INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG/M15175
Antalya, January 2008

Title **Depth Map Estimation Software**
Sub group **Video**
Authors **Olgierd Stankiewicz** (ostank@multimedia.edu.pl) **and Krzysztof Wegner,**
Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

This document is in response to N9468 "Call for Contributions on FTV Test Material" [6], in particular for "Depth Map Estimation & View Synthesis Software" paragraph.

1 Introduction

We propose hybrid approach for the depth map estimation problem. Our solution exploits modified optical-flow algorithm as the main iterative computation core and hierarchical shape-adaptive block matching for the first guess of disparity map. This approach overcomes drawbacks of the traditional block matching technique and basic optical flow.

In our approach:

- computational power required by classical block-matching is reduced
- low-accuracy estimation provided by the block-matching step is enhanced by estimation based on the optical flow technique,
- 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
- iterative nature of optical flow technique allows propagation of depth information across flat or untextured regions

Although currently our software can be used only for generation of disparity maps from stereo pairs and cannot exploit information from multi-view video sequences, it is worth to notice that both components may be easily extended to support extraction of depth maps from more than two views. Block matching can be extended by simple modification of SAD-based matching scheme that reflects relative positions of cameras. SAD computation must respect pixel values in all images and disparity changes over views. Optical flow can be upgraded by extension of gradient computation equations.

Another advantage of proposal is that it can probably be effortlessly implemented in hardware. Both block-matching and optical-flow techniques are known to have efficient implementations.

Resultant disparity map can be used to produce depth map if exact camera locations are known.

2 Description of the algorithm

2.1 Overview

Our hybrid approach consists of two main components - optical-flow and block matching (Fig.1).

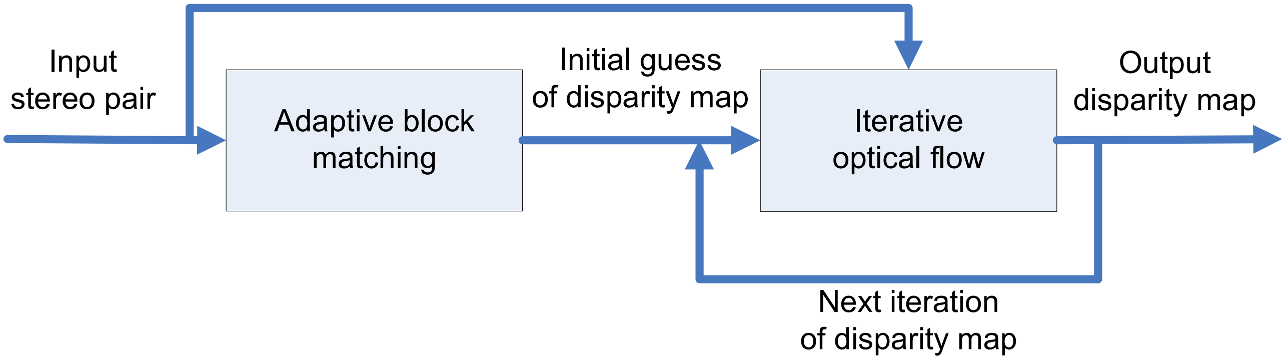


Figure1. Block diagram of the algorithm proposed

The block matching technique provides the initial guess of the depth map. Unfortunately, classical block-matching omits non-local information and fails to extract depth information from flat, untextured regions. Moreover, disparity values are quantified and only pixel-accurate. That is why we propose hierarchical algorithm that starts with low-resolution images and gradually traverses toward full-resolution. This not only allows for more-global estimation of depth map but also reduces complexity of matching step. Out algorithm is also shape-adaptive with respect to block size and block positioning.

The next step is optical flow that starts with disparities coming from block matching, and iteratively improves quality and accuracy of resultant depth map. We decided to use classical gradient-based optical flow and introduce some improvements specific to depth map estimation. These improvements reduce computational complexity, improve reliability of the iteration scheme and also impose some constraints on resultant disparity map.

In fact, both steps are performed in hierarchical mode as described below.

2.2 Hierarchical shape-adaptive block matching

Block matching stages deliver initial guess of disparity map for optical flow. It starts with low-resolution image pair which comes from decimation of original image pair. Decimation process (by factor of 2 at each step) proceeds until longer dimension of the image falls below 32 pixels. Reduction of image resolution employs simple averaging for low-pass filtration of visual content.

After the decimation is done, block matching algorithm starts. Pixels surrounding considered point in reference image, for which disparity is being computed, are compared with corresponding pixels from second image. Disparity value for which SAD (Sum of Absolute Differences) is the least is chosen as output.

In first step, whole range of possible disparities is considered. Restoration of image pair resolution at each step goes with interpolation of disparity map. A correction is also added with respect to possible inaccuracy from previous step (Fig. 2).

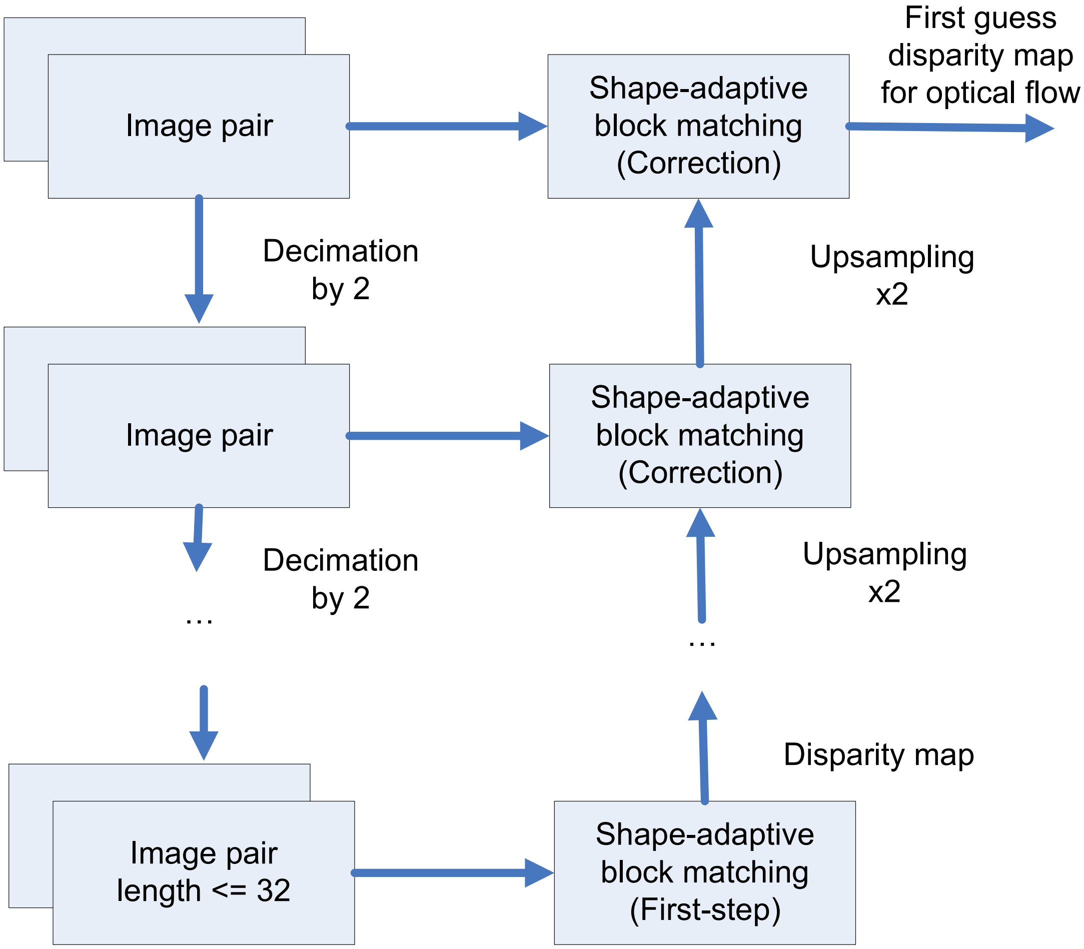


Image pair
Shape-adaptive block matching (Correction)
First guess disparity map for optical flow
Decimation by 2
Upsampling x2
Image pair
Shape-adaptive block matching (Correction)
Decimation by 2
Upsampling x2
…
…
Disparity map
Image pair length <= 32
Shape-adaptive block matching (First-step)

Figure2. Block diagram of hierarchical block matching

Hierarchical approach leads to reduction of computational complexity but also allows for extraction of disparity basing on higher-order textural content. This gain is utilized for extension of block-matching task. Specifically, all block positions relative to the hook-point are considered, and the best one is chosen. Classically, only block center is used as the hook-point. Such an approach is motivated by occurrence of object edges which introduce misfit (Fig. 3).

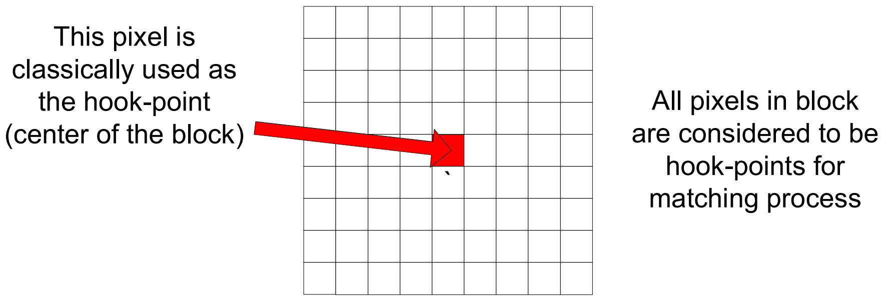

This pixel is classically used as the hook-point (center of the block)
All pixels in block are considered to be hook-points for matching process

Figure 3. Adaptive hook-point selection for exemplary 9x9 block size

Another point about ‘shape-adaptiveness’ of proposed block-matching technique is that is references size of block considered. For each disparity value, matching starts with block of size 9x9 pixels and is reduced until normalized SAD value reaches minimum. Reduction of block size stops when any of following conditions turns out to be true: either when 3x3 block is reached – because matching of single pixels (1x1 block) introduces extraordinary noise into disparity map, or when normalized SAD value goes below 1 – because it turned out that this condition is homogenous with “match is good enough”. Both of these help to avoid erroneous block matches.

2.3 Optical-flow

We employ algorithm based on standard gradient, iterative optical flow, similar to [1]. Because we assume that image pair is rectified [2], disparities are only one-dimensional vectors, and thus optical-flow scheme can be simplified. Because vertical component of disparity vector is assumed to be zero, it does not have to be computed.

Computation at each iteration is derived from linear approximation based on gradients existing in corresponding points in stereo pair. These gradients – horizontal gradient and inter-picture gradient are used for calculation of disparity correction.

The main improvement that we propose is that current disparity value compensates positions of corresponding pixels in stereo pair. Pixels of right image $R(y,x)$ are matched with pixels from left image $L(y',x')$ using compensated coordinates y',x', which are real and in particular might not be integer.

Each iteration consists o following steps:

1. Disparity image is filtered to achieve expected level of smoothness. The mask of the FIR filter is presented in the Fig. 4.

$$d_{filtered}(y,x) \leftarrow \frac{1}{12}\Big[\ 2\cdot d_i(y-1,x)+2\cdot d_i(y+1,x)+2\cdot d_i(y,x-1)+2\cdot d_i(y,x+1)$$
$$d_i(y-1,x-1)+d_i(y-1,x+1)+d_i(y+1,x-1)+d_i(y+1,x+1)\ \Big],$$

where:

$d_i(y,x)$ - disparity value at i-th iteration

$d_{filtered}(y,x)$ - filtered disparity value at coordinates y,x

$$\frac{1}{12}\begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 0 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

Figure 4. Mask of low-pass FIR filter used for filtration of disparity map

2. Pixel coordinates are compensated with respect to current disparity vector. Thanks to that, pixels $R(y,x)$ and $L(y',x')$ correspond to each other.

   $x' = x + d_x(y,x)$

   $y' = y$

3. For each pixel horizontal and inter-image gradients are computed

   $$g_x(y,x) = \frac{1}{4\Delta}\left(R(y,x+\Delta) - R(y,x-\Delta) + L(y',x'+\Delta) - L(y',x'-\Delta)\right)$$

   $$g_t(y,x) = L(y',x') - R(y,x)$$

   where:

   Δ - gradient approximation step, for experiment it was set to 0.01

   $g_x(y,x)$ - horizontal gradient for coordinates y,x (in right picture)

   $g_t(y,x)$ - inter-picture gradient for coordinates y,x (in right picture)

4. Disparity correction:

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

   where:

   α - noise uncertainty factor; for experiment it was set to 5

   β - iteration step size; for experiment it was set to 0.5

   Note that if α =0, equation yields to

   $$d_{i+1}(y,x) \leftarrow d_{filtered}(y,x) - \beta \cdot g_t(y,x) \cdot \frac{1}{g_x(y,x)},$$

   which represents first order gradient-based linear equation solving iteration scheme. Non-zero α factor was introduced to cancel disparity corrections on small gradients that are assumed to have relatively high noise level.

5. Additional constraints:

   $$\mathit{if} \quad d_{i+1}(y,x) < 0 \quad \mathit{then} \quad d_{i+1}(y,x) \leftarrow \frac{1}{2} d_{i+1}(y,x)$$

   $$\mathit{if} \quad d_{i+1}(y,x) > \mathit{image\,width} \quad \mathit{then} \quad d_{i+1}(y,x) \leftarrow \mathit{image\,width}$$

The reason behind the last step (additional constraints) is to eliminate values impossible in real-world scenes (but attainable by optical-flow algorithm) in disparity map.

3 Performance

The algorithm has been examined using commonly known test images and video sequences. The results were assessed basing on subjective quality. The technique of [4] was used to obtain ground-truth disparity images presented in Fig.5 and Fig.6. It is intrusive technique and thus is unusable in case of FTV of computer vision applications.

The main problem with proposed algorithm is its incapability to recognize regions that are occluded in scene which yields with false disparity values. These artifacts are visible in Fig.5, especially between leafs of the plant.

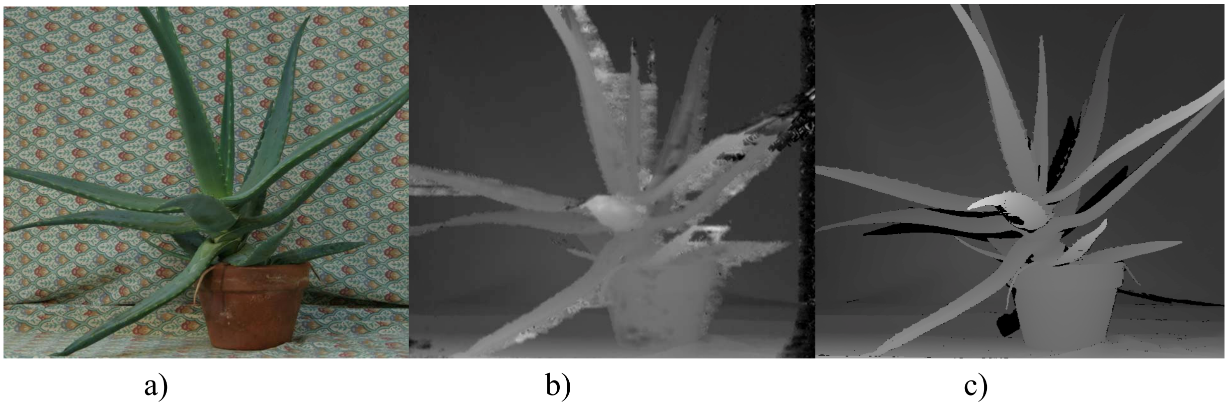


Figure 5. Experiment performed on “Aloe” scene
a) original “Aloe” image [3]
b) disparity map obtained with proposed algorithm
c) ground-truth image obtained using the technique of [4]

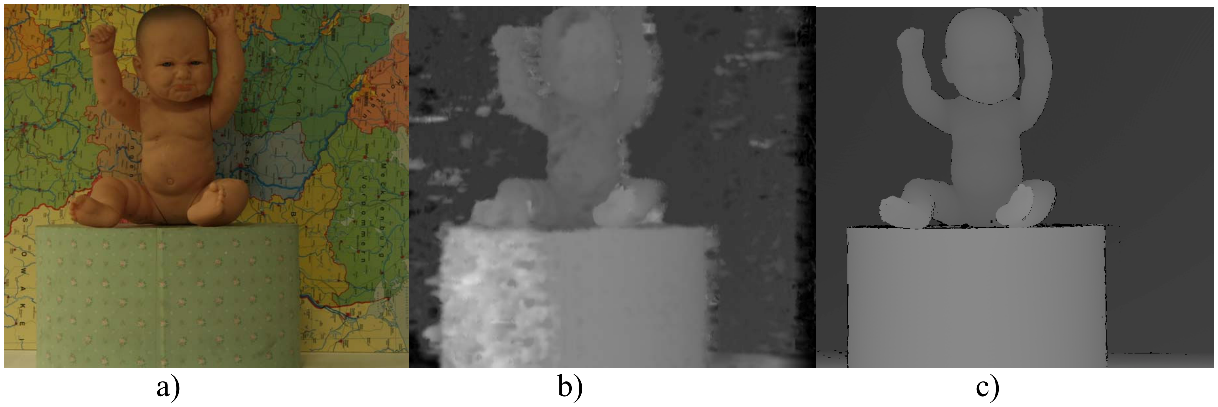
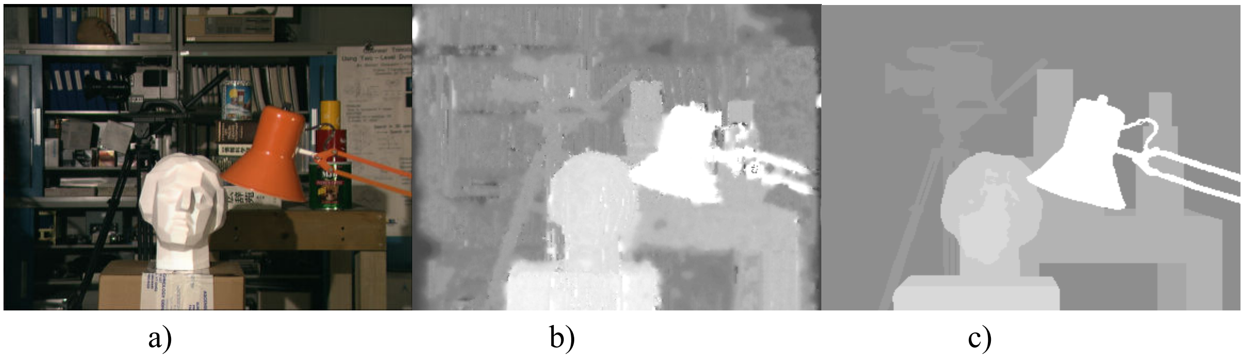


Figure 6. Experiment performed on “Baby” scene
a) original “Baby” image [3]
b) disparity map obtained with proposed algorithm

c) ground-truth image obtained using the technique of [4]

Figure 7 shows that proposed algorithm is capable to reveal background information. Although standard output formats for disparity representation do not allow fractal precision, the output of the algorithm if sub-pixel accurate.



a) b) c)

Figure 7. Experiment performed on Tsukuba University Head Dataset
a) original "Left" image
b) disparity map obtained with proposed algorithm
c) ground-truth image

4 User interface

The application takes two sequences in YUV 4:2:0 format as an input, and produces YUV 4:2:0 disparity output sequence as an output (luminance only).

```
DepthGen.exe
         -il left_sequence
         -ir right_sequence
         -od output_disparity_sequence
         -bl left_image_bitmap
         -br right_image_bitmap
         -bd output_disparity_bitmap
         [-sc disparity_scale_value]
```

`-il` – file name for left-view input sequence in YUV 4:2:0 format
`-ir` – file name for right-view input sequence in YUV 4:2:0 format
`-od` – file name for output disparity sequence in YUV 4:2:0 format
`-bl` – file name for left-view input bitmap in BMP RGB format
`-br` – file name for right-view input sequence in BMP RGB format
`-bd` – file name for output disparity sequence in BMP RGB format
`-sc` – real factor which is used for multiplication of resultant disparity map; default is 1.0. it does not have impact on disparity map generation, but only on representation of final result

5 Conclusions

We proposed an original approach to depth map estimation problem, which exploits two well known techniques: block-matching and optical flow. Both of the techniques were modified and some major enhancements were introduced.

One of the biggest advantages of proposed approach is that it estimates disparity across flat and untextured regions. Iterative nature of 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. Sub-pixel accuracy supported by gradient-based core of optical flow, provides good background estimation which is important for distant scenes. Moreover there are no major contraindications for this technique to be implemented efficiently in hardware e.g. both block matching and optical flow computations are related to local neighbourhood of calculated element only, and thus both are 'cache-friendly'.

One of the main drawbacks of the algorithm is computational complexity which is related to large number of iterations required for convergence. 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]. Another disadvantage of the proposed algorithm is its poor performance on occluded regions. We anticipate that resilience for occlusion problems might be assured by adding extrapolation step that would repair occluded regions basing on the correct neighbour regions. This will be task of our future work.

6 References

[1] Horn BKP, Schunck BG. "Determining Optical Flow: A Retrospective." Artificial Intelligence 336 (10 1993): 162-163.

[2] W. Matusik, H. Pfister, T. Weyrich, A. Vetro, "Calibration and Rectification Procedures for Multi-Camera Systems" ISO/IEC JTC1/SC29/WG11 M11435, Palma de Mallorca, Oct 2004.

[3] Middlebury Stereo Vision Page, http://vision.middlebury.edu/stereo/

[4] D. Scharstein and R. Szeliski. „High-accuracy stereo depth maps using structured light" In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003),* volume 1, pages 195-202, Madison, WI, June 2003.

[5] M. Tanimoto, T. Fujii and K. Suzuki, "Multi-view depth map of Rena and Akko & Kayo", ISO/IEC JTC1/SC29/WG11, M14888, October 2007.

[6] "Call for Contributions on FTV Test Material", ISO/IEC JTC1/SC29/WG11, MPEG 2007/N9468, Shenzhen, China, October 2007.