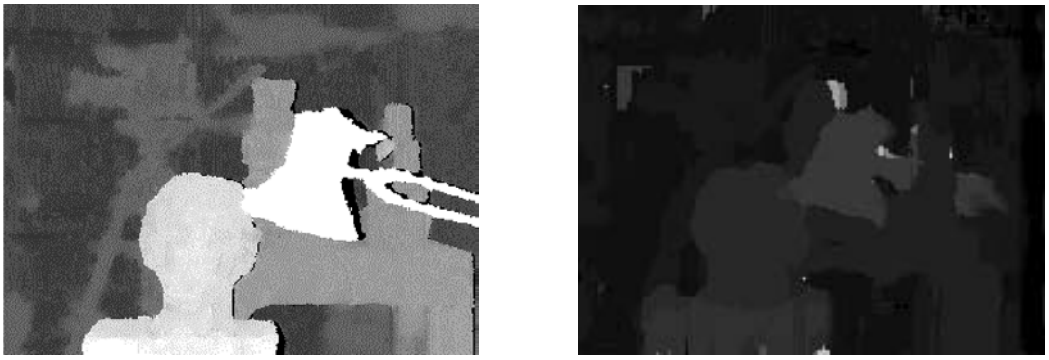
In figure 7 the computed depth map can be seen for the images presented in figure 6.

7. Discussion

As it has previously been discussed in this paper, the matching process satisfies several of the different problems that arise in methods ranging from stereovision to image retrieval. The main contribution of the proposed algorithm is to solve the matching problem by comparing characteristics of regions instead of the regions themselves. Since standard and real images are segmented into no greater than 255 regions, the computational efficiency increases with the size of the image as opposed to both windowed-areas and visual cue methods.

When designing the comparison steps, some decisions were taken, to set the partial thresholds. The value of the threshold limits the searching process, however this is not critical. The most significant parameter is the result from the voting process, this is automatically minimized during the vector comparison and, thus, matching is optimized among all the different matching possibilities.

The results obtained for the computed depth maps perceptually correspond to the true depth map. However, errors are still evident: It can be seen that the gray scale segmentation based on brightness remains sensitive to noise. Other errors arise due to problems of matching small or undifferentiated figures such as the tins located behind the lamp in figure 4. However, the difference in distance between both table legs and also between the different parts of the lamp is represented, where we can see their (truncated) curvature. Figure 8 presents some of the best results in gray scale depth maps estimations from the same pair of images of a cooperative algorithm [28] and with a windowing matching [29].



(a) (b)

Fig. 8. (a) Cooperative algorithm result taken from [28] and (b) a windowing algorithm result from [29]

This first compared algorithms works in an iterative way and, although it is a very good estimation of the depth map, it achieves this result after between 4 and 7 iterations. The second example uses 8x8 windows to find matches. In this sense, we must find a compromise between complexity and accuracy, depending on the final application of the algorithm.

The same results can be found for the Teddy image in [30].

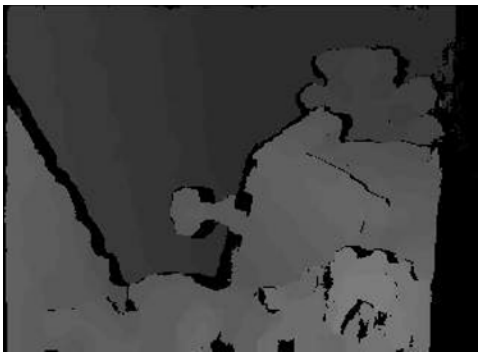


Fig. 9. Teddy computed depth map in [30]

The algorithm proposed in [30] preprocess the image with background subtraction (derivative operation) and, then, applies a 11x11 pixels window for the matching process. Because of that, the results obtained are very accurate, paying the price of a high complexity and, hence, computational load.

In example of figure 5, several errors regarding the estimated depth are present, such as the background gray levels, and the problems with depth estimation when the gray level is continuous, which can clearly be seen for the case of the floor. Considering this, the level of contrast is important in deeply relevant regions to obtain an adjusted measure of the depth.

Another relevant point to be made is related to the unmatched and non-segmented regions within this image (drawn in black). The majority of this lack of information is due to the minimum area condition required to segment a

region. The body of the second figure in these images has a complex and detailed structure that presents problems when identification is required.

Most of the black regions are then not matched and also not segmented (i.e. when the area is smaller than the minimum allowed), thus the errors propagate from this discrimination to the computation of the depth map. A critical case of this error is presented in figure 7.b, where the left written panel has not been matched. These sets of images are rich in color, however their truncation into black and white produces a deficient segmentation and, hence, deficient matching.

As stated in the depth estimation section, the images which represent the computed depth maps are based on the left of the original images. The true depth map is also provided over the left geometry of the image pairs.

One of the main shortcomings to the proposed method is based on the segmentation quality. This particular area has been thoroughly investigated (see, for example, [31]) where adaptive region growing algorithms have been proposed to perform image segmentation similarly to the segmentation process carried out by human vision. The problems that arise with this technique are again based on the computational load required by the complex algorithms and the dependence on the image structure. Moreover, they are usually context dependent and need, then, human interaction. The scope of the research work presented in this paper has enforced the implementation of a contextual independent algorithm resulting in a higher matching error rate. A compromise must be made between the segmentation, matching quality and computational complexity. If a more efficient segmentation performance is required, color segmentation can be implemented, however the main disadvantage to this process is that three times more calculations (one per channel) are required [32,33]. Figure 10 presents an example of color matching in [22].

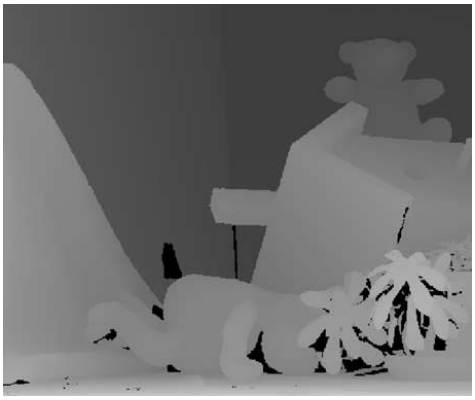


Fig. 10. Teddy color based depth map estimation in [22]

The proposed algorithm achieves, depending on the specific image pairs, an acceptable compromise between the two aforementioned factors where no false matching is observed also no interpolation, and further post processing tasks are required. The truncation preprocessing has been presented apart from the algorithm, just for a better comprehension goal, but it is implemented on-the-fly in the region growing algorithm, thus, another advantage of

the proposed algorithm is based on the fact that part of the preprocessing is carried out during the processing. No additional loops are then required in the program.

Moreover, no differential filter is required (as in [5,21,30]) for feature extraction and, in contra to these procedures, every segmented and recognized pixel is linked to a region, so the final construction of the depth map consists of searching within a Look-Up-Table (where the abscissa-differences are stored for each of the matched regions) the depth value of the pixel is related to the identifier of the footprint image, and store this value in the depth map image.

In contrast to other well known methods which perform image matching [22], we have replaced the task of comparing moving windows in both images by comparing two vectors that contain close to 100 terms in the Tsukuba images, where the resolution of this image has been the largest in this study (346x464 pixels). The comparative process, due to its nested structure, allows the time spent on many operations to be saved when certain tests have not been passed. However, we obtained an unexpected result when analyzing the time required to compute the matching of close to 90 region vectors in the second example, which was greater than that of the time required for the Tsukuba image. The explanation for this result is based on the number of steps required to be carried out in the nested comparing structure. If region descriptors are similar, the comparing structure needs to go through several steps, thus generating an increased computational load even when the number of descriptors is lower than that of the first example.

The main disadvantage of this algorithm is the dependence of the matching quality in terms of the region growing quality. This dependence has been shown in figure 6 and its results, where deficient information (because of the gray scale preprocessing) is given to the region growing procedure and, hence, the matching is not satisfactory.

Finally, it can clearly be seen that if image retrieval is to be implemented based on this proposal; only a small amount of adjustments must be considered for the comparison process. These are based on avoiding the angle and relative differences between other parameters.

8. Conclusions

An effective way of solving the problems that arise due to matching based on stereovision has been proposed and tested in this work. The main problems and solutions to performing this task have been identified, and are shown to be critical for the majority of computer vision applications.

From the results obtained and presented here, it has been shown that this area of investigation should be further developed to ascertain the limitations and for its development concerning new proposals. The algorithm presented in this work forms part of a preliminary approach to the current problem of depth mapping and motivates for further research to improve the segmentation process, thus retrieving other relevant characteristics to optimize the comparison process.

Acknowledgements

We would like to acknowledge the student grant offered by the Universidad Carlos III de Madrid and Spanish Center for Subtitling and Audiodescription (CESyA) which has allowed this research work to be performed.

References

[1] A.R.J. François, G.G. Medioni.: Interactive 3D model extraction from a single image. Image and Vis. Comp. 19, 317-328 (2001)

[2] K.E. Ozden, K. Schindler, L. van Gool.: Simultaneous Segmentation and 3D Reconstruction of Monocular Image Sequences. Computer Vision, 2007.ICCV 2007.IEEE 11th Int. Conf. on. 1-8 (2007)

[3] A. Saxena, S.H. Chung, A.Y. Ng.: 3-D Depth Reconstruction from a Single Still Image. Int. J. of Comp. Vis. 76, 53-69 (2008)

[4] A.S. Malik, T.-S. Choi.: Depth Estimation by Finding Best Focused Points Using Line Fitting. Lecture Notes in Comp. Sci, Imag. and Signal Process.. 5099, 120-127 (2008)

[5] Y. Jia, Y. Xu, W. Liu, C. Yang, Y. Zhu, X. Zhang, L. An.: A Miniature Stereo Vision Machine for Real-Time Dense Depth Mapping. LNCS, Comp. Vis. Syst. 2626, 268-277 (2003)

[6] J.-P. Pons, R. Keriven.: Multi-View Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score. Int. J. of Comp. Vis. 72 (2), 179-193 (2007)

[7] H.K.I. Kim, K. Kogure, K. Sohn.: A Real-Time 3D Modeling System Using Multiple Stereo Cameras for Free-Viewpoint Video Generation. LNCS, Imag. Anal. Recognit. 4142, 237-249 (2006)

[8] T.S. Douglas, S.E. Solomonidis, W.A. Sandham, W.D. Spence.: Ultrasound image matching using genetic algorithms. Med. and Biol. Eng. and Comp. 40, 168-172 (2002)

[9] L. Yao, L. Ma, D. Wu.: Low Cost 3D Shape Acquisition System Using Strip Shifting Pattern. LNCS, Digit. Hum. Model. 4561, 276-285 (2007)

[10] J.P.O. Evans.: Stereoscopic imaging using folded linear dual-energy x-ray detectors. Meas. Sci. and Technol. 13, 1388-1397 (2002)

[11] S.M. Seitz, J. Kim.: The Space of All Stereo Images. Int. J. of Comp. Vis. 48 (1), 21-38 (2002)

[12] T. Tuytelaars, L.v. Gool.: Matching Widely Separated Views Based on Affine Invariant Regions. Int. J. of Comp. Vis. 59 (1), 61-85 (2004)

[13] J. Yu, L. Weng, Y. Tian, Y. Wang, X. Tai.: A Novel Image Matching Method in Camera-calibrated System. Cybern. and Intell. Syst., 2008 IEEE Conf. on. 48-51 (2008)

[14] C. Schimd, A. Zisserman, R. Mohr.: Integrating Geometric and Photometric Information for Image Retrieval. LNCS, Shape, Contour and Group. in Comp. Vis. 1681, 217-233 (1999)

[15] L. Szumilas, H. Wildenauer, A. Hanbury.: Invariant Shape Matching for Detection of Semi-local Image Structures. LNCS, Imag. Anal. Recognit. 5627, 551-562 (2009)

[16] Y. Xia, A. Tung, Y.W. Ji.: A Novel Wavelet Stereo Matching Method to Improve DEM Accuracy Generated from SPOT Stereo Image Pairs. Int. Geosci. and Remote Sens. Symp. 7, 3277-3279 (2001)

[17] M.S. Islam, L. Kitchen.: Nonlinear Similarity Based Image Matching. Int. Fed. for Inf. Process. 228, 401-410 (2004)

[18] J. Williams, M. Bennamoun.: A Non-linear Filtering Approach to Image Matching. Proceedings of the 14th Int. Conf. on Pattern Recognit. 1 (1), 3 (1998)

[19] Ch. Wang, M.L. Gavrilova.: A Novel Topology-Based Matching Algorithm for Fingerprint Recognition in the Presence of Elastic Distortions. LNCS, Comp. Sci. and Its Appl. ICCSA. 3480, 748-757 (2005)

[20] Z. He, Q. Wang.: A Fast and Effective Dichotomy Based Hash Algorithm for Image Matching. LNCS, Adv. in Vis. Comp. 5358, 328-337 (2009)

[21] G. Pajares, J.M. Cruz, J.A. López-Orozco.: Relaxation labeling in stereo image matching. Pattern Recognit. 33, 53-68 (2000)

[22] M. Bleyer, M. Gelautz.: A layered stereo matching algorithm using image segmentation and global visibility constraints. J. of Photogramm. & Remote Sens. 59, 128-150 (2005)

[23] V.N. Radhika, B. Kartikeyan, G. Krishna, S. Chowdhury, P.K. Srivastava.: Robust Stereo Image Matching for Spaceborne Imagery. IEEE Trans. on Geosci. and Remote Sens. 45 (9), 2993-3000 (2007)

[24] P. Revuelta Sanz, B. Ruiz Mezcua, J.M. Sánchez Pena.: Efficient Characteristics Vector Extraction Algorithm using Auto-seeded Region-Growing. 9th IEEE/ACIS Int. Conf. on Comp. and Inf. Sci. ICIS 2010. (In Press) (2010)

[25] N. Ouerhani, A. Bur, H. Hügli.: Linear vs. Nonlinear Feature Combination for Saliency Computation: A Comparison with Human Vision. LNCS, Pattern Recognit. 4174, 314-323 (2006)

[26] Intel Corporation Motion Analysis and Object Tracking. Computer Vision Library, 2001, pp. 2-11-2-15.

[27] G. Egnal, R.P. Wildes.: Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches. IEEE Trans. on Pattern Anal. and Mach. Intell.. 24 (8), 1127-1133 (2002)

[28] C.L. Zitnick, T. Kanade.: A Cooperative Algorithm for Stereo Matching and Occlusion Detection. IEEE Trans. on Pattern Anal. and Mach. Intell. 22 (7), 675-684 (2000)

[29] S.-H. Lee, J. Yi, J.-S. Kim.: Real-Time Stereo Vision on a Reconfigurable System. Lecture Notes in Computer Science, Embed. Comp. Syst.: Archit., Model., and Simul. 3553, 299-307 (2005)

[30] S.-K. Han, M.-H. Jeong, S.-H. Woo, B.-J. You.: Architecture and Implementation of Real-Time Stereo Vision with Bilateral Background Subtraction. LNCS, Adv. Intell. Comp. Theor. and Appl. 4681, 906-912 (2007)

[31] Y.-L. Chang, X. Li.: Adaptive Image Region-Growing. IEEE Tans. on Imag. Process. 3 (6), 868-872 (1994)

[32] X.L. Wang, L.J. Wang.: Color image segmentation based on Bayesian framework and level set. Proc. of 2008 Int. Conf. on Mach. Learn. and Cybern. 1-7, 3484-3489 (2008)

[33] Y.-H. Kuan, C.-M. Kuo, N.-C. Yang.: Color-based image salient region segmentation using novel region merging strategy. IEEE Trans. on Multimed. 10 (5), 832-845 (2008)