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

ISO/IEC JTC1/SC29/WG11
MPEG2013/M31518
October 2013, Geneva, Switzerland

Source **Poznań University of Technology**
Nagoya Industrial Science Research Institute
Status **Proposal**
Title **Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television**
Author **Olgierd Stankiewicz, Krzysztof Wegner, Masayuki Tanimoto, Marek Domański**

Introduction

At the 105th meeting in Vienna in July 2013 a new Ad hoc group on Free-viewpoint Television (FTV) has been founded.

One of the mandates of the new founded FTV group is to study available technologies for Free-viewpoint Television. During the recent years various formats for representation of 3D data have been developed. Currently, the most popular option that has been investigated during the works on second phase of FTV, is MVD representation. Therefore, for the sake of evaluation of the future proposals for new data formats and their comparison with existing state-of-the art, a reference software for scene analysis and synthesis is desired.

During the works on compression technology for 3D video in the second phase of FTV, a software called Depth Estimation Reference Software has been developed. The software has been initially proposed by Nagoya University and then improved jointly by the group, up to version 5.1. This software has been thoughtfully tested for dense linear arrangements of cameras, but not for other camera arrangements. A practical verification with new test sequences has shown that some required features are missing.

This paper presents a new version of Depth Estimation Reference Software. It has been enhanced with extended support for arbitrary camera arrangement. The range of supported disparities has been extended to 16-bit. Previously used 8-bit depth representation allowed only for 256 different values of disparities, which in our research turned to be not enough for circular camera arrangement. Also, support for moving (changing in time) camera parameters has been added.
The detailed description of introduced extensions and enhancement can be found below.

1. Introduced Extensions

1.1. Epipolar line search

In linear camera arrangement rectified imaged are assumed to be provided. In the case of sparse or circular camera arrangement, rectification often cannot be performed either because is not technically attainable at all (e.g. due to positioning of the cameras) or because it would distort the image at unacceptable level (Fig. 1). As rectification is transformation which aligns epipolar lines to horizontal rows in all camera, the more skewed is the camera arrangement, the more distorted is the rectified image.

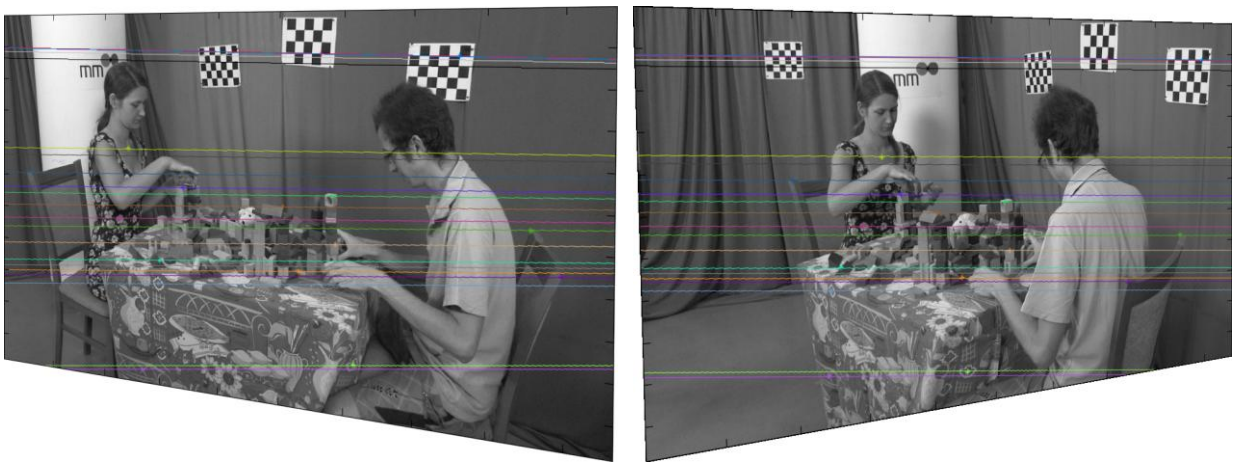

Fig. 1. Exampled rectified pair of images from “Poznan_Game” sequence with unacceptable distortion caused by rectification.

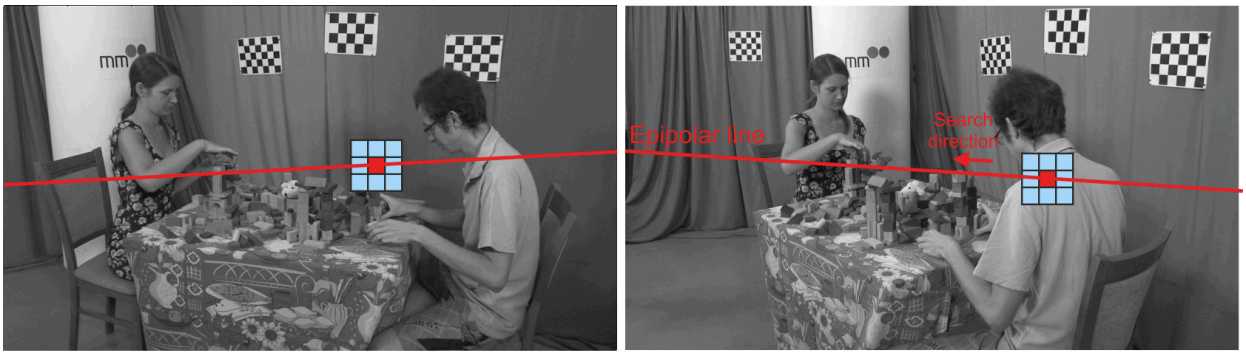

Fig. 2. Explanation of epipolar line search.

Therefore we have introduced epipolar line search which doesn’t require rectified image to be provided. Pixel, block, soft segment matching is done along epipolar line calculated based on camera parameters (Fig. 2).

```
MatchingMethod     4   # 0...Conventional,      1...Disaprity-based,
                         2...Homography-based, 3...Soft-segmentation-based,
                         4 Epipolar Line-based
```

1.2. Vertical up-sampling.

In DERS 5.1, sub-pixel disparity search is performed by up-sampling of the reference image. As the search has been performed only along horizontal rows (in rectified images all disparities are horizontal) only horizontal up-sampling was performed (Fig. 3).

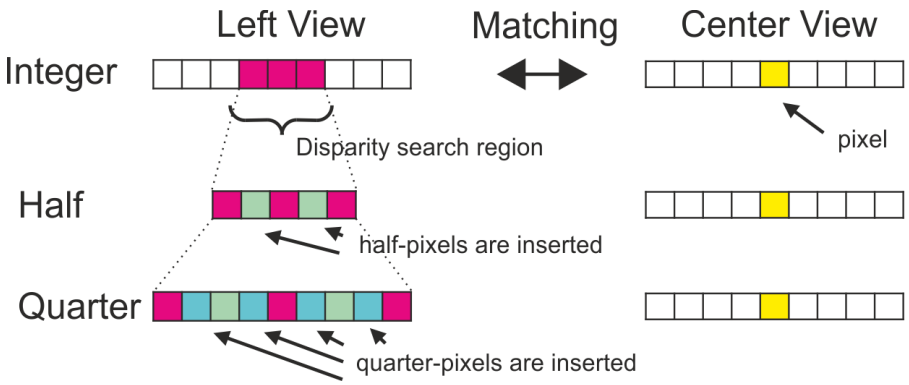

Fig. 3. Search in horizontal direction only, performed in DERS 5.1.

In generalized case, search is performed along epipolar lines. Those can point to any point in the image and thus any sub-pixel position. Therefore, as the disparities are not solely horizontal, also the up-sampling needs to be performed in both directions. Support for such bi-directional up-sampling has been added (Fig.4).

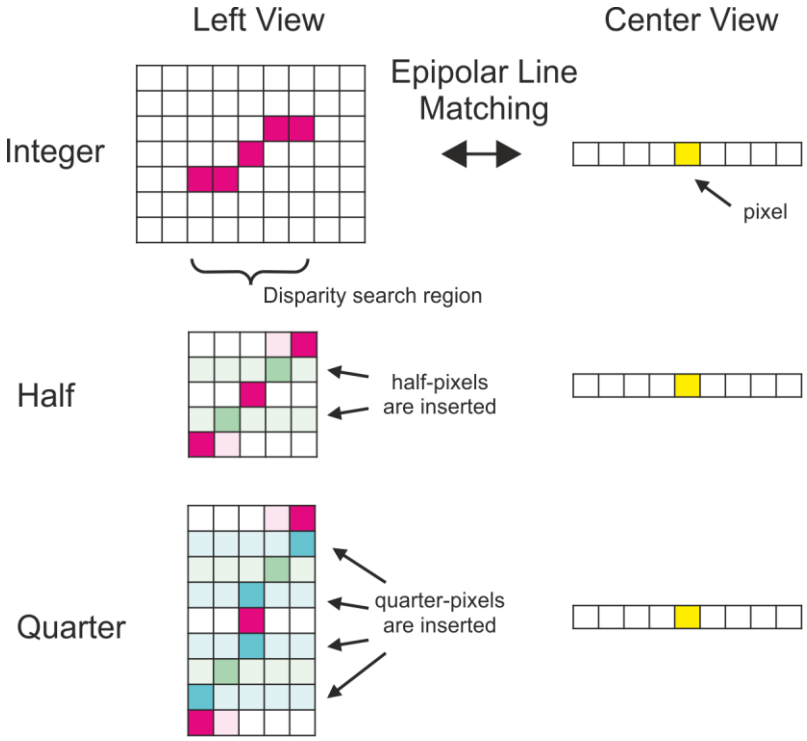

Fig. 4. Explanation of vertical up-sampling.

```
Filter                    1   # 0...(Bi)-linear, 1...(Bi)-Cubic, 2...MPEG-4 AVC 6-tap
VerticalFilter            1   # 0...(Bi)-linear, 1...(Bi)-Cubic, 2...MPEG-4 AVC 6-tap
```

1.3. Extended depth range/density - 16 bit depth maps

DERS version 5.1 supports only 8 bit depth maps. This corresponds to 256 possible disparities which can be represented in an estimated depth map.

Z-near and Z-far normalization allows to use the software for any depth range but wider depth range results in more quantized depth maps. In previously used data sets (with linear camera arrangement) this was found to be satisfying as the number of considered disparity values (even with sub-pixel precision) not exceeded 256 much.

In sparse or circular camera arrangement, number of disparity values that need to be represented can be much larger. Therefore, a new version of DERS have been modified in order to support up to 65536 depth values (16 bit) with possibility of further easy enhancement if necessary.

Modification of bit-depth of generated depth maps can be simply modified by usage of compilation flag in *version.h* file.

```
#define DEPTH_16BIT      //Enables 16 bit depth maps

#ifdef DEPTH_16BIT                //If 16bit depth is on define appropriate c++ types
  typedef unsigned short DepthType; //defines 16bit variable for storing disparity values
  typedef unsigned char ImageType;  //define 8bit variable for storing image samples
  #define MAX_DEPTH (256*256)       //Maximum disparity value
  #define MAX_LUMA 256              //Maximum luminance value
#else
  typedef unsigned char DepthType;
  typedef unsigned char ImageType;
  #define MAX_DEPTH 256
  #define MAX_LUMA 256
#endif
```

1.3 Support for stereo case

By default DERS requires 3 views for depth estimation. Often only 2 view are available and thus we have added support for stereo input.

By default 3-view matching is performed.

```
MatchDirection                    3               # 1...Left 2...Right 3...Both
```

2. Example

Exemplary configuration file for enhanced Depth Estimation Reference Software.

```
#=============== INPUT PARAMETERS ===============
DepthType                         0               # 0...Depth from camera, 1...Depth from
the origin of 3D space
SourceWidth                       1920            # Input frame width
SourceHeight                      1080             # Input frame height
StartFrame                        0              # Starting frame #
TotalNumberOfFrames               1              # Total number of input frames
LeftCameraName                    param_cam2     # Name of left camera
CenterCameraName                  param_cam3     # Name of center camera
RightCameraName                   param_cam4     # Name of right camera
MatchDirection                    3               # 1...Left 2...Right 3...Both
```

```
MinimumValueOfDisparitySearchRange     160             # Minimum value of disparity search range.
This value is not always same as all pairs of views.
MaximumValueOfDisparitySearchRange     250             # Maximum value of disparity search range.
This value is not always same as all pairs of views.
MinimumValueOfDisparityRange           160             # Minimum value of disparity range. This
value is smaller than or equal to minimum value of disparity search ranges for all views.
MaximumValueOfDisparityRange           250              # Maximum value of disparity range. This
value is larger than or equal to maximum value of disparity search ranges for all views.
SmoothingCoefficient                   1.0                # Smoothing coefficient to compute
depth maps
FileLeftViewImage                      v1920x1080_seq14_cam2.yuv    # Name of left input video
FileCenterViewImage                    v1920x1080_seq14_cam3.yuv    # Name of center input video
FileRightViewImage                     v1920x1080_seq14_cam4.yuv    # Name of right input video
FileOutputDepthMapImage                v1920x1080_depth_16bps.yuv # Name of output depth map file
FileCameraParameter                    parametry1234.txt    # Name of text file which includes
camera parameters

BaselineBasis                          1  # 0...minimum baseline, 1...maximum baseline, 2...left
baseline, 3...right baseline
Precision                              1  # 1...Integer-pel, 2...Half-pel, 4...Quater-pel
SearchLevel                            1  # 1...Integer-pel, 2...Half-pel, 4...Quater-pel
Filter                                 0  # 0...(Bi)-linear, 1...(Bi)-Cubic, 2...MPEG-4 AVC 6-tap
VerticalFilter                         0  # 0...(Bi)-linear, 1...(Bi)-Cubic, 2...MPEG-4 AVC 6-tap

MatchingMethod                         4  # 0...Conventional, 1...Disaprity-based,
                                          # 2...Homography-based, 3...Soft-segmentation-based,
                                          # 4 Epipolar Line-based

#========== Temporal Enhancement ==========
TemporalEnhancement                    0       # 0...Off, 1...On
Threshold                              2.0     # Threshold of MAD

#========== Size of Matching Block ==========
MatchingBlock                        3        # 1...Pixel matching, 3...3x3 block matching

#========== Segmentation ==========
ImageSegmentation                      0    # 0...Off, 1...On
SmoothingCoefficient2                  1.00 # Smoothing coefficient to compute depth maps
SegmentationMethod                       3  # 1...mean shift algorithm, 2...phyramid segmentation,
                                            # 3...K mean clustering
MaxCluster                              64  # Positive Integer Value

#========== Soft-Segmentation ==========
SoftDistanceCoeff                      100            # SoftSegmentation Distance Coefficient 21
SoftColorCoeff                         100            # SoftSegmentation Color Coefficient 10
SoftBlockWidth                         11             # SoftSegmentation Block Width 11
SoftBlockHeight                        11             # SoftSegmentation Block Height 11

#========== Occlusions ==========
Occlusion                              0              # 0...Off, 1...On

#====================================================
#========== Semi-automatic Depth Estimation ==========
#====================================================
DepthEstimationMode                  0                # 0...automatic Depth Estimation;
                                                      # 1...Semi-automatic Nagoya;
                                                      # 2...Semi-automatic ETRI

#---- For DepthEstimationMode = 1  ------
TemporalWeight                         0.5            # Temporal mode Cost weight;
                                #(weight)*cost_current_frame + (1.0-weight)*cost_previous_frame
FileCenterManual              Poznan_Street_Depth\manual\Poznan_Street_00_1920x1088_rec_cam05
                                   #Path and filename prefix of the manual input files

#---- For DepthEstimationMode = 2  ------
RefreshFrame                           10              # Period of refresh frame
ThresholdOfDepthDifference             10              # Threshold value of depth difference
MovingObjectsBSize                     0               # 0: small  1: medium  2: large
MotionSearchBSize                      0               # 0: narrow 1: medium 2: wide
```

MatchDirection

Unsigned Int (1-3), default: 3

Specifies the matching direction.1 means that matching is done between left and center view, 2 means that matching is done between right and central view, 3 means that matching is done between both left-center and right-center views. For stereo only you should use either value 1 or 2.

`VerticalFilter`

Unsigned int (0-2), default: 1

Specifies the up-sampling filter to generate image signals on sub-pixel vertical positions. 0 means (bi-) linear filter, 1 means (bi-) cubic filter, and 2 means filter which used in MPEG-4 AVC.

`MatchingMethod`

Unsigned Int (0-3), default: 2

Specifies the method how to identify corresponding pixels. 0 means the conventional method, where disparities are given directly, 1 means disparity-based method where disparities are calculated from z values, 2 means homography-based method, 3 means soft-segmentation method and 4 means epipolar line search.

Fig. 5. Exemplary depth map for "Poznan_Game" sequence, obtained using enhanced Depth Estimation Reference Software.

3. Software SVN

The software can be accessed from our SVN server:

https://svn.multimedia.edu.pl/ders

user: mpeg-ftv
pass: ftvftv

4. Conclusion

We presented an enhanced Depth Estimation Reference Software that is natural extension of previously developed DERS software. We propose to use it as a reference in further development and quality assessment of the new proposals of FTV formats.

5. References

[1] M. Tanimoto, T. Fujii, and K. Suzuki, "Multi-View Depth Map of Rena and Akko & Kayo", ISO/IEC JTC1/SC29/WG11, M14888, October 2007.

[2] M. Tanimoto, T. Fujii and K. Suzuki, "Improvement of Depth Map Estimation and View Synthesis," ISO/IEC JTC1/SC29/WG11, M15090, January 2008.

[3] M. Tanimoto, T. Fujii and K. Suzuki, “Reference Software of Depth Estimation and View Synthesis for FTV/3DV,” ISO/IEC JTC1/SC29/WG11, M15836, October 2008.

[4] M. Tanimoto, T. Fujii and K. Suzuki, “Depth Estimation to Improve Boundary Clarification”, ISO/IEC JTC1/SC29/WG11, M15837, October 2008.

[5] 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.

[6] 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.

[7] http://opencv.jp/opencv-1.0.0_org/docs/index.htm .