SIMILARITY MEASURES FOR DEPTH ESTIMATION

Krzysztof Wegner, Olgierd Stankiewicz

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology,
Polanka 3, 60-965 Poznań, Poland,
email: [kwegner , ostank]@multimedia.edu.pl

ABSTRACT

This paper deals with similarity measures for stereoscopic depth estimation. These measures are used for matching of image pairs, which is the first step of the estimation process. We analyze influence of these similarity measures on performance of depth estimation with use of commonly known measures and compare the results with some novel proposals. The performance is judged by increase of quality of view synthesis, which is the main aim of this paper. Experimental results over a variety of moving material demonstrate that considerable gain can be attained without any modifications to estimation core and with tuning of matching stage only. Finally, some guidelines on design of well performing similarity measures are given. For the sake of paper, the whole work is described in context of belief-propagation algorithm, but the results and conclusions apply in general for many other state-of-the art optimization techniques.

Index Terms— depth map estimation, free-view television, belief propagation, similarity measure

1. INTRODUCTION

The modeling of 3D scenes from sets of views is an important task in many modern applications. In particular, 3D television (3DTV) and free-view television (FTV) are in the spotlight as some of the next-generation broadcasting systems providing 3D experiences [1]. In such applications, there is a need for depth information, which can be used together with color information for synthesis of virtual views. The quality of synthesized views impacts performance of such systems in at least two ways: through quality of intermediate views delivered to the user, and through quality of inter-view prediction in coding. Thus, it is desirable to have depth maps of high quality. Moreover, broadcasting systems require real-time processing, so computational complexity is also a serious factor.

One of the passive 3D depth sensing methods is stereo matching, which has been subject of extensive computer vision research during recent years. Stereo matching techniques estimate depth through computation of the disparity map between input images.

In general, modern techniques employ three steps of the estimation: direct matching of the images, optimization of the solution and post-processing.

Post-processing is employed for final refinement of the produced depth map. Depending on the application, it can be used for smoothing, denoising or increasing precision. In general, post-processing stage does not necessary improve quality with respect to fidelity, but usually is used for improvement of performance in the final application [2].

As for optimization algorithms, some of the commonly used are Belief Propagation (BP) [3] and Graph Cuts (GC) [4,5]. These techniques employ iterative processing of the depth model with use of message passing (BP) or structural modification of the graph representing possible solutions (GC). Apart from the specific features of an exact algorithm, the aim of these tools is to produce solution with respect to a global goal function. In our paper we focus on the belief-propagation algorithm (BP) because it is more regular and thus is more suitable for experimentation. Gains obtained through improvement of optimization algorithm are currently considered to be the most interesting regions of research on depth map estimation techniques [6].

The first step of the processing, direct matching of the stereo pair, is the feeding point for optimization algorithms mentioned before. Typically, it is done with use of image similarity measures like SAD or SSD (Sum of Absolute/Squared Differences). Such approaches are computationally efficient and usually are the methods of first choice but in other applications are known to be outperformed by other methods [7]. Currently there is lack of research considering selection of this measure and its influence on performance of depth estimation. Analysis of this influence is in the focus of this work.

2. BELIEF PROPAGATION FRAMEWORK

In our experimental work we are using the belief propagation (BP) algorithm. Belief propagation is

a commonly used in computer vision [9], iterative optimization algorithm for functions on a graphical model. In case of depth estimation, disparity map is modeled as a 2-dimensional Markov-field. Each point of the scene (for which disparity is sought) is represented by a single node of the field. Nodes of the mesh communicate with others by message passing mechanism (Figure 1). Each message contains information about beliefs of node, specifically beliefs about all similarity costs related with all possible disparities for considered point (Figure 2).

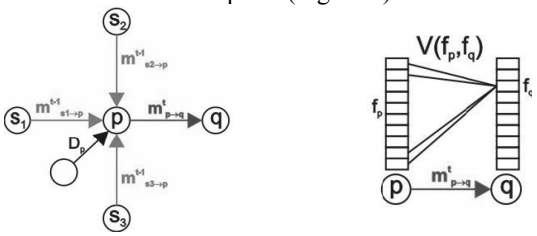


Figure 1. Message passing scheme in BP algorithm, where: $m^t_{s\to d}$ – message passed in t-th iteration from node s to node d, $V(f_p, f_q)$ – cost of belief change from disparity f_p to disparity f_q.

The beliefs attained from neighboring nodes and self beliefs are mixed together to produce new beliefs of the node for the next iteration. The process of message passing is repeated until convergence. In the end, the beliefs with the highest likelihood are chosen as a final result.

At each iteration and for each node, messages are computed with the following update equation:

$$m_{pq}(x_q) = \min_{x_p \in \kappa} \left\{ V_{pq}(x_p, x_q) + V_p(x_p) \quad + \sum_{r:\ r\neq q,\ (r,p)\in\varepsilon} m_{rp}(x_p) \right\} \qquad (1)$$

where: $V_{pq}(x_p,x_q)$ – transition cost in node q between disparity x_p and x_q insisted by node p,
$V_p(x_p)$ – observation in node p about disparity x_p,
$m_{pq}(x_q)$ – message from node p to q about disparity x_q.

Transition cost $V_p(x_p, x_q)$ is introduced to increase resultant disparity map smoothness. Its direct role is to handicap changes of the belief of the node. In this work we use so called 'Pot Model' of transition cost:

$$V_{pq}(x_p, x_q) = \begin{cases} 0 & for \quad x_p = x_q \\ \alpha & for \quad x_p \neq x_q \end{cases} \qquad (2)$$

Use of such a model provides constant discontinuity punishment α for change of node beliefs and zero otherwise.

3. SIMILARITY MEASURES

Following similarity measures are presented for single component case. In our implementation, the final measure is in fact an equally weighted sum of the measures derived for respective color components separately. This allows us to take the RGB color space into consideration.

3.1. Absolute Difference

Absolute Difference (AD) (3) similarity measure is well known and generic solution for image matching. Currently, AD is the most commonly used measure in state-of-the-art depth estimation algorithms [6,8,9]. AD measure is computed from differences between compared pixels in matched images *L(x,y)* and *R(x',y')*.

$$AD(x,y) = |L(x,y) - R(x',y')| \qquad (3)$$

In spite of computational efficiency, AD measure has some significant drawbacks: it is not discriminative for bad solutions and is vulnerable to noise. Because only single points are matched, this measure does not provide any information about texture of matched objects.

3.2. GRADIENT

We propose the following GRADIENT similarity measure for comparison of images:

$$GRADIENT(x,y) = \left|\frac{\partial}{\partial x}L(x,y) - \frac{\partial}{\partial x}R(x',y')\right| + \left|\frac{\partial}{\partial y}L(x,y) - \frac{\partial}{\partial y}R(x',y')\right| \qquad (4)$$

The measure is constructed from sum of absolute differences between horizontal and vertical gradients of matched images. In out implementation, the gradients are computed with use of Sobel operator.

GRADIENT similarity measure is designed to convey information about edges. This allows for improved depth estimation around objects boundaries.

3.3. RANK

RANK measure exploits non-parametric RANK transform [7]. The results of RANK transform *RT{P}* (computed at each point of the input image) is number of neighboring points (neighborhood *N(P)*), which have lesser value *I(P')* than value *I(P)* of currently processed point *P* (5):

$$RT\{P\} = \left\|\{P' \in N(P) \mid I(P') < I(P)\}\right\| \qquad (5)$$

We formulate RANK similarity measure as absolute difference between RANK transforms of two compared images:

$$RANK(x,y) = |RT\{L\}(x,y) - RT\{R\}(x',y')| \qquad (6)$$

The main feature of RANK similarity measure is its robustness and immunity to noise. Because outcome of RANK is constructed from neighboring pixels, it is depended on the texture of matched objects.

3.4. CENSUS

CENSUS measure exploits non-parametric CENSUS image transform [7] which, similarly to RANK transform, considers neighborhood of processed point. The results of

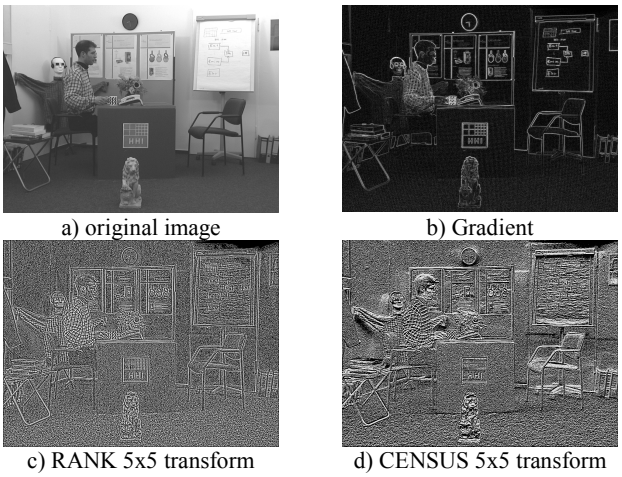
CENSUS transform (computed at each point of the input image) is a map of bits gathered in word *CT{P}*. Each bit in position *k* is a flag indicating whether neighboring point P_k has lesser value $I(P_k')$ than value *I(P)* of currently processed point *P*:

$$CT\{P\}[k] = \{P_k' \in N(P) \mid I(P_k') < I(P)\} \quad (7)$$

We formulate CENSUS similarity measure as *hamming* distance between CENSUS transforms of two compared images:

$$CENSUS(x,y) = hamming(CT\{L\}(x,y), CT\{R\}(x',y')) \quad (8)$$

The main advantage of CENSUS similarity measure over RANK is that it takes spatial distribution of texture into consideration. Likewise to RANK, CENSUS is immune to noise, but is slightly more computationally complex.



a) original image b) Gradient

c) RANK 5x5 transform d) CENSUS 5x5 transform

Figure 2. Examples of transforms used for similarity computation.

3.5. Mixtures of similarity measures

Stand-alone use of basic measures presented above is not efficient (Figure 4, Results) and can even bring decrease of synthesis quality. Thereby, we propose some novel similarity measures, which are mixtures (products) of these measures. In particular we propose following formulas:

- $AD(x,y) \cdot RANK(x,y)$
- $AD(x,y) \cdot GRADIENT(x,y)$
- $GRADIENT(x,y) \cdot RANK(x,y)$
- $AD(x,y) \cdot CENSUS(x,y)$
- $GRADIENT(x,y) \cdot CENSUS(x,y)$

Because the basic measures are multiplied together (modulated), there is no need for weighting, which would require optimization of the weights (e.g. in case of weighted sums). Our previous work has shown that such optimal weights are highly dependent on content of the scene.

The motivation behind use of mixtures is to enrich monotonic character of AD (Figure 5) with discriminative and noise robustness features of RANK and CENSUS transforms and with edge sensitivity of GRADIENT measure. We suppose that these three matching components are essential for good performance of the matching process.

4. EXPERIMENTS

Evaluation of depth map quality is not a straightforward task. First of all, in general, the references are not available. There are attempts to build image database [6] with ground-truth depth maps, but still these depth-maps are produced with off-line techniques (like Structured Lighting [10]) which have different specifics. This makes comparison difficult and inadequate. Secondly, the expectations for depth maps vary between applications and thus depth maps cannot be judged regardless of the final application. In this paper, we focus on 3D television in which depth maps are used for the view synthesis purposes. Therefore, we use quality evaluation based on the view synthesis (Figure 3). Such method has been successfully used by MPEG-FTV experts [1]. The experiments were performed with use of depth estimation software [2] and view synthesis software [11] as follows:

- depth corresponding to neighboring original views NL (left) and NR (right) are estimated,
- views SL and SR (at positions of OL and OR) are synthesized from NL+depth and NR+depth,
- newly synthesized views SL-SR are compared with OL-OR pair - subjectively and by PSNR.

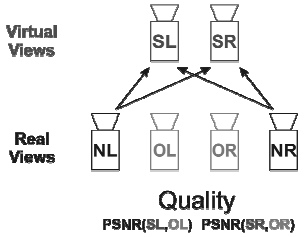


Figure 3. Setup of experiments for depth-estimation and view-synthesis similarity measure evaluation.

For testing purposes we have used a set of 5 color multi-view video test sequences, containing different scenes, kindly provided by [12,13,14] for scientific purposes.

5. RESULTS

Figure 4 shows that use of GRADIENT, RANK and CENSUS similarity measures in stand-alone mode does not bring any gain. The results are worse from about 1 dB to over 4 dB than for AD measure reference. This yields from fact, that none of these basic measures convey information about three principal components for matching. GRADIENT measure is discriminative only for edges, while RANK and CENSUS measures are suitable only for highly textured regions.

However, standalone use of the basic similarity measures is deprecated, it can be noticed that use of proposed novel mixtures bring considerable gain up to 1.2 dB.

The best results are attained with use of $AD(x,y)\cdot CENSUS(x,y)$ and $AD(x,y)\cdot GRADIENT(x,y)$. These measures contain information about textures, edges and lightness in the neighborhood of compared pixels.

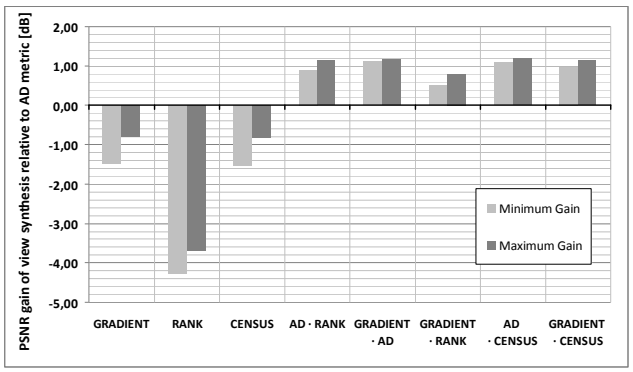


Figure 4. Gains in view synthesis attained with use of specific similarity measures used in depth estimation with respect to AD.

The reason behind such results can be explained with use of Figure 5. The winning similarity measures are those that are highly discriminative (for wrong disparity values) and have very narrow minimum around the correct disparity value.

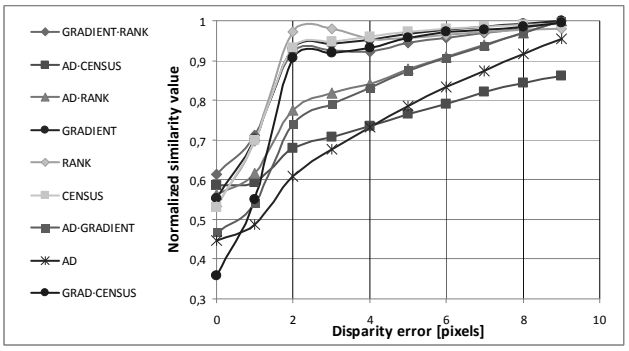


Figure 5. Graphs of similarity values for disparities around optimal disparity value (an average over all pixels in all sequences).

6. SUMMARY

In this paper, we have analyzed influence of similarity measure selection on depth estimation and view synthesis process. Use of novel mixtures has been proposed and it has been shown that some of them bring considerable gain over commonly used AD similarity measure. Also, some general concussions about similarity measures for depth estimation have been given. It has been shown that design of such measures must take various factors into account, like noise vulnerability, edge localization and texture sensitivity.

7. ACKNOWLEDGEMENT

This work was supported by the public funds as a research project in years 2007-2009.

11. REFERENCES

[1] A. Smolic, K. Müller, P. Kauff, T. Wiegand, “Considerations about the Vision of a 3D Video Standard”, ISO/IEC JTC1/SC29/WG11, MPEG 2008/M16130, Lausanne, Switzerland, February 2009.

[2] O. Stankiewicz, K. Wegner "Depth Map Estimation Software version 3", MPEG 2008/M15540, Hannover, Germany, July 2008.

[3] A. Klaus, M. Sormann, K. Karner, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, International Conference on Pattern Recognition 2006.

[4] J.S. Yedidia, W.T. Freeman, Y. Weiss, “Understanding Belief Propagation and Its Generalizations”, Exploring Artificial Intelligence in the New Millennium, ISBN 1558608117, Chap. 8, pp. 239-236, January 2003 (Science & Technology Books).

[5] http://www.tanimoto.nuee.nagoya-u.ac.jp/ - MPEG-FTV web-page, Tanimoto Laboratory, Nagoya University.

[6] “Middlebury Webpage”, http://vision.middlebury.edu/stereo/.

[7] R. Zabih, J. Woodfill, "Non-parametric Local Transforms for Computing Visual Correspondence", European Conference on Computer Vision, Stockholm, Sweden, May 1994, pages 151-158.

[8] D. Scharstein, R. Szeliski, „A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”, International Journal of Computer Vision 2002.

[9] J. Sun, N.N. Zheng, H.Y. Shum, “Stereo matching using belief propagation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7):787–800, 2003.

[10] D. Scharstein, R. Szeliski, „High-accuracy stereo depth maps using structured light”, Computer Vision and Pattern Recognition 2003, volume 1, pages 195-202, June 2003.

[11] M. Gotfryd, K. Wegner, M. Domański "View synthesis software and assessment of its performance", ISO/IEC JTC1/SC29/WG11, MPEG 2008/M15672, Hannover, Germany, July 2008.

[12] I. Feldmann, M. Müller, F. Zilly, R. Tanger, A. Smolic, P. Kauff, T. Wiegand, „HHI Test Material for 3D Video”, ISO/IEC JTC1/SC29/WG11, MPEG 2008/M15413, Archamps, France, April 2008.

[13] Yo-Sung Ho, Eun-Kyung Lee, and Cheon Lee “Multiview Video Test Sequence and Camera Parameters”, ISO/IEC JTC1/SC29/WG11, MPEG 2008/M15419, Archamps, France, April 2008.

[14] M. Tanimoto, T. Fujii, N. Fukushima, “1D Parallel Test Sequences for MPEG-FTV”, MPEG 2008/M15378, Archamps, France, April 2008.