D. Akimov, A. Shestov, A. Voronov, D. Vatolin, "Automatic left-right channel swap detection,"
International Conference on 3D Imaging (IC3D), pp. 1–6, 2012. doi:10.1109/IC3D.2012.6615126

AUTOMATIC LEFT-RIGHT CHANNEL SWAP DETECTION

Dmitry Akimov, Alexey Shestov, Alexander Voronov, Dmitriy Vatolin

Department of Computational Mathematics and Cybernetics
Lomonosov Moscow State University
Moscow, Russia

ABSTRACT

Automatic analysis of stereo-video quality plays an important role in the process of capturing, converting and editing video in 3D format. Although several low-level stereo-video quality metrics were proposed, a lot more challenging problems of high-level stereo-video analysis, such as left-right channel swap detection, are still practically unsearched. The result of channel swap is very disturbing, but it is not always obvious even to a human observer what is wrong in such sequence. In this paper we represent a fully automatic algorithm for left-right channel swap detection. Experimental results for real video sequences demonstrate the effectiveness of the proposed technique.

Index Terms— Stereo vision, stereo image processing, image analysis.

1. INTRODUCTION

Every small mistake in the process of a 3D movie production may lead to a lot of serious issues in the final product. For example, sometimes left and right views can be occasionally swapped in some scenes of the movie.

Our goal was to detect such scenes in the movies. Actually, it is often difficult to detect artefacts of this type even for a human. It is a common issue that a viewer understands that something is wrong but is not able to determine a problem. The correctness of views arrangement can be checked on the basis of foreground-background segmentation and inter-view optical flow analysis. Currently we have performed early stage research, implemented initial version of the algorithm for swapped views detection and tested it on our test set. The testing results were used to perform the precision-recall, reveal drawbacks of the proposed techniques and determine further directions of algorithm improvement.

2. RELATED WORK

Most of the existing stereo quality metrics are focused on estimation of stereoscopic artefacts that were caused by different distortions in views or in depth map, for example color and focus distortions, vertical disparity, object boundaries distortion. Several authors [7, 3] tried to use the existing 2D quality metrics such as PSNR, SSIM, VQM, UQI, C4, RRIQA to perform quality evaluation.

(a) Left view

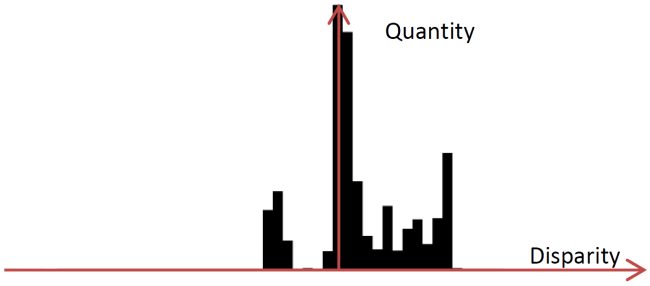
(b) Right view

Fig. 1: The left view occlusions are always leftwards the object (a), and on the right view — rightwards (b). The frame is taken from the movie "Pirates of the Caribbean: On Stranger Tides".

In [2, 10, 8] authors used different metrics for monoscopic and stereoscopic artefacts. The disparity estimation algorithms were used to compute binocular artefacts. Some interesting concepts were introduced in [9] by Mike Knee. Their main algorithm idea is that in most scenes objects at the center and the bottom of the screen are generally nearer than objects at the top and sides.

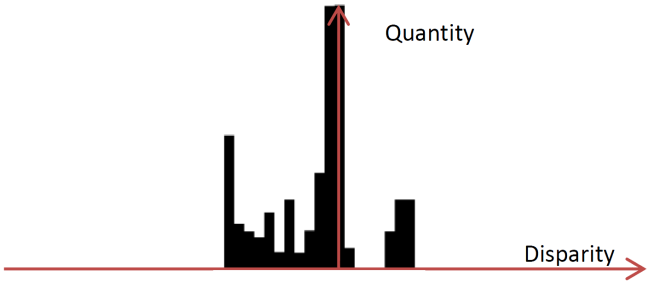
A detection algorithm that analyses correlation of measured disparity distribution with the above template is proposed in the paper mentioned above. Author also mentions

(a) A left view of the source frame with marked occlusions



(b) Histogram of the x-coordinate of the left view OF vector field

(c) OF vector field of the left view



(d) Histogram of the x-coordinate of the right view OF vector field

Fig. 2: An example of OF vector x-coordinate histograms (b, d) for the source frame with edges (a) and corresponding OF vector field (c). Frame taken from the movie “Pirates of the Caribbean: On Stranger Tides”. OF vector field is visualised using Middlebury color benchmark [1].

that a potentially more reliable method of left-right channel swap detection can be based on the fact that closer objects are expected to occlude objects that are further away. In the proposed technique we will use this fact as the main rule of detection procedure and add new disparity analysis as a support method in cases with lack of occlusions.

3. CONCEPTS AND ASSUMPTIONS

In this section we will introduce the main terminology that is used in the paper, and will describe the main ideas and assumptions which determine the range of the cases where the proposed algorithm can be applied.

3.1. Necessary definitions

Binocular disparity refers to the difference in image location of an object seen by the left and right eyes, resulting from the eyes’ horizontal separation. We will call it disparity further.

We treated disparity as the motion vector field between views. We will use the terms “motion vector” and “disparity” as synonyms.

Occlusions are the regions which are presented in only one view of a stereo image and are closed by foreground objects in the other view.

Left-right consistency (or LRC) is the confidence measure for inter-view Optical Flow [5]. Denoting the motion vector in the point p in the left image as $\overrightarrow{v}$ and the motion vector in the point $p + \overrightarrow{v}$ in the right image as $\overrightarrow{u}$, in ideal situation $\overrightarrow{u}+\overrightarrow{v} = \overrightarrow{0}$. So the greater $||\overrightarrow{u}+\overrightarrow{v}||$ is — the less confident is the motion vector in the point p.

3.2. Assumptions

All the algorithm assumptions and ideas are based on the two facts of binocular stereography.

The first fact is that in the left view occlusions are always leftwards the object and in the right view — rightwards (see Figure 1). Indeed, disparity depends on depth monotonically, so in the right view (left view) disparity value of a closer object is always larger (less) than disparity value of further object (see Figure 4). Let us treat the disparity as the motion vectors. While moving from the right view to the left (from the left to the right) the closer object has positive (negative) motion relative to the further objects, e.g. moves rightwards (leftwards).

The second fact is that negative parallax regions are one-third of the zone of stereo comfort perception, and positive parallax regions are two-third of this zone (see Figure 2).

According to the first fact the density of the image edges rightwards and leftwards the occlusions is the measure of like-

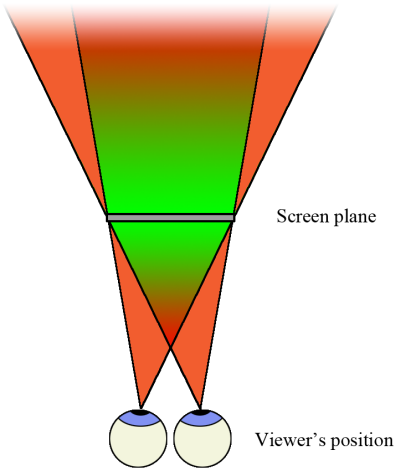


Fig. 3: Stereo perception zones. Areas of comfort stereo perception are marked with green color. Red and orange colors correspond to highly uncomfortable areas.

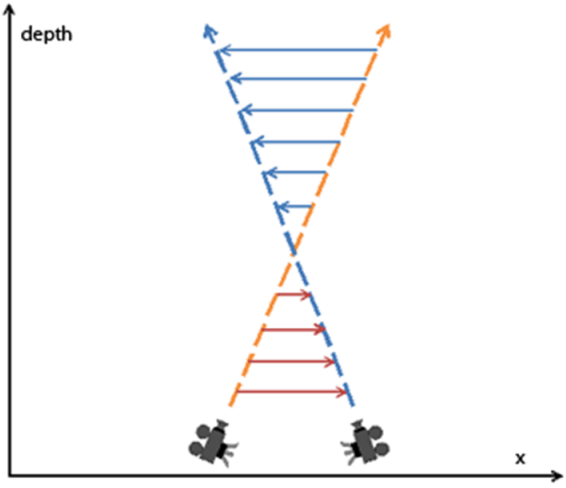


Fig. 4: Dependence of disparity on object depth. Objects with lower depth value have lower disparity (considering disparity as a signed value).

lihood whether the current view is left or right.

But there are some frames, where occlusions are so thin, that they can’t be detected by the algorithm at all or we can’t say if edges are rightwards or leftwards them. In such cases there mustn’t be big foreground objects, which depth values are significantly different from the background depth. So, according to the second fact, we can expect that the left view has more positive disparity values than negative and vice versa for the right view. In such cases the barycentre of the left view disparity histogram must have positive coordinate and the barycentre of the right — negative (see Figure 3). So the difference of the barycentres can be used as an indicator of the views arrangement correctness.

These considerations lead us to the following pipeline:

1. Estimation of the necessary data.
2. Occlusion-based decision making.
3. Histogram-based decision making.
4. Methods combining.

4. ALGORITHM

4.1. Preprocessing step

For each view the necessary data estimation is performed:

4.1.1. Disparity

To estimate disparity we used the Optical Flow (OF) algorithm described in [11].

4.1.2. Occlusions

To estimate occlusion map the left-right consistency metric was calculated. The resulting occlusion areas were estimated using LRC thresholding and median filtering.

4.1.3. Edge detection

We tried simple and reliable Canny algorithm [4] and it produced satisfactory results. In future we will probably use more complicated edge detectors.

4.1.4. gradient calculation

The next important data is image gradient estimated in L*a*b* color space using Sobel filter [6]. We use absolute gradient values as a confidence measure for edges.

4.1.5. Disparity histogram

In Figure 2, examples of histograms of left and right disparity x-coordinates are presented: this is the case of thin occlusions. That is why the barycentre of the left OF vector field histogram must be rightwards the barycentre of the right OF vector field histogram.

4.2. Occlusion-based decision making

In this step we calculate how well the occlusion boundaries align with the image edges. We use the following considerations:

- wide occlusions are more confident than thin occlusions

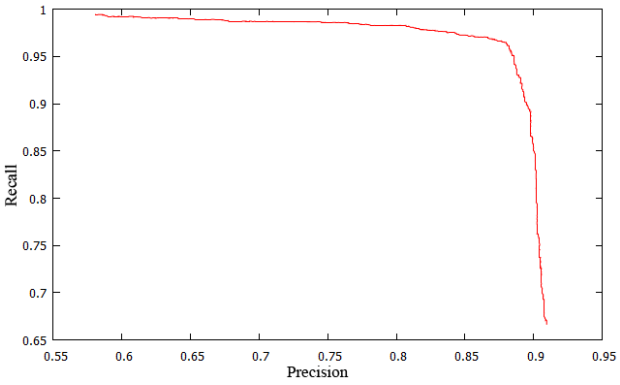


Fig. 5: Precision-recall diagram obtained by varying the *probability threshold* with all other parameters fixed.

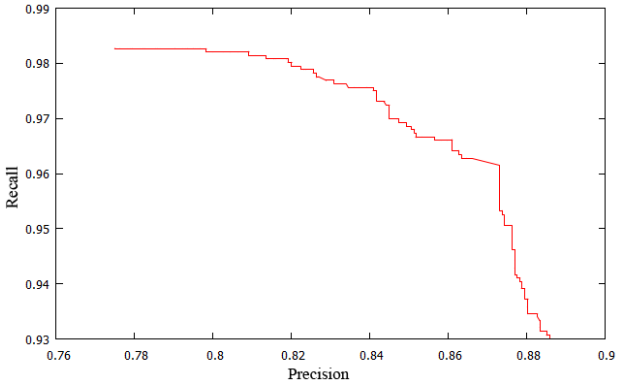


Fig. 6: Precision-recall diagram obtained by varying the *distance threshold* with all other parameters fixed.

- edges with higher values of gradient magnitude are more confident
- the closer an occlusion is to an edge — the more likely they correspond to one object's border.

According to it, for each view two values are calculated:

$$LS = \int\limits_{Bound_R} \frac{width_{occ}(p) \cdot edge\_conf(p_{edge})}{\|p - p_{edge}\|_x} \mathrm{d}p \quad (1)$$

$$RS = \int\limits_{Bound_L} \frac{width_{occ}(p) \cdot edge\_conf(p_{edge})}{\|p - p_{edge}\|_x} \mathrm{d}p \quad (2)$$

$$p_{edge} = \underset{p' \in edge\ mask}{\arg\min} \|p - p'\|_x \quad (3)$$

where $\|\cdot\|_x$ denotes x-coordinate of a vector, $Bound_L$ and $Bound_R$ are the left and right occlusion boundaries respectively and $width_{occ}(x)$ is the width of the occlusion. So, we obtain four numbers: LS and RS of the left view (LS_L, RS_L) and LS and RS of the right view (LS_R, RS_R). Then we estimate the probability that views are not swapped:

$$LR_{sum} = LS_L + RS_L + LS_R + RS_R \quad (4)$$

$$P_{LR} = \frac{LS_L + RS_R}{LR_{sum}} \quad (5)$$

and compare it with the *probability threshold*. If it's less we decide that views of the current frame are swapped.

4.3. Histogram-based decision making

For this step we need disparity histograms of left view and right view (see Figure 2). We estimate the following values:

$$moment_l = \sum_{v_x=v_{xmin}}^{v_{xmax}} hist_{lx}(v_x)^\gamma \quad (6)$$

$$moment_r = \sum_{v_x=v_{xmin}}^{v_{xmax}} hist_{rx}(v_x)^\gamma \quad (7)$$

$$histogram\ difference = moment_l - moment_r \quad (8)$$

Where $hist_{lx}$ and $hist_{rx}$ are the disparity x-coordinate histograms for the left and the right views respectively, is a tunable parameter, it is chosen according to the testing results. Then we compare *histogram difference* with *distance threshold*. If it's less, we decide that views are swapped at the current frame.

4.4. Histogram-based decision making

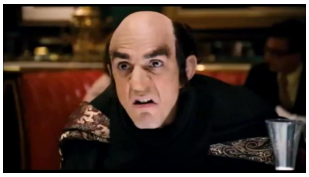
We have two methods for the decision making whether views are swapped or not: occlusion method and histogram method. Occlusion method is suitable if there are some wide reliable occlusions in the scene, histogram method — in the other case. So, for each scene we must decide what method we will use. We take LR_{sum} (4) as an indicator of occlusions presence in the frame. If the LR_{sum} is less than threshold, the histogram method is used, if it is higher, the occlusions method is used.

As it turned out, the algorithm was mostly sensitive to the *distance threshold* and *probability threshold*. That is why these two parameters were used for algorithm tuning and precision-recall regulation.

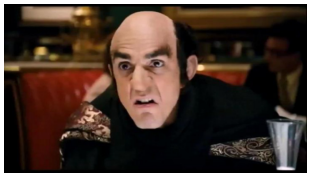
5. EXPERIMENTS

5.1. Algorithm tuning

For our test set we took 780 random frames from 13 movies, 60 frames from the each movie. These frames were manually checked. For our task it is easy to obtain samples with swapped channels, we need only to swap views of good frames.

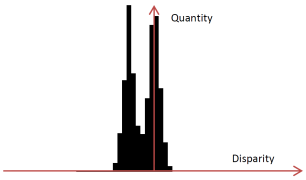


(a) "Left" view in the trailer (which indeed must be the right view)



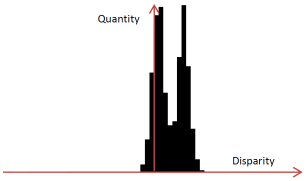
(b) "Right" view in the trailer (which indeed must be the left view)

(c) "Left" view OF vector field



(d) Histogram of the "left" view OF vector field

(e) "Right" view OF vector field



(f) Histogram of the "right" view OF vector field

Fig. 7: Example of frame (a-b) with swapped channels detected by the proposed algorithm, corresponding OF fields (c, e) and OF vector histograms (d, f). Frame is taken from "The Smurfs" movie trailer. OF vector field is visualised using Middlebury color benchmark [1].

So we obtain a test set of 1560 frames — 780 frames with non-swapped views and 780 frames with swapped views.

We searched for optimal parameter values in order to minimize false negatives, because our main goal is to find frames with swapped channels in real movies. So in our tests we preferred to preserve high recall values and in some cases loose in precision.

Results of our tests are presented as precision-recall diagrams (see Figure 5, 6).

First, we ran our algorithm on the test set varying probability threshold from 0.01 to 1.0 with all other parameters fixed and obtained the precision-recall diagram, which is presented in Figure 5.

Next, we ran our algorithm on the test set varying distance threshold from -2.0 to 12.0 with all parameters fixed and obtained the next precision-recall diagram (see Figure 6).

5.2. Further improvements

After analysing the results we determined some ways to improve our algorithm. First of all, we can combine methods for decision making in a more complicated way, than simple thresholding. Next, we can use the consideration that a color of pixels in occlusion must be similar with background color, not with the object color. The last idea refers to object behaviour in the scene. There is the fact that on a left view the object with the lowest depth value must have the highest disparity value (if we consider disparity as a signed value, not only its modulus) and other way round on a right view (see Figure 4). The decision depends on the position of detected background and foreground areas. We can use rough occlusions produced by Motion Estimation for the consequent frames (not between views) and their color similarity with background for local background/foreground segmentation. Also we can use segmentation from motion for objects detection.

6. CONCLUSION

An algorithm for automatic swapped views detection is proposed. It is based on occlusion detection and motion vectors histogram. The algorithm was tested on 780 frames from 13 movies. The precision-recall diagrams were constructed using two parameters. The complexity is estimated. The drawbacks are analysed and further directions are proposed.

7. ACKNOWLEDGEMENTS

This work is partially supported by the Intel/Cisco Video-Aware Wireless Network (VAWN) Program and by grant 10-01-00697a from the Russian Foundation of Basic Research.

8. REFERENCES

[1] S. Baker, D. Scharstein, P. J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[2] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and B. G. Akar. Towards compound stereo-video quality metric: a specific encoder-based framework. In *in Image Analysis and Interpretation*, pages 218–222, 2006.

[3] P. Campisi, P. Callet, and E. Marini. Stereoscopic images quality assessment. In *in 15th European Signal Processing Conference*, pages 2110–2114, 2007.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:679–698, 1986.

[5] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1127–1133, August 2002.

[6] H. Farid and E. P. Simoncelli. Differentiation of discrete multi-dimensional signals. *IEEE Trans Image Processing*, 13:496–508, 2004.

[7] C. T. E. R. Hewage, T. S. Worrall, S. Dogan, and M. A. Kondoz. Prediction of stereoscopic video quality using objective quality models of 2-d video. *Electronics Letters*, 44(16):963–965, 2008.

[8] L. Jin, A. Gotchev, A. Boev, , and K. Egiazarian. Validation of a new full reference metric for quality assessment of mobile 3dtv content. In *in 19th European Signal Processing Conference*, pages 1894–1898, 2011.

[9] M. Knee. Getting machines to watch 3d for you. *SMPTE Moion Imaging Journal*, 121:52–58, 2012.

[10] A. Mittal, K. A. Moorthy, J. Ghosh, and C. A. Bovik. Validation of a new full reference metric for quality assessment of mobile 3dtv content. In *in Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, pages 338–343, 2011.

[11] A. S. Ogale and Y. Aloimonos. Shape and the stereo correspondence problem. *Int. J. Comput. Vision*, 65(3):147–162, December 2005.