STEREOSCOPIC DEPTH ESTIMATION USING FUZZY SEGMENT MATCHING[1]

Krzysztof Wegner, Olgierd Stankiewicz, Marek Domański

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology

ABSTRACT

Stereo matching techniques usually match segments or blocks of pixels. This paper proposes to match segments defined as fuzzy sets of pixels. The proposed matching method is applicable to various techniques of stereo matching as well as to different measures of differences between pixels. In the paper, embedment of this approach into the state-of-the-art depth estimation software is described. Obtained experimental results show that the proposed way of stereo matching increases reliability of various depth estimation techniques.

Index Terms— Stereo matching, depth estimation, soft segmentation, disparity calculation, fuzzy set

1. INTRODUCTION

Stereoscopic depth estimation is an essential research topic in 3D video technology and has made significant progress in recent years. For depth estimation, many approaches have been studied in the references [1-5]. Here, we deal with the stereo matching algorithm that searches for corresponding pixel pairs in the left- and right-eye images, where both pixels originate from projections of the same object point in the 3-D world. The disparity between the two corresponding pixels may be directly used for the depth calculations unless one of the projections is occluded. Several practical techniques have been already proposed to overcome the occlusion problems [1-3]. Often, more than two images are used in order to minimize the occlusion effects, but even in such cases, search for corresponding pixels from an image pair is a basic operation.

The stereo matching approach itself has several variants that include substantially different algorithms, like those using belief propagation [5] or graph cuts [6]. In all these techniques, the disparity-per-pixel maps are obtained by matching blocks of pixels from a pair of images. The most widely used are: rectangular blocks with constant size (e.g. 3×3 or 5×5) [4] and blocks with adaptively selected size [7]. More sophisticated approaches exploit matching of irregular blocks obtained by segmentation of images [2]. For example, it is the case of the state-of-the-art technique implemented in Depth Estimation Reference Software (DERS) [4] used for 3D video standardization activities within Moving Picture Expert Group (MPEG) affiliated by International Organization for Standardization (ISO).

In this paper, we extend our initial proposal [8] of matching fuzzy segments (that are fuzzy sets of pixels, called also soft segments). Within this paper, we are going to show that substitution of classical segments by fuzzy segments usually improves reliability of the depth estimation techniques. The proposed approach is applicable to various techniques of stereo matching as well as to different measures of differences between pixels.

2. MEASURES FOR STEREO MATCHING

For matching of hard (not fuzzy) segments, various measures of pixel differences or pixel similarity are used [9]. The pixel difference may be measured by the absolute value of the luminance difference

$$PixDiff(P_L,P_R)=\left|Y(P_L)-Y(P_R)\right|, \tag{1}$$

where P_L and P_R are pixels from the left- and right-view images, respectively. $Y(\bullet)$ is the value of luminance. Of course, this measure may be also calculated using other color components.

For the pixel difference measure defined by Eq. 1, the respective segment (block) difference measure is Sum of Absolute Differences (SAD) for luminance. Except of this widely used measure, another option is to use

$$PixDiff\ (P_L,P_R)=\left(Y(P_L)-Y(P_R)\right)^2 \tag{2}$$

that is related to Sum of Squared Differences (SSD). Both SAD and SSD are used to measure differences or similarities between segments or blocks of pixels. Let assume that the segments under comparison are $segC_L$ and $segC_R$ from the left- and right-eye images, respectively. These segments are centered on points $C_L, C_R \in \Re^2$. We are considering disparity estimation, therefore $C_L = C_R + d$, where $d = [d_H, 0]$ and d_H is a potential disparity value under checking. For both SAD and SSD, the respective general formula for segment difference may be written as:

$$SegDiff(segC_L,segC_R)=\sum_{(P_L,P_R),\ \ P_L=P_R+d} PixDiff(P_L,P_R), \tag{3}$$

[1] This work was supported by the public funds.

where the sum is calculated over all pixel pairs (P_L, P_R) such that $P_L \in segC_L \subset \Re^2$, $P_R \in segC_R \subset \Re^2$, and $P_L = P_R + d$. Such matching of segments (blocks) is shown in Fig. 1.

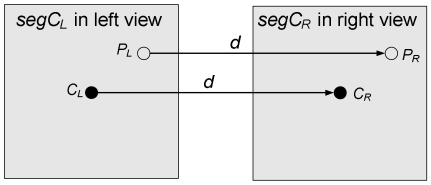

Fig. 1. Matching of segments $segC_L$ and $segC_R$ (blocks) with pixels P_L and P_R in the left and right views, respectively.

For stereo matching, application of other pixel difference measures also have been studied [9]. The matching of fuzzy segments, proposed in this paper, may be used with any of those pixel difference measures. Nevertheless, Eq. 3 should be modified for fuzzy segments, as it will be considered in the next section.

3. MAIN IDEA OF THE PAPER

The main idea of the paper consists in stereo matching of fuzzy segments. A fuzzy segment *segC* is a fuzzy set of pixels around pixel C. In practical approach discussed further in this paper, assumed is that *segC* is a rectangular set of pixels, e.g. 32×32 or 64×64. For each pixel $P \in segC$, a membership function $m(P, segC)$ is defined such that

$$0 \le m(P, segC) \le 1. \tag{4}$$

This membership function $m(P, segC)$ should meet two requirements:

- $m(P, segC)$ is decreasing, for increasing distance between pixel P and the fuzzy segment center point C;
- $m(P, segC)$ is decreasing, for increasing difference of luminance (or color) between pixel P and the fuzzy segment center point C.

Obviously, for a segment *segC*, many different membership functions $m(P, segC)$ may satisfy the above-mentioned requirements. For our implementation of stereo matching, we have proposed the following exponential formula (5):

$$m(P, segC) = e^{-\frac{|Y(C)-Y(P)|}{c_c}} \cdot e^{-\frac{|C-P|}{c_p}} , \tag{5}$$

where:

$Y(P)$, $Y(C)$ – luminance at pixel P, and at the fuzzy segment center point C, respectively,
$| C - P |$ – distance between pixel P and the fuzzy segment center point C,
c_c – color similarity coefficient,
c_p – pixel proximity coefficient.

Our experiments have revealed, that $c_c=40$ and $c_p=10$ is a good choice for the most of test data sets in use. Of course, also other membership functions may be used.

In stereo matching, values of the membership functions $m(P_L, segC_L)$ and $m(P_R, segC_R)$ are needed for segments from both the left and right image. In order to speed up stereo matching itself, the values of the membership function may be pre-calculated for individual segments. Note that the membership function depends on a single segment only.

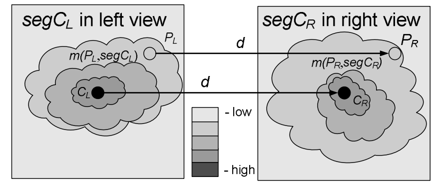

Fig. 2. Matching of fuzzy segments $segC_L$ and $segC_R$ with pixels P_L and P_R and their membership functions $m(P_L, segC_L)$ and $m(P_R, segC_R)$ in the left and right views, respectively.

Actual fuzzy segment matching is performed between the views (Fig. 2). For each pair of fuzzy segments from the two views its segment difference measure may be calculated:

$$SegDiff\ (segC_L, segC_R) = \frac{\sum\limits_{\substack{P_L,P_R \\ P_L=P_R+d}} m(P_L, segC_L)\cdot m(P_R, segC_R)\cdot PixDiff(P_L, P_R)}{\sum\limits_{\substack{P_L,P_R \\ P_L=P_R+d}} m(P_L, segC_L)\cdot m(P_R, segC_R)} \tag{6}$$

where:

$SegDiff(segC_L, segC_R)$ – difference between segments $segC_L$ and $segC_R$,
$PixDiff(P_L,P_R)$ – pixel difference metric for a pixel pair P_L and P_R.

The pixel difference metric $PixDiff(P_L,P_R)$ can be chosen freely among those known from references [9]. In particular, the metric given by Eq. 1 is often used. This metric has been used by also the authors for their experiments.

The main idea of the paper is to replace the segment difference measure from Eq. 3 by that defined by Eq. 6.

For stereo matching, this new segment difference metric may be used in a standard way. For example, for a pixel C_L from the left image, minimization of $SegDiff(segC_L, segC_R)$ on a set of pixels C_R will lead to estimation of the value d that is the disparity at C_L.

Note that the proposed fuzzy segment matching can be seen as matching using weighted windows with weights adapting to the content of images.

4. EXPERIMENTAL RESULTS

Although proposed fuzzy segment matching tool can be applied to many depth estimation algorithms, for the sake of experiment we have implemented it into two state-of-the-art depth estimation techniques. These are Depth Estimation Reference Software (DERS) [4] from Nagoya University and Depth Estimation Software version 3 (PUTv3) [5] from Poznan University of Technology, which are based on graph-cuts and belief-propagation algorithms, respectively. These techniques have been chosen because of their usage within standardization activities of ISO/IEC MPEG group and availability of their source codes for research.

We have assessed our fuzzy segment matching tool with use of two datasets: Middlebury [13] stereoscopic still images and multiview video test sequences that are used by MPEG in their standardization activities [9-12].

In the case of still images, our tool was compared with competitive Middlebury techniques [13] using ground-truth depth maps. Table 1 show *bad-pixels* for some of the Middlebury techniques and also for proposed techniques. It shows, that fuzzy-segment matching tool, implemented in PUTv3, provides depth maps of quality that is competitive to other state-of-the-art techniques. On the other hand, proposed fuzzy segment matching increases bad-pixels ratio of DERS to about 2 percent points at most. The segment size for our technique was experimentally chosen to 16×16.

Table 1. Middlebury [13] algorithms ranking.

Algorithm	Bad-pixels (Non-occ) [%]	
	Tsukuba	Cones
CoopRegion	0.87	2.79
AdaptOvrSegBP	1.69	3.60
DERS	5.02	7.22
PUTv3	4.17	4.93
DERS + fuzzy seg	4.05	4.95
PUTv3 + fuzzy seg	1.77	2.40
GC+occ	1.19	5.36

In case of 3D video, depth maps are used for view synthesis and thus it is appropriate to assess the quality of depth maps by evaluation of quality of synthesized views. Therefore, the obtained depth maps have been used for synthesis of a virtual view (Fig. 3) with use of MPEG View Synthesis Reference Software (VSRS) [4]. Finally, PSNR (for luminance) was calculated for this synthesized view with respect to its original reference view.

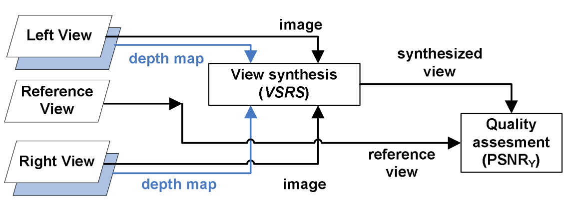

Fig. 3. Depth map quality assessment
by assessment of quality of the synthesized view.

Unfortunately, implementations and raw results of state-of-the-art algorithm are hardly available, and thus, as a reference for our tests, we could only employ before-mentioned DERS and PUTv3 techniques, with matching downgraded to simple block matching (Eq. 3). Please note, that optimization algorithms (Belief Propagation and Graph Cuts) were enabled in all cases.

For each sequence, depth maps have been estimated with DERS and PUTv3, using pixel matching and 3×3 bivalent block matching (see Eqs. 1 and 3) as well as the proposed fuzzy segment matching. The window size of 3×3 is close to optimum for matching with bivalent windows.

For proposed fuzzy segment matching, the optimum segment size has been chosen experimentally and individually for each sequence. For sequences: Book Arrival, Leaving Laptop [10] it was 64×64, for Newspaper [11] it was 32×32 and for sequences: Poznan Car Park, Poznan Street, Poznan Hall [12], it was 48×48.

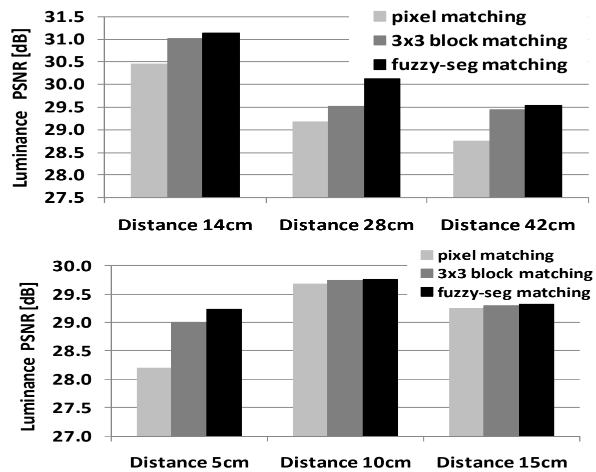

Fig. 4. View synthesis quality versus distance between cameras, for *"Poznan_Car Park"* [12] (top) and *"Newspaper"* [11] test sequences (bottom) and *DERS* software.

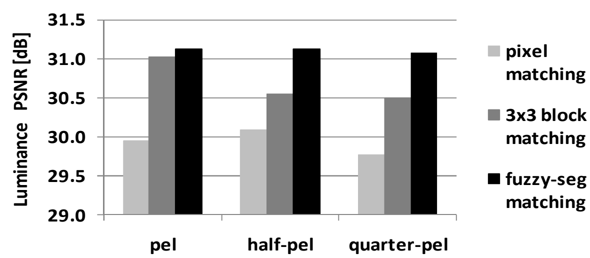

Fig. 5. View synthesis quality with various precisions of disparity maps, for "Poznan_CarPark" [12] test sequence (DERS software, distance between cameras = 14cm).

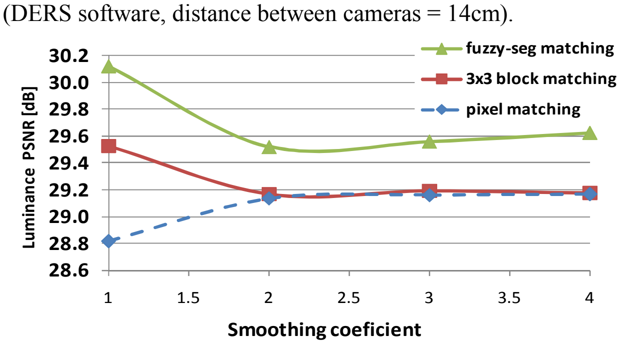

Fig. 6. View synthesis quality versus depth map smoothing coefficient for "Poznan Car Park" [12] test sequence (DERS software, distance between cameras = 14cm, pixel precision).

The comparison has been performed for various values of the depth estimation parameters, like: distance between cameras selected for depth estimation, smoothing coefficient, disparity precision. The obtained results show that the proposed fuzzy segment matching technique ("fuzzy-seg") outperforms classical approaches (pixel and

3×3 block matching) over variety of camera distances (Fig. 4) and for various accuracy of the disparity map (Fig. 5). Over wide range of smoothing coefficient values in DERS (Fig. 6), the gain is up to 0.6dB (with respect to 3×3 block matching) and from 0.1dB to about 1.0dB (with respect to pixel matching). The results prove similar improvements for both depth estimation techniques employed (Fig. 7).

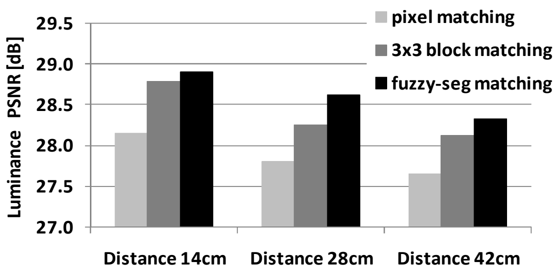

Fig. 7. View synthesis quality versus camera distance, for "Poznan_CarPark" [12] test sequence (PUTv3 software).

Table 2 summarizes our experiments on video. It shows averaged results of all test sequences used. For each sequence, the best results have been taken into account among those obtained for different values of the camera distance, the disparity map precision and the smoothing coefficient.

Table 2. Averaged results of view synthesis quality.

Matching \ Depth est.	*DERS* [4]	*PUTv3* [5]
Pixel matching	31.21 dB	28.21 dB
3×3 block matching	31.77 dB	28.98 dB
Proposed fuzzy segment matching	31.93 dB	29.43 dB

5. CONCLUSSIONS

In the paper, a new fuzzy segment matching tool have been proposed for depth estimation. This new variant of stereo matching has been already implemented in DERS software and is used by MPEG in 3D video standardization activities.

For the purpose of this paper, the new matching methodology has been tested as a tool implemented into two state-of-the art depth estimation techniques. The quality of depth estimation has been assessed using the quality of synthetic video from an intermediate point of view. The experiments yielded the conclusion that the proposed way of matching results in considerable improvements of the depth maps. The gain for still images, rated as bad-pixels ratio, is from about 1 to about 2 percent points. For video, quality of synthesized view, with respect to 3×3 bivalent block matching, ranges from about 0.1dB to 1.0dB of PSNR (luminance), and typically is of about 0.3dB. This improvements can be observed under various conditions expressed by the camera distance used for depth estimation, the precision of disparity maps and the depth estimation algorithm-specific parameters like smoothing coefficient. Nevertheless, the proposed technique can be implemented together with any image-matching based depth estimation algorithm. Despite of fact that the abovementioned improvement has been obtained at the cost of increased complexity, the fuzzy segment tool is already included in the software used in research aimed at standardization.

6. REFERENCES

[1] D. Min, K. Sohn "Cost Aggregation and Occlusion Handling With WLS in Stereo Matching", Image Processing, IEEE Transactions on Volume 17, Issue 8, pp. 1431 – 1442, 2008.

[2] M. Bleyer, M. Gelautz, "A Layered Stereo Algorithm using Image Segmentation and Global Visibility Constraints" Proceedings of the 2004 International Conference on Image Processing (ICIP 2004), Singapore, Volume 5, 2004.

[3] A. Agarwal, A. Blake, "Dense Stereo Matching over the Panum Band", IEEE Transactions Pattern Analysis Machine Intelligence, vol. 32, pp. 416-430, 2010.

[4] M. Tanimoto, T. Fujii, K. Suzuki, "Video Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching", ISO/IEC JTC1/SC29/WG11 MPEG/M16092, Lausanne, Switzerland, 2009.

[5] O. Stankiewicz, K. Wegner „Depth Map Estimation Software Version 3", ISO/IEC JTC1/SC29/WG11 MPEG/M15540, Hannover, Germany, 2008.

[6] Y. Deng, Q. Yang, X. Lin, X. Tang, "Stereo Correspondence with Occlusion Handling in a Symmetric Patch-Based Graph-Cuts Model", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 29, Issue 6, pp. 1068 – 1079, 2007.

[7] K. Wegner, O. Stankiewicz, "A Hybrid Technique for Stereoscopic Depth Estimation in Video", International Conference on Signals and Electronic Systems, Kraków, Poland, 2008.

[8] O. Stankiewicz, K. Wegner, "An Enhancement of Depth Estimation Reference Software with use of soft-segmentation", ISO/IEC JTC1/SC29/WG11 MPEG/M16757, London, UK, 2009.

[9] K. Wegner, O. Stankiewicz, "Similiarity Measures for Depth Estimation", 3DTV-Conference 2009, The True Vision Capture, Transmission and Display of 3D Video, Potsdam, Germany, 2009.

[10] I. Feldmann, M. Mueller, A. Smolic, T. Wiegand et al., „HHI Test Material for 3D Video", ISO/IEC JTC1/SC29/WG11 MPEG/M15413, Archamps, France, 2008.

[11] Yo-Sung Ho, E.-K. Lee, C. Lee, "Video Test Sequence and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG/M15419, Archamps, France, 2008.

[12] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG/M17050, Xian, China, 2009.

[13] "Middlebury Stereo Datasets" http://vision.middlebury.edu/stereo/data/.