Hybrid Approach to Depth Estimation

Olgierd STANKIEWICZ,
Chair of Multimedia Telecommunications
and Microelectronics
Poznan University of Technology,
Polanka 3, 69-965 Poznań, Poland,
email: ostank@multimedia.edu.pl

Krzysztof WEGNER,
Chair of Multimedia Telecommunications
and Microelectronics
Poznan University of Technology,
Polanka 3, 69-965 Poznań, Poland,
email: kwegner@multimedia.edu.pl

***Abstract*—In this paper, we propose a new hybrid technique for the disparity map estimation from a stereo pair of images. The original concepts of our proposal are: use of dual-path disparity estimation, a new hierarchical shape-adaptive block matching. optical flow extended with novel disparity compensation step and improved iteration scheme, novel occlusion detection and disparity extrapolation schemes. The main advantages of proposal are: low computational complexity, sub-pixel accuracy of estimation, propagation of depth information across flat or untextured regions and good detection of occluded regions.**

I. Introduction

The 3D reconstruction of video scenes is an important task in many multimedia applications, including free-view television (FTV), depth-based segmentation and surveillance systems. Most of these applications make use of auxiliary depth information, commonly achieved by depth map estimation. Desired depth maps can be computed from disparity map [10].

Most of known techniques for the disparity map estimation (e.g. [4,5,6,7]) are only pixel accurate and require large amounts of memory and computational power. For example, complexity of state-of-the-art Belief Propagation based techniques [4] is proportional to image dimensions and number of disparity levels which is $O(N^3)$. Moreover, commonly used classical block-matching technique fails to extract disparity information from flat, weakly textured regions.

In this paper, we propose a hybrid technique that aims at computationally and memory efficient sub-pixel accurate disparity estimation.

II. Algorithm Overview

We assume processing of a rectified [11] stereo video pair, which comes from linearly positioned camera array. Such an arrangement is typical for currently developed systems [10]. The original idea of our proposal (Fig. 1) comprises:

- novel dual-path simultaneous disparity estimation, that allows for exchange of data between the views,
- use of hybrid approach exploiting block matching algorithm and sub-pixel refinement by optical flow,
- a new hierarchical shape-adaptive block matching algorithm,
- a novel occlusion detection method based on preliminary results from block matching algorithm from another path,
- a disparity compensation step and improved iterative scheme for optical flow algorithm,
- a new disparity refinement technique taking advantage of available information from another path.

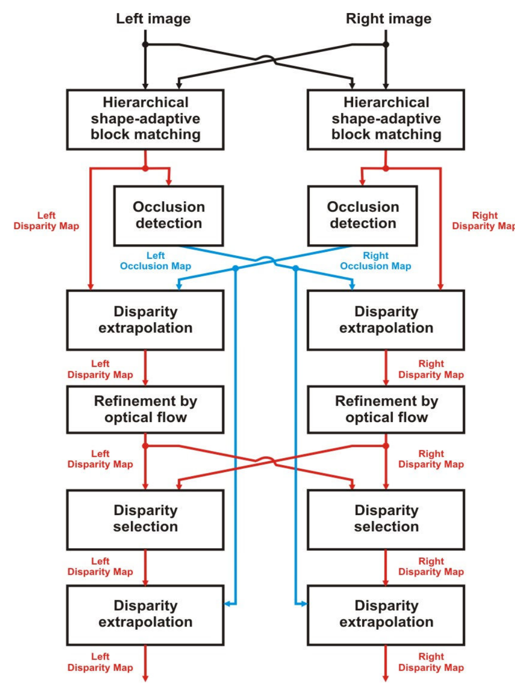

Figure 1 – Block diagram of the algorithm proposed

The block matching technique provides the initial guess of the disparity map. To overcome drawback of classical block matching, we propose a hierarchical algorithm that starts with low resolution images and gradually traverses towards full resolution. Hierarchical approach not only allows for estimation of disparity map based on large area of image content but also reduces complexity of matching step. Our algorithm is also shape-adaptive with respect to block size and block positioning.

Disparity values obtained by block matching are quantized with respect to a given pixel grid. Improvement of resolution would require drastic computational costs. To avoid that, we propose that the next step is refinement by optical flow, which iteratively improves quality and accuracy of resultant disparity map. In this paper we use gradient-based approach to optical flow (similar to [1]) and introduce some improvements specific to disparity map estimation. These improvements reduce computational complexity, improve reliability of the iterative scheme and also impose some constraints on resultant disparity map.

III. Description of the Algorithm

A. Hierarchical shape-adaptive block matching

Aim of block matching algorithm in our proposal is to deliver initial guess of disparity map for further processing. We propose a new hierarchical approach to this task. Hierarchical block matching iteratively estimates disparity map starting from lowest resolution and improving accuracy in subsequent steps. Moreover size of the blocks employed by our technique is adaptively selected.

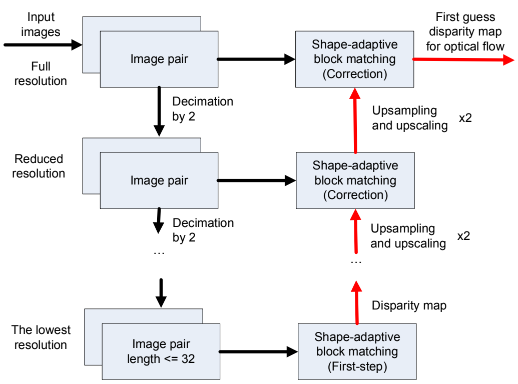

Fig. 2. Block diagram of hierarchical block matching

The algorithm starts with successive decimation of original image pair (in our implementation the size of the smallest image was 32×24 pixels). Block matching algorithm (Fig. 2) starts with the lowest-resolution image pair. In the first step, the disparity is estimated in the full range implied by horizontal resolution of decimated image.

In consecutive iterations image resolution is doubled (less decimated images are used) and disparity map obtained from previous iteration is up-sampled and up-scaled. Disparity map is then only updated: search is limited to range, which depends on existence of edge in disparity map obtained from previous level. If there is an edge, disparity is examined: from disparity value on the left side of the edge to disparity value on the right side. In flat regions only small neighborhood of current disparity is considered.

Estimation of disparity for each pixel in image is done independently. In basic block matching, block of pixels in small neighborhood centered at current pixel is matched with pixels from second image (Fig.3. Upper pair). We introduce a method for adaptively selecting shape of matched blocks. In our proposal, currently analyzed point is not necessarily the center of matching window and can occupy any location inside the rectangular matching window (Fig.3 Lower pair). Location inside the window is chosen by lowest NSAD (Normalized Sum of Absolute Differences) criterion [12]. Such an approach is motivated by occurrence of mismatch regions, which may result from occluded regions.

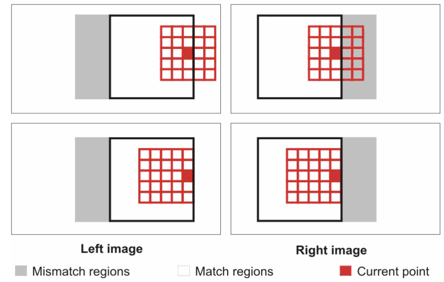

Fig. 3. Adaptive hook-point selection for exemplary 5×5 block size and its behavior in case of mismatch region (grey) and match region (white). Matching selects lower pair instead of upper

Size of the block may also vary. For each disparity value, matching starts with block of large size (in our experiments: 9×9 pixels) and is reduced until NSAD value reaches minimum. Reduction of block size stops when any of following conditions turns out to be true: either when 3×3 block is reached or when NSAD value goes below arbitrarily chosen threshold. This condition is same as “match is good enough”.

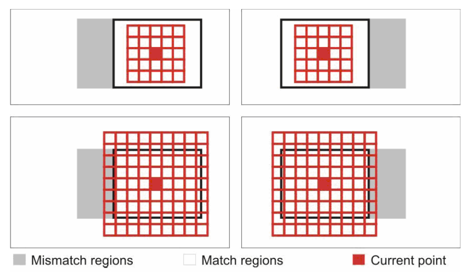

Fig. 4. Adaptive selection of block size between exemplary 5×5 and 9×9 in context of its behavior over small objects (white)

Varying size of block helps in the case of occurrence of small details in images (where small block size is chosen) (Fig. 4) while not limits scope of matching in other cases (where it is best to perform matching in blocks as big as possible).

B. Occlusion detection

Not all pixels in the left image have corresponding pixels in the right image (and vice versa) because of occlusion (Fig. 5). Occluded regions containing artifacts should be detected in order to cancel erroneous values. For this task we introduce a new technique described below.

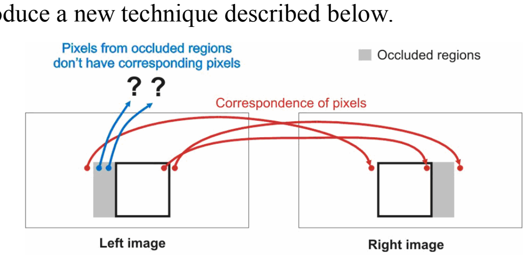

Fig. 5. Correspondence between pixels in the left and the right image with respect to occluded regions

Our experiments revealed that straight-forward use of NSAD (as pixels in occluded regions mismatch and imply relatively big values of NSAD) it is not efficient, because it is uncertain whether big NSAD values refer to occluded regions or result from existence of noise.

In our proposal, occlusion is detected basing on "Pixel Use Count". "PUC" map counts how many times given pixel have been matched as corresponding one to pixel in second image by the block matching process. If pixel has never been used, we estimate that it is occluded. Such an approach gives occlusion information in crossed scheme shown in Fig. 1: left path generates occlusion information for right disparity map and so on.

C. Disparity extrapolation

After corrupted disparity values lying in occluded regions are identified in occlusion detection phase, they have to be replaced. We assume that only background can be occluded (object closer cover objects that are further) and that the nearest background disparity is a good estimation for those regions. This nearest background disparity can be found in the neighborhood next to the left of considered pixel in left image disparity map, and in neighborhood next to the right of considered pixel in right image disparity map. Our algorithm exploits sliding window (in our implementation of size 20×1 pixels) hooked at considered pixel (Fig. 6). Average of non occluded disparities in this window replaces disparity value at the considered occluded point.

Disparity extrapolation is exploited twice in each path of our algorithm (Fig. 1): before refinement by optical flow and as a final stage. First use is necessary to minimize confusion of gradient computation. It is crucial for good optical flow stage operation. The second use employs extended scheme of extrapolation with disparity selector described in point 3.5. Unfortunately, the second extended scheme is not well suited for operation before optical flow stage because the disparity map at that time is too messy which would that would introduce major artifacts at selector block.

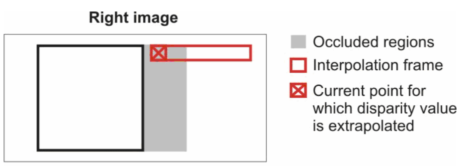

Fig. 6. Disparity extrapolation frame in occluded regions for exemplary right image

D. Refinement by optical flow

We employ technique based on gradient, iterative optical flow algorithm [1], however we introduce two novel improvements:

I. Disparity compensation - in our technique pixels of right image R(x,y) are matched with pixels from the left image L(x',y') using compensated coordinates x',y', which are real and in particular might not be integer.

II. New interative scheme, for calculation of disparity correction, which is derived from linear approximation of images and horizontal and inter-image gradients (1).

$$d_i(x,y) \leftarrow d_{i-1}(x,y) - \beta \cdot g_t(x,y) \cdot \frac{g_x(x,y)}{g_x(x,y)^2 + \alpha} \quad , (1)$$

where:

$d_i(x,y)$ – disparity value for coordinates x,y, at i-th iteration,
$g_x(x,y)$ - horizontal gradient for compensated coordinates x,y,
$g_y(x,y)$ - inter-picture gradient for compensated coordinates x,y,
α - noise uncertainty coefficient (for experiment α =5),
β- iteration step size (for experiment β=0.5).

Note that if $\alpha = 0$, Equation (1) yields (2):

$$d_i(x,y) \leftarrow d_{i-1}(x,y) - \beta \cdot g_t(x,y) \cdot \frac{1}{g_x(x,y)} \quad , (2)$$

Non-zero α coefficient was introduced to cancel disparity corrections over regions with small gradients that are assumed to have relatively high noise level. This allows for better performance over untextured regions.

Finally, negative disparity values are reduced by half. The reason behind it is to eliminate values impossible in real-world scenes (but attainable by optical flow algorithm) in disparity map, but not too strictly to keep the iterative process fluent. If disparity values are negative it means that images are swapped.

E. Disparity selection

After the optical flow refinement stage the disparity map is smooth, devoid of major artifacts but overtly continuous at depth edges. Moreover, disparities at previously occluded regions are only initial and weak estimates – those occluded regions occur at left edges of objects in the left path of processing and at right edges of objects in the right path. One can notice that opposite edges are of better quality (sharper). We propose that disparity information is exchanged between processing path (Fig. 1) and better estimation is chosen according to NSAD criterion. It is possible because NSAD values are propagated from block matching module. Resultant disparity map goes to another disparity extrapolation block, where pixels occluded in both paths are extrapolated.

IV. Experiment

The algorithm has been examined using commonly known test images and video sequences. The results were assessed mainly basing on subjective quality but also on objective measures:

- 'Bad Pixel count' [9] – which measure percentage of erroneous pixels, classified with respect to ground-truth disparity map and arbitrary difference threshold,

- NBP-SAD and NBP-SSD [12] (Normalized Bad Pixel – SAD/SSD) - which measure normalized magnitude / energy of errors with respect to ground-truth disparity map,

- PSNR of resynthesized view [12] – express quality of disparity map as error of resynthesized image with comparison to original view.

V. Results

Table 1 shows results attained by our technique compared with quality measures mentioned above. Among other techniques, out technique has the worst bad-pixel ratio, but the best NBP-SAD value. It means that our technique produces many bad-pixels, but with relatively small errors versus ground-truth.

The values of PSNR and NBP-SSD show that our technique is competitive to other algorithms and referring to them, places at about the middle of the ranking.

venus	Bad Pixels	NBP-SAD	NBP-SSD	PSNR [dB]
AdaptOvrSegBP [7]	0,13%	2,32	5,72	32,84
PlaneFitBP [8]	0,10%	2,26	5,88	34,45
DoubleBP [5]	0,15%	2,37	6,01	34,68
SubPixDoubleBP [6]	0,15%	2,41	6,15	32,04
Our proposal	**5,95%**	**2,21**	**6,31**	**27,51**
AdaptingBP [4]	0,13%	2,65	7,15	31,31
SSD +MF [3]	3,63%	3,14	10,82	34,51
Ground truth [2]	0,00%	-	-	31,46

cones	Bad Pixels	NBP-SAD	NBP-SSD	PSNR [dB]
AdaptingBP [4]	2,30%	2,19	5,94	29,01
AdaptOvrSegBP [7]	3,42%	2,25	6,74	30,14
PlaneFitBP [8]	restrict4sports4h	2,77	9,67	31,42
DoubleBP [5]	3,31%	3,31	15,74	32,51
SubPixDoubleBP [6]	3,20%	3,32	15,97	30,74
SSD +MF [3]	8,64%	3,12	17,07	30,02
Our proposal	**17,44%**	**3,61**	**24,07**	**28,63**
Ground truth [2]	0,00%	-	-	31,19

tsukuba	Bad Pixels	NBP-SAD	NBP-SSD	PSNR [dB]
Our proposal	**8,01%**	**2,50**	**9,28**	**29,96**
AdaptOvrSegBP [7]	1,58%	4,16	24,95	31,51
SubPixDoubleBP [6]	1,22%	4,10	25,46	29,81
DoubleBP [5]	0,89%	5,15	33,75	35,32
SSD +MF [3]	5,25%	5,28	34,78	33,19
AdaptingBP [4]	1,10%	5,25	35,71	29,62
PlaneFitBP [8]	0,96%	5,38	36,61	34,87
Ground truth [2]	0,00%	-	-	35,31

Table 1. Comparison of algorithm performance; sorted by NBP-SSD – normalized energy of errors

Fig. 7 shows that proposed algorithm is capable to re-veal background information. Although standard output for-mats for disparity representation do not allow fractional precision, the output of the algorithm is sub-pixel accurate without any additional computational or memory costs

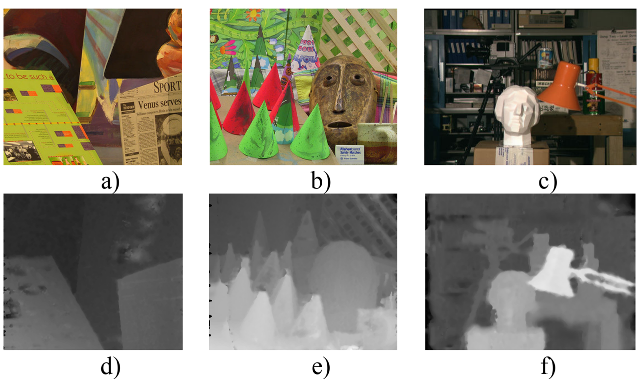

Fig. 7. Results of experiment performed on “Venus” scene (a,d,g), “Cones” scene (b,e,h) and “Tsukuba” [9] (c,f,i). Original left-view images (a,b,c), disparity maps obtained with algorithm proposed (d,e,f)

VI. Conclusions

We have proposed an original technique for depth map estimation that exploits parallel dual-path processing for the left and the right view. We also have proposed a new hierarchical shape-adaptive block matching that reduces computational cost relative to classical full search block-matching algorithm. Improvements to optical flow algorithm have been introduced – disparity compensation step, which helps to avoid local-minima problem and new iterative scheme which allows for better performance over untexturred regions. Also a new occlusion detection algorithm has been introduced which provides good accuracy basing on information coming from complementary path. Also some simple but efficient disparity extrapolation schemes have been proposed. Resultant disparity map coming from our algorithm can be used to produce depth map if exact camera parameters and locations are known [10].

One of the biggest advantages of proposed approach is sub-pixel accuracy supported by gradient-based core of optical flow. Thanks to that our algorithm provides good background estimation which is important for distant scenes.

Another advantage of presented technique is that it estimates disparity across flat and untextured regions. Iterative nature of our new optical flow core allows propagation of information about depth across these regions. The proposed algorithm works efficiently even in absence of specific features (sharp edges, corners etc.) but exploits information that they carry.

One of the main drawbacks of the algorithm is computational complexity which is related to large number of iterations required for convergence. Fortunately it is has limited dependence on resolution of processed image pair. It is also noticeable, that edges between objects in disparity maps are not preserved very well. It is caused mainly by filtering step, responsible for smoothness constraint [1].

Although currently our experiments has been per-formed only for generation of disparity maps from stereo pairs and had not exploited information from multiview images, it is worth to notice that out technique may be easily extended to support extraction of depth maps from more than two views. It will be task for our future work.

To summarize, in our approach:

- low-accuracy estimation provided by the block-matching step is enhanced by estimation based on the optical flow technique,
- iterative nature of optical flow technique allows propagation of depth information across flat or untextured regions,
- the usage of the initial guess overrides the local-minima problem encountered by the estimation based only on the pure optical flow
- the number of iterations typically needed for the optical flow technique is reduced due to good initial guess.

References

[1] B.K.P. Horn, B.G. Schunck. “*Determining optical flow*”. Massachusetts Institue of Technology, 1980.

[2] D. Scharstein and R. Szeliski. “*High-accuracy stereo depth maps using structured light*”. IEEE CVPR 2003, vol-ume 1, pages 195-202, Madison, WI, June 2003.

[3] D. Scharstein, R. Szeliski. „*A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*”, IJCV 2002.

[4] A. Klaus, M. Sormann, K. Karner. „*Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*”, ICPR 2006.

[5] Q. Yáng, L. Wang, R. Yang, et al. „*Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling*”. CVPR 2006.

[6] Q. Yang, R. Yang, J. Davis, D. Nistér. “*Spatial-depth super resolution for range images*”. CVPR 2007.

[7] Anonymous, “*Stereo reconstruction with mixed pixels using adaptive over-segmentation*”, CVPR 2008.

[8] Anonymous. “*Near real-time stereo for weakly-textured scenes*”. CVPR 2008.

[9] “*Middlebury Stereo Datasets*” http://vision.middlebury.edu/stereo/data/.

[10] “*Description of Exploration Experiments in 3D Video*”, MPEG 2008/N9596, Antalya, Turkey, January 2008.

[11] M. Pollefeys, R. Koch, L. Van Gool, “*A simple and efficient rectification method for general motion*”, Proc. International Conference on Computer Vision 1999, pp. 496-501.

[12] O. Stankiewicz, K. Wegner, “*Depth Map Estimation Software version 2*”, MPEG 2008/M15338, Archamps, France, April 2008.