Methodology for stereoscopic motion-picture quality assessment

Alexander Voronov, Dmitriy Vatolin, Denis Sumin, Vyacheslav Napadovsky, Alexey Borisov

Lomonosov Moscow State University, Moscow, Russia

ABSTRACT

Creating and processing stereoscopic video imposes additional quality requirements related to view synchronization. In this work we propose a set of algorithms for detecting typical stereoscopic-video problems, which appear owing to imprecise setup of capture equipment or incorrect postprocessing. We developed a methodology for analyzing the quality of S3D motion pictures and for revealing their most problematic scenes. We then processed 10 modern stereo films, including *Avatar*, *Resident Evil: Afterlife* and *Hugo*, and analyzed changes in S3D-film quality over the years. This work presents real examples of common artifacts (color and sharpness mismatch, vertical disparity and excessive horizontal disparity) in the motion pictures we processed, as well as possible solutions for each problem. Our results enable improved quality assessment during the filming and postproduction stages.

Keywords: stereoscopic video, quality assessment, stereo matching, color mismatch, sharpness mismatch, vertical disparity

1. INTRODUCTION

In this work we address the problem of stereoscopic-video quality control during filming and postproduction. Numerous factors can cause visual discomfort when viewing stereoscopic films. Basically, users experience discomfort when they see unusual effects in the stereoscopic picture that are impossible under natural viewing conditions. These types of unwanted artifacts have been studied and are considered to be problematic.

Some S3D-content issues and possible artifacts are described in [1]. The authors discuss the issues that arise throughout the production pipeline from the video creator to the viewer: during content creation, format conversion, encoding and decoding, and rendering for a specific display device. More-detailed classification of issues appearing in S3D video is presented in [2] with attention paid specifically to mobile 3DTV.

Since S3D content can be created in different ways—including filming with a stereo camera, conversion from 2D using depth maps and computer-graphics rendering—certain artifacts are typical for certain methods of stereoscopic-video creation. The common problem, which is applicable to all the methods, is depth-budget control. The necessity of depth-budget control relates to vergence-accommodation conflict. Avoiding frequent use of excessive disparity, which causes visual fatigue, is preferred. Frequent transitions from scenes with significant negative disparity to scenes with significant positive disparity, and vice versa, may also cause discomfort. Furthermore, objects with large positive disparity may fail to converge, especially on a movie screen, when the object disparity exceeds the convergence limits.

Another problem that is common to all the methods of S3D-content creation is stereo-window violation [3]. This artifact appears when an object with significant negative disparity is situated on the left or right boundary of the frame so that a large part of the object is visible only to one eye. This effect can annoy viewers. Possible solutions include changing the vergence point to eliminate negative disparity for the object and application of the floating window technique during postproduction; in this last case, the borders of each view are adjusted independently to hide the artifact.

Authors e-mail: {avoronov, dmitriy, dsumin, vnapadovsky, aborisov}@graphics.cs.msu.ru

In this paper we mostly address the problem of quality control for video captured using a stereo camera. Typical problems specific to this type of content are the following:

- geometry distortion,
- color mismatch,
- focus mismatch,
- time desynchronization,
- excessive or insufficient disparity,
- stereo-window violation.

These problems are caused by inaccurate camera setup or synchronization. Most of them can be fixed during the postproduction stage. Geometry distortion is a result of camera misalignment. Correction of small geometry distortions is possible through affine transforms of the views; the parameters of the transform can be estimated using image rectification [4, 5]. There are a variety of reasons which may cause color mismatch between views. For instance, adjustments to the cameras' brightness and color settings must be synchronized. This type of artifact is the easiest one to fix as long as differences in the camera settings don't yield over- or underexposure in one view. Relevant algorithms are presented in [6] and [7]. More-complex color-mismatch problems can be caused by inaccurate selection of light and polarizing filters. Problems with different polarization may arise when using mirror stereo rigs. Differences in color caused by filters and mirror rigs are usually nonuniform, so locally adaptive color-correction approach is required to fix this problem. Suitable algorithm is described in [8]. In [9], the authors provide physical background on the focus-mismatch problem and propose a metric based on sharpness computation for each pixel in a view, using the estimated disparity map.

Some solutions enable on-the-fly S3D-video quality evaluation during filming [10, 11]. These assistance systems provide depth-budget control and warnings about geometry distortions and stereo-window violation. Such systems restrict an algorithm's computational complexity because they require real-time processing. These types of solutions still don't eliminate the need for quality control of postproduction-stage results.

In this paper, we propose a system for offline quality analysis that is intended for evaluation of an entire video: it provides an average quality score for the video and identified scenes with the most noticeable artifacts. Also, the proposed system can evaluate the quality of individual shots—a feature that can be beneficial during filming and postproduction.

In Section 2, we describe our methodology for analyzing disparity distribution in a film and for identifying scenes with the most noticeable distortions. In Section 3 we present results from application of this methodology to modern S3D films. Section 4 offers ideas on how the methodology can be extended and improved. Finally, we conclude with a short summary of our work in Section 5.

2. PROPOSED METHOD

Here we describe our methodology for S3D-video quality assessment. First, stereoscopic video is provided as input; the output is a set of metric values for each frame in the video. The metric value is a number that quantifies the noticeability of distortion in the frame: the greater the value, the more noticeable the detected distortion. We implemented the following metrics: color mismatch, sharpness mismatch and vertical disparity. To analyze horizontal disparity we use a disparity distribution histogram and values for the maximum positive and negative horizontal disparities. These values enable detection of scenes with excessive disparity. The distribution histogram for each frame is useful for more-complex analysis.

A list of scenes containing the most noticeable distortions is returned for each metric. Finally a quality comparison of different videos can be undertaken using the average metric values.

2.1 Horizontal-disparity distribution analysis

Input for the horizontal-disparity distribution metric is a stereo image or single stereo-video frame. The output consists of a horizontal-disparity distribution histogram and the maximum negative and positive horizontal-disparity values for the image.

The metric-calculation algorithm involves three steps:

1. Color-independent stereo matching
2. Determining and masking unreliable areas in the image
3. Calculating the horizontal-disparity histogram for the unmasked areas and estimating extreme horizontal-disparity values

The calculated metric values help determine which scenes can most benefit from changing the depth budget. Although precisely setting the horizontal parallax before shooting a scene is important, available algorithms enable parallax (and thus disparity) correction. The problem of parallax correction is close to that of view reconstruction. This approach essentially inherits difficult subtasks like the background-restoration problem; therefore, significantly changing the parallax while preserving stereo-picture quality is complicated.
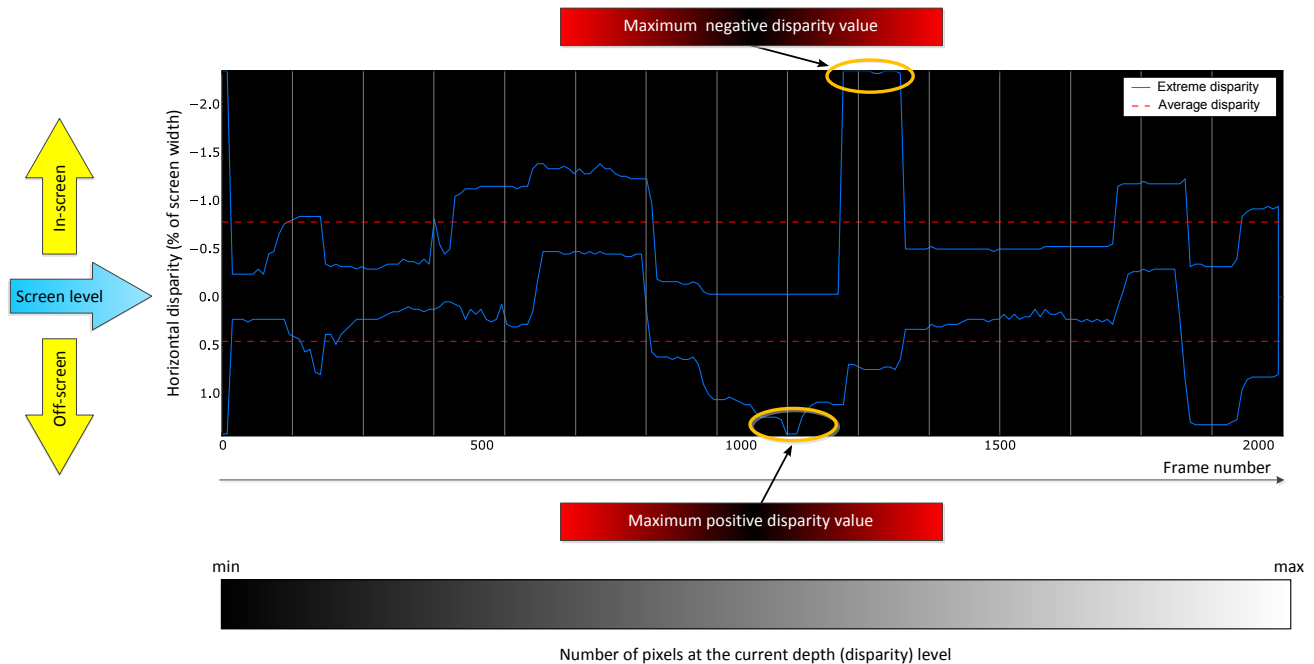

Figure 1: Horizontal-disparity distribution diagram for "MSU Views" video sequence. This diagram helps in identifying scenes with excessive horizontal disparity and enables comparison of depth budgets in different video sequences.

2.1.1 Color-independent stereo matching

Initially we chose the block-based approach described in [12] as a stereo matching algorithm. This approach provides a dense disparity map, what is an advantage over the feature-based approach. The main advantage over per-pixel stereo matching is lower computational complexity. The disparity quality obtained using block-based matching is suitable for our estimation purposes.

To improve the reliability of the matching results for color-mismatched video, we modified the block-matching metric to make it more robust when dealing with local brightness mismatch. Details of this modification are described in [13].
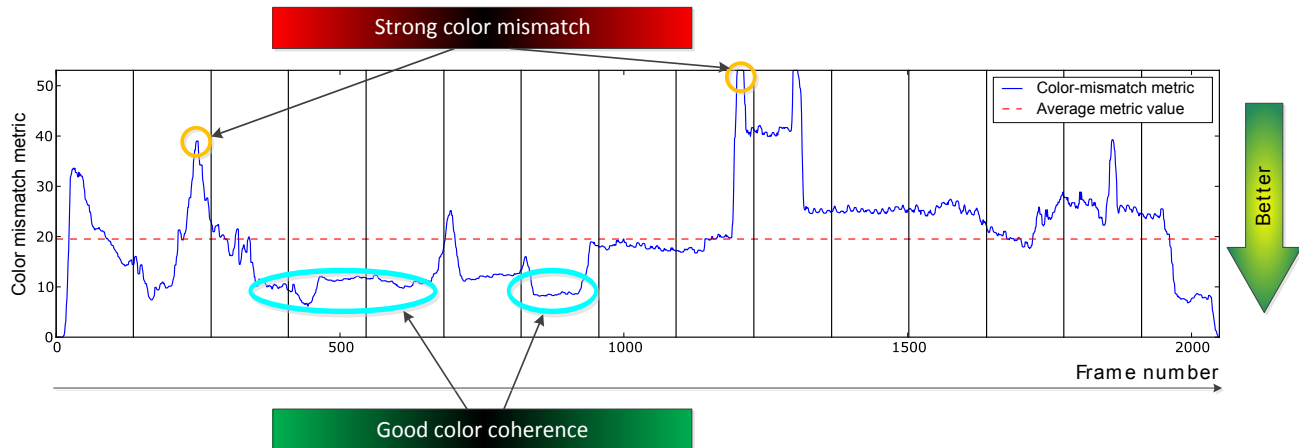
Figure 2: Color-mismatch metric diagram for "MSU Views" video sequence. This diagram helps to identify the scenes with the most noticeable distortion.

2.1.2 Determining unreliable areas and calculating the metric

Owing to possible errors in the block-matching algorithm while processing uniform areas, we exclude these areas from consideration. For each block we estimate the variance of the Y-component in the YUV color space; blocks with a variance less than a given threshold are marked as unreliable. Only reliable blocks are taken into account when estimating disparity distribution histogram.

On the basis of the horizontal-disparity distribution histogram, we estimate the extreme values for the horizontal disparity. To make the estimation process more robust, we exclude 5% of the values from the boundaries of the histogram. The remaining values are considered to be the maximum positive and negative horizontal disparity.

We measure horizontal disparity as a percentage of screen width. This choice of units is related to the fact that we processed only Blu-ray 3D content, which is oriented to home-theater devices. Most home-theater displays have an aspect ratio (AR) of 16:9; the AR of Blu-ray content is usually the same or higher (we are unaware of any Blu-ray release with an AR less than 16:9). As a result we decided to use screen width as the denominator for this metric value. This approach may be unsuitable for movie-theater screens; the best approach may be recalculation of the metric values in terms of parallax if the screen dimensions and viewer position are known.

This metric for extreme horizontal disparity is helpful when analyzing the disparity distribution over the video. The main drawback of such and approach is that the proposed metric doesn't reflect the size of the object with extreme disparity. This fact motivated our use of distribution histograms in conjunction with extreme-values diagram. Figure 1 shows an example diagram.

2.2 Color-mismatch detection

This metric enables detection of scenes with noticeable color differences between views. The main idea is to perform stereo matching of the views, then reconstruct one view from the other on the basis of matching data and compare colors from the original and reconstructed views. For stereo matching we use the same algorithm described in Subsection 2.1. Differences between the original and reconstructed images are measured as the Mean Square Error (MSE) in the RGB color space. We exclude from consideration 5% of image pixels with the largest differences so that incorrect stereo matching of object boundaries and occlusion areas will not affect the results. The metric value is dimensionless.

A detailed description of the color-mismatch detection algorithm can be found in [13]. Figure 2 presents an example per-frame diagram.

The resulting metric values can be used to correct color mismatches. Most postprocessing software packages for stereoscopic pictures have tools for inter-view color correction. Common methods are described in the introduction.

(a) Left view of source frame.

(b) Right view of source frame.

(c) Difference between the original right view and stereo-matching-based left-to-right reconstruction (contrast increased).

(d) Checkerwise blending of the original right view and the left-to-right reconstruction.
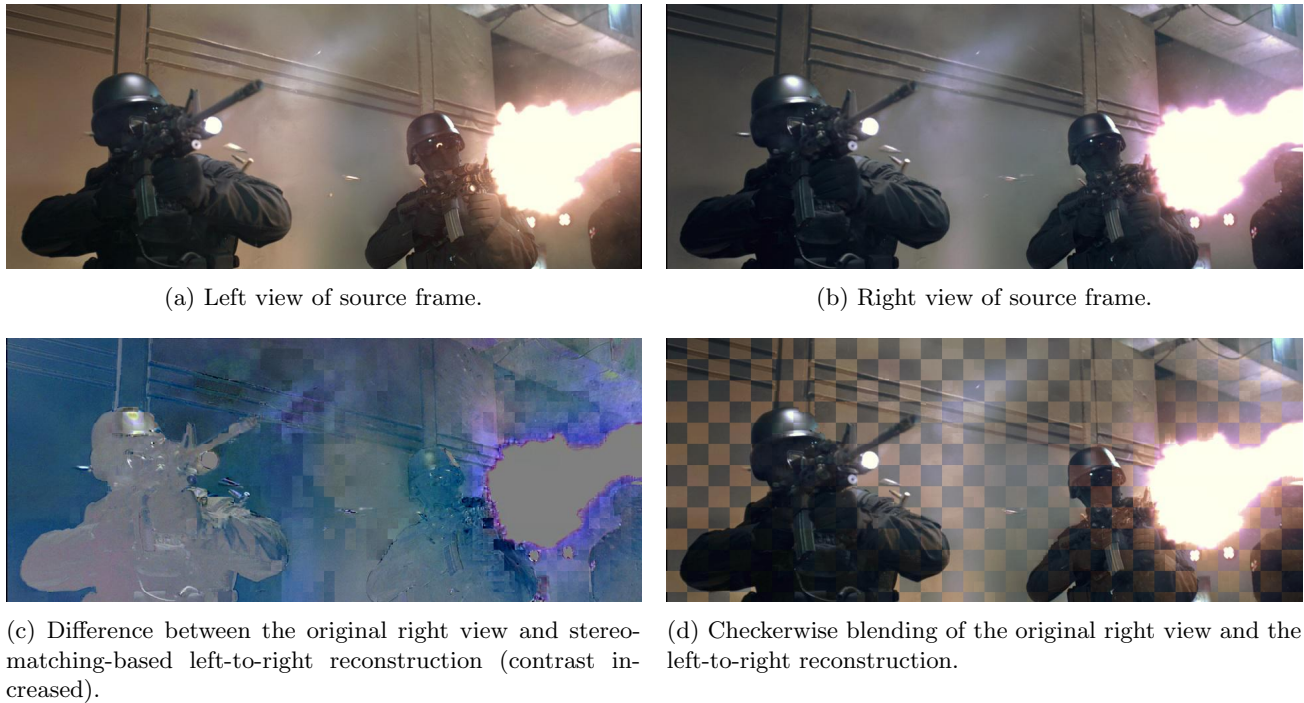
Figure 3: Example of scene from *Resident Evil: Afterlife* detected using color-mismatch metric. Checkerwise-blending visualization in (d) demonstrates noticeable color differences between views.

2.3 Sharpness-mismatch detection

We use the term *sharpness mismatch* instead of *focus mismatch* because the corresponding metric is designed to detect differences in high frequencies; these differences can be caused not only by focus mismatches but also by inaccurate postprocessing, differences in motion blur and asymmetric compression.

The idea of the metric is similar to that of color mismatch. We perform stereo matching, then reconstruct one view from another and compare the original and reconstructed views. The primary difference relative to color-mismatch estimation is that we compare only high-frequency information from the views. High frequencies for each view are estimated as the absolute difference between the original image and the result of bilateral filtering for that image. Figure 4 illustrates the entire process. The resulting value depends on the strength of the sharpness difference and on the size of the area containing the sharpness mismatch. The metric value is dimensionless.

A detailed description of the sharpness-mismatch detection algorithm is presented in [13].

Sharpness-mismatch problems can be avoided during the filming stage, and methods of sharpness-mismatch correction can be applied during the postprocessing stage. In postprocessing, sharpness-mismatch problem can be solved with blurring one view or sharpening the other. The second method is more complicated.

2.4 Vertical-disparity detection

Global vertical disparity is a measure of geometric camera-calibration accuracy. This kind of artifact appears when the cameras in a rig have different angles relative to the horizontal plane. Usually the problem is more complicated, and cameras may have different tilt or perspective distortion. It is unclear how to compare scenes containing tilt, perspective distortion and vertical disparity, which is why at this stage we estimate only the vertical disparity, which is easy to understand.
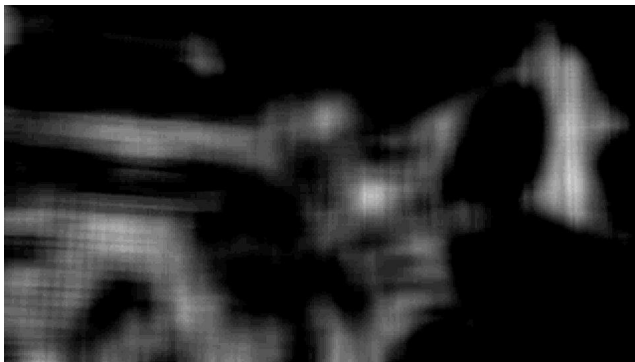
We estimate the vertical disparity using the stereo-matching data. The first two steps of the algorithm are the same as those for horizontal-disparity estimation: we perform stereo matching and then mask unreliable
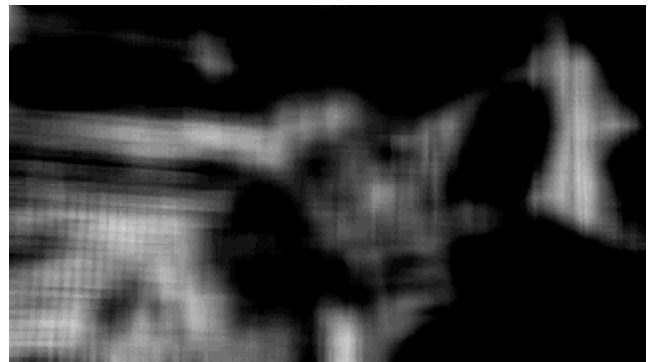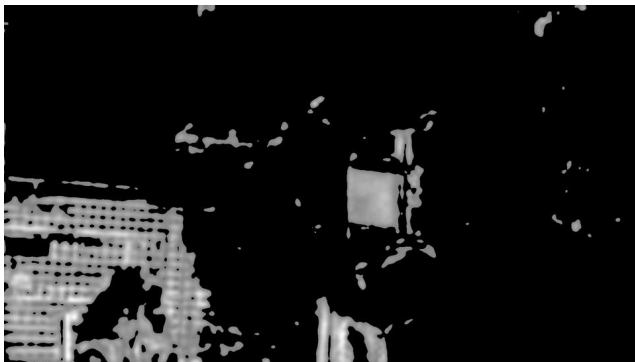
(a) Left view of original image.

(b) Right view of original image.

(c) Estimated high-frequency map of the left view.

(d) Estimated high-frequency map of the right view.

(e) Penalty-map visualization demonstrating which areas have noticeable differences in sharpness according to a comparison of high-frequency maps.

(f) Magnified area with the most noticeable sharpness difference.

Figure 4: Example of scene from *Avatar* with detected sharpness mismatch. The area with the most noticeable difference in sharpness is marked with red rectangle in (a) and (b); a magnification of this area is presented in (f).

Table 1: Results of the full-length-film evaluation. The largest values for each metric are marked in bold. The positive and negative disparity values describe the strength of the "3D effect." High metric values for vertical disparity and for color and sharpness mismatch mean the film is not of high quality.

Film title	Pos. disparity (% of screen width)	Neg. disparity (% of screen width)	Vertical disparity (‰ of screen width)	Color mismatch	Sharpness mismatch
Avatar	1.13510	0.57940	0.05654	7.37639	8.71903
Dolphin Tale	0.79032	0.17189	0.59346	**29.22981**	11.93237
Galapagos: The Enchanted Voyage	**1.18344**	**1.38288**	**0.75352**	22.31848	12.37332
Hugo	0.58173	0.98574	0.18485	3.08822	9.32534
Into the Deep	**1.64675**	**2.09470**	0.62676	**27.20178**	**15.95367**
Pirates of the Caribbean: On Stranger Tides	0.61308	0.32259	0.12376	2.83375	5.36008
Resident Evil: Afterlife	0.62952	0.45495	0.32623	16.88972	8.67114
Sanctum	1.10979	0.80092	0.50303	8.12971	10.32646
Step Up 3D	0.85358	0.74246	**0.73433**	20.28920	**12.85025**
The Three Musketeers	0.81952	0.04928	0.19528	5.06023	5.11177

blocks. For the third step we estimate the value of the global vertical disparity as the median disparity among the reliable blocks.

Also, note that local vertical disparity is admissible—for example, in the case of toed-in cameras. The metric produces no "false alarms" in such a case because the median of the vertical disparity is close to zero when scenes are filmed properly using toed-in cameras.

During the filming stage, geometric distortions must be avoided by means of precise stereoscopic camera calibration. In the postprocessing stage, tools for correcting vertical disparity or inter-view tilt can be used.

We use per mil of screen width as the unit for this metric. The justification for our choice of screen width as the denominator is the same as that for horizontal disparity in Subsection 2.1: for home-theater content, the frame width is typically fixed, but the frame height may vary with the aspect ratio of the image.

3. RESULTS

To understand how often the problems described above appear in real videos, we performed quality evaluation for a set of recent Blu-ray 3D releases. We chose Blu-ray 3D as the content with the best publicly available quality. Since our selected releases are oriented toward home-theater displays, we assume the comparison of the films' depth budget is fair.

The list of evaluated films and their average metric values are presented in Table 1. Diagrams depicting the results of the evaluation are presented in Appendix A. The results of the evaluation demonstrate that the quality of the top films is increasing over time. Another trend is that recent films are more conservative in terms of depth budget compared with the oldest films in our test set.

Processing speed is about five days for a three-hour film using an Intel Core i7 processor running at 2.8GHz. Since we don't assume any temporal dependencies between results for different frames, processing can be parallelized effectively among several computers.

4. FURTHER WORK

Development of our proposed system can follow two main directions: research on new metrics and improvement of processing speed. To complete the set of metrics for processing captured video, development of a metric for stereo-window violation and a metric for time desynchronization must be undertaken, at a minimum.

Another major task in metric development is creation of metrics for converted stereo. Such metrics will be more complicated than those for captured stereo video because numerous issues arise in the 2D-to-3D conversion process. We plan to estimate some depth-map features and their correspondence with the source views. It will then be possible to estimate the quality of edge processing and occlusion filling. This kinds of metrics will enable complex quality assessment of converted films.

The main opportunity to improve the algorithm's processing speed is through porting of the most time-consuming tasks to a GPU.

Owing to the complexity of the problem studied in this work, we performed no complicated subjective testing nor any analysis of correlation between human perception of artifacts and our results. The results of human-perception tests may depend on the type and quality of the display device; also, the results are susceptible to viewers' visual acuity and their ability to perceive 3D. This topic should therefore be addressed separately. One of the first tasks in this case is to determine the threshold of artifact noticeability for each metric.

We plan to share the results of our work with professional stereographers and researchers to invite feedback on the problem and recommendations on how we can improve our system. Our final goal is creation of a powerful automatic system for S3D-content quality assessment; such a system may help improve the visual quality of future stereo films.

5. CONCLUSION

In this paper we presented our methodology for quality evaluation of stereoscopic films. We described algorithms for estimating horizontal-disparity distribution and for detecting artifacts in captured video. To demonstrate the utility of the proposed methodology, we evaluated recent Blu-ray 3D releases and presented the results. Possible directions of further work include development of new quality metrics, improvement of processing speed and performance of subjective testing to determine the noticeability of artifacts to viewers.

ACKNOWLEDGMENTS

REFERENCES

[1] Huynh-Thu, Q., Callet, P. L., and Barkowsky, M., "Video quality assessment: From 2D to 3D – challenges and future trends," in [*17th IEEE International Conference on Image Processing (ICIP)*], 4025–4028 (Sept. 2010).

[2] Boev, A., Hollosi, D., Gotchev, A., and Egiazarian, K., "Classification and simulation of stereoscopic artifacts in mobile 3DTV content," in [*Proc. SPIE Stereoscopic Displays and Applications XX*], **7237**(1), 72371F–1– –72371F–12 (2009).

[3] Tseng, K., Huang, W., Luo, A., Huang, W., Yeh, Y., and Chen, W., "Automatically optimizing stereo camera system based on 3d cinematography principles," in [*3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4, IEEE (2012).

[4] Guo-Yu, L., Xu, C., and Wei-Gong, Z., "A robust epipolar rectification method of stereo pairs," in [*Measuring Technology and Mechatronics Automation (ICMTMA), International Conference on*], **1**, 322–326, IEEE (2010).

[5] Zhou, J. and Li, B., "Image rectification for stereoscopic visualization without 3d glasses," *Image and Video Retrieval* , 495–498 (2006).

[6] Fecker, U., Barkowsky, M., and Kaup, A., "Time-constant histogram matching for luminance and chrominance compensation of multi-view video sequences," in [*Picture Coding Symposium (PCS)*], (November 2007).

[7] Rosenthal, J.-C., Zilly, F., and Kauff, P., "Preserving dynamic range by advanced color histogram matching in stereo vision," in [*Proceedings of International Conference on 3D Imaging (IC3D)*], (December 2012).

[8] Yu, J.-J., Kim, H.-D., Jang, H.-W., and Nam, S.-W., "A hybrid color matching between stereo image sequences," in [*3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4 (May 2011).

[9] Devernay, F., Pujades, S., and Ch.A.V., V., "Focus mismatch detection in stereoscopic content," in [*SPIE Proceedings on Stereoscopic Displays and Applications*], **8288** (Jan. 2012).

[10] Zilly, F., Müller, M., Eisert, P., and Kauff, P., "The stereoscopic analyzer–an image-based assistance tool for stereo shooting and 3D production," in [*IEEE International Conference on Image Processing*], 4029–4032 (2010).

[11] Luo, A.-C., Huang, W.-J., Chen, W.-C., Lin, C.-W., and Chen, S.-W., "A stereoscopic content analysis system with visual discomfort-aware," in [*Proceedings of International Conference on 3D Imaging (IC3D)*], (December 2012).

[12] Simonyan, K., Grishin, S., Vatolin, D., and Popov, D., "Fast video super-resolution via classification," in [*Proceedings of IEEE International Conference on Image Processing*], 349–352 (2008).

[13] Voronov, A., Vatolin, D., Sumin, D., Napadovsky, V., and Borisov, A., "Towards automatic stereo-video quality assessment and detection of color and sharpness mismatch," in [*Proceedings of International Conference on 3D Imaging (IC3D)*], (December 2012).

APPENDIX A. FILM COMPARISON DIAGRAMS
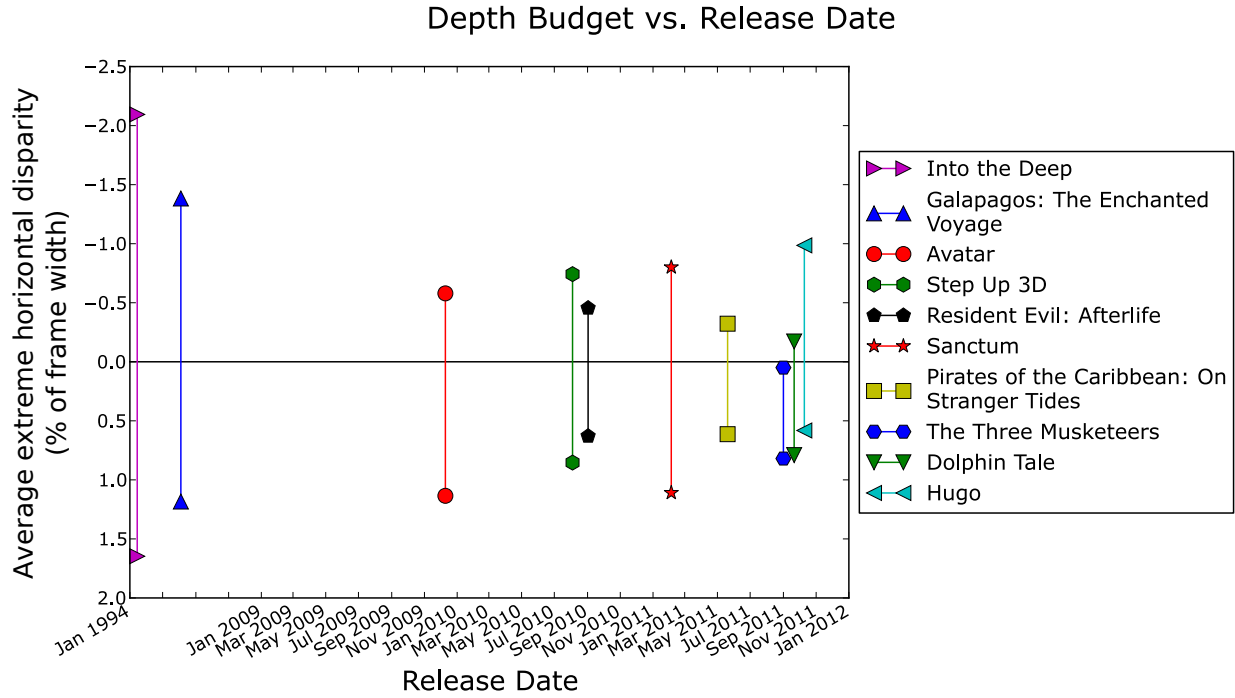
A.1 Average value diagrams

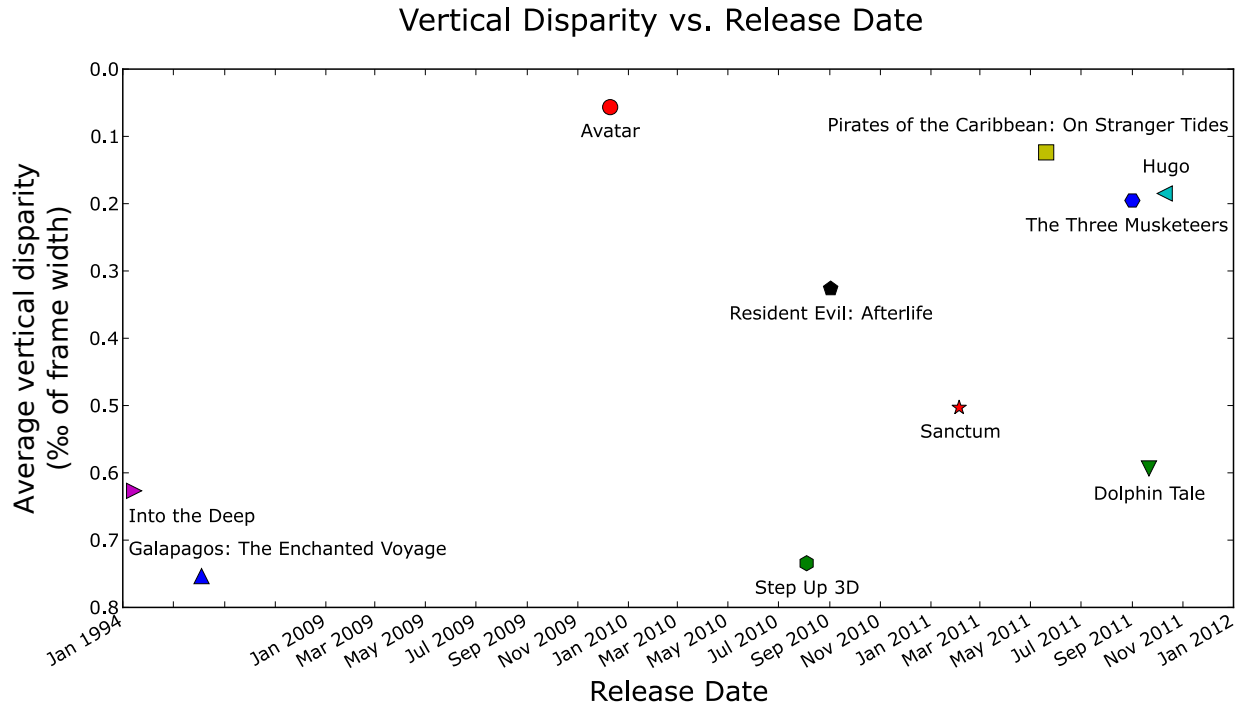Figure 5: Diagram illustrating average positive and negative horizontal disparity for the evaluated films.

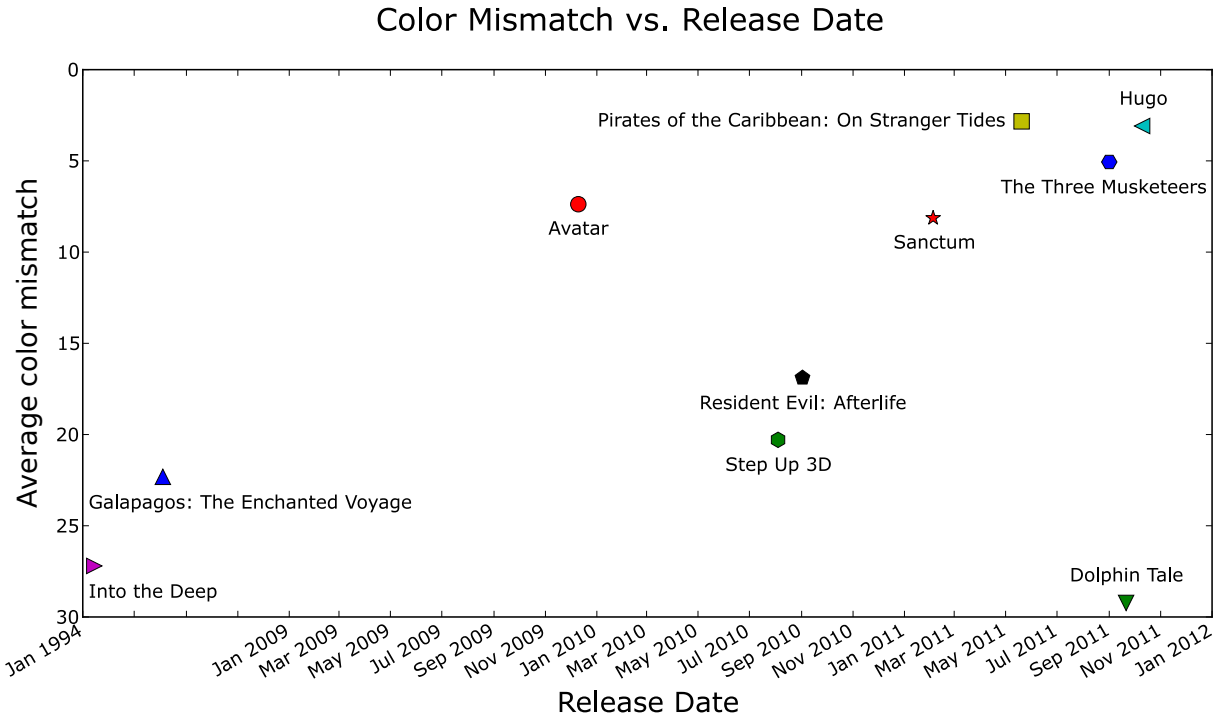Figure 6: Diagram illustrating average vertical disparity for the evaluated films.

Color Mismatch vs. Release Date

Figure 7: Diagram illustrating average color-mismatch metric value for the evaluated films.

Sharpness Mismatch vs. Release Date
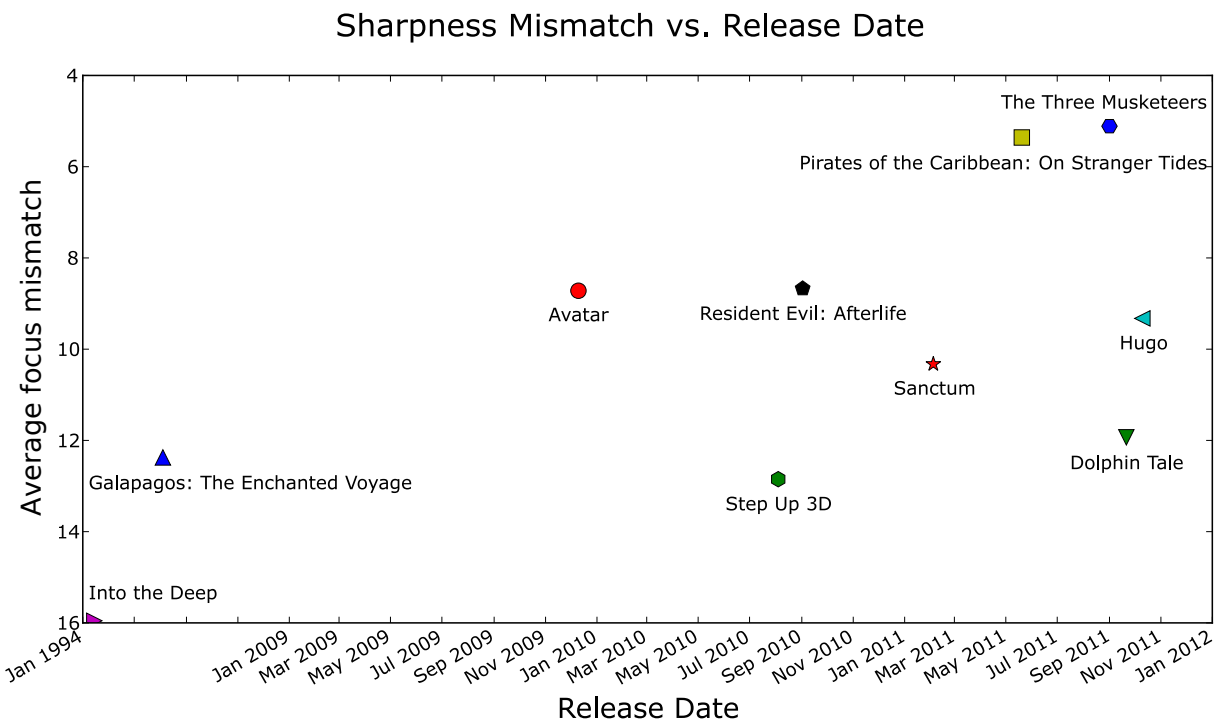
Figure 8: Diagram illustrating average sharpness mismatch value for the evaluated films.

A.2 Integral histograms

Integral histograms illustrate the number of good-quality and poor-quality scenes in each film. Figure 9 shows an example integral histogram for the artifact metric. This histogram comprises three films: A, B and C. Film A has more frames with low metric values. Since our distortion metrics yield greater values in cases of higher distortion, film A is thus better than films B and C.

Consider point P at coordinates (x, y), for example. According to the integral histogram, point P indicates that y percent of the frames in the film have a metric value no greater than x. This particular histogram indicates that 40% of the frames in film A have a value no greater than 0.1.
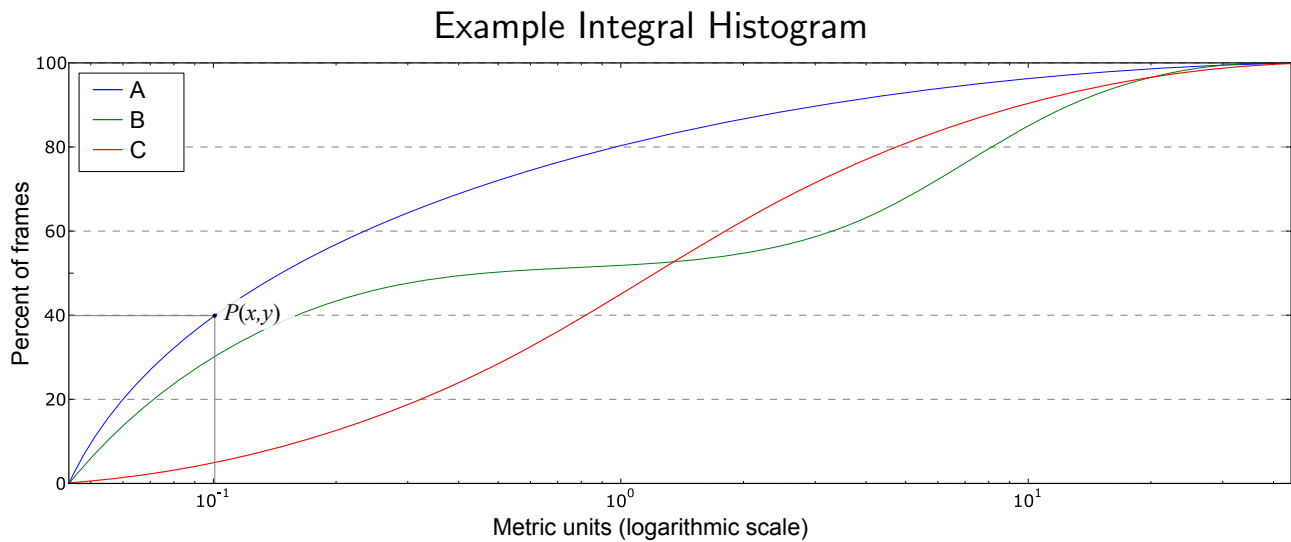
Figure 9: Example of an integral (cumulative) histogram.

In the case of using integral histograms for artifact analysis, the metric lines for films with better quality are "closer" to the top-left corner of graph. When analyzing depth budget, "flatter" films are lie "closer" to this top-left corner. To estimate depth budget for a given frame we use difference between extreme positive and negative disparity levels in that frame.

The intersection of lines B and C is also worth mentioning. This feature means that film B has more scenes with a low metric value, but it also has scenes with very high values. When analyzing film quality, these results tend to indicate that artifacts in film B are more noticeable.

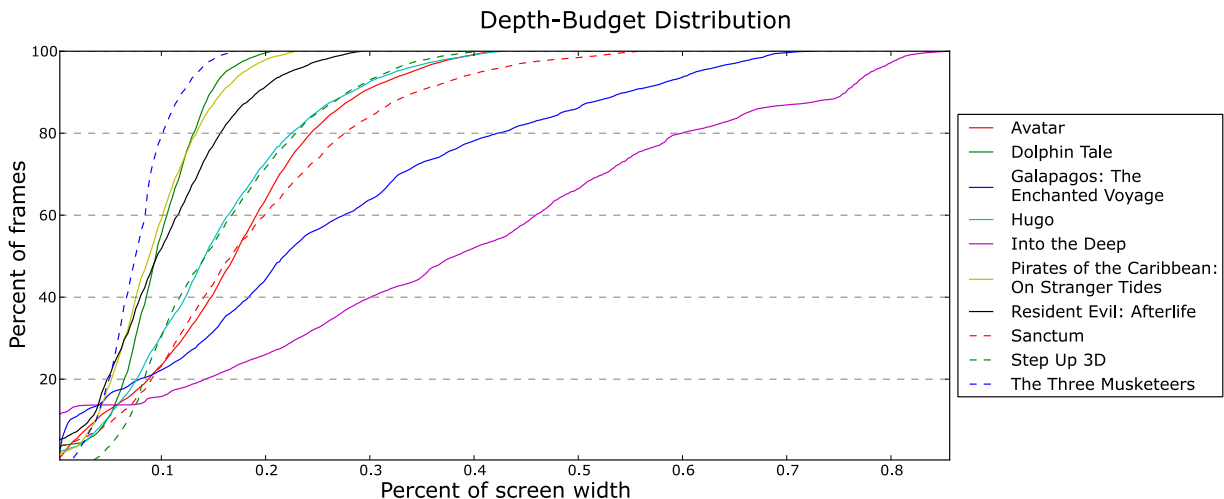Below we present an integral histogram for each metric.

Figure 10: Diagram illustrating distribution of depth budget for the evaluated films. Depth budget for each frame is presented as difference between extreme values of positive and negative disparities.
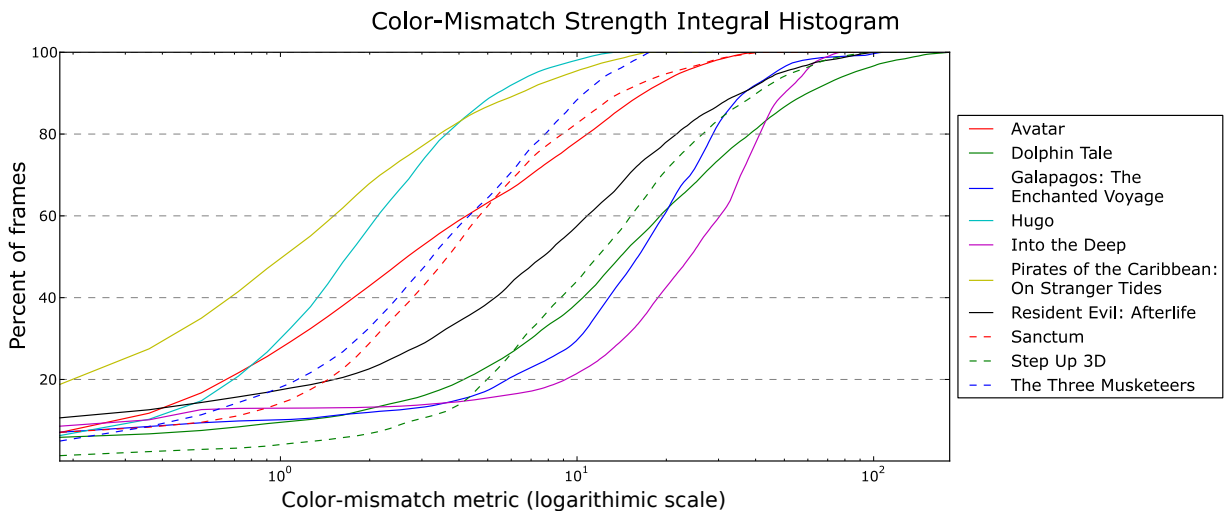

Figure 11: Diagram illustrating distribution of color-mismatch metric value for the evaluated films.
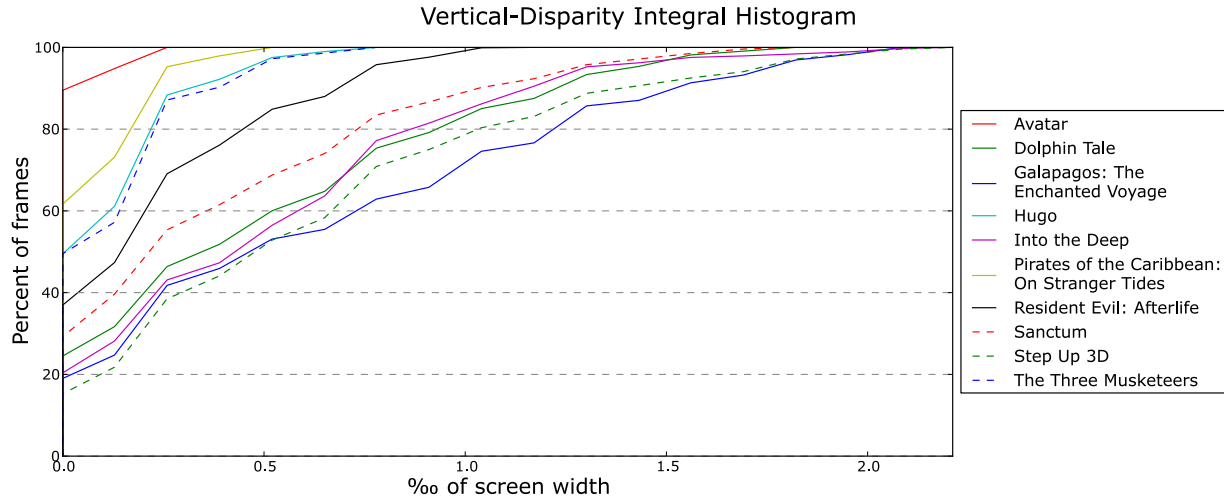
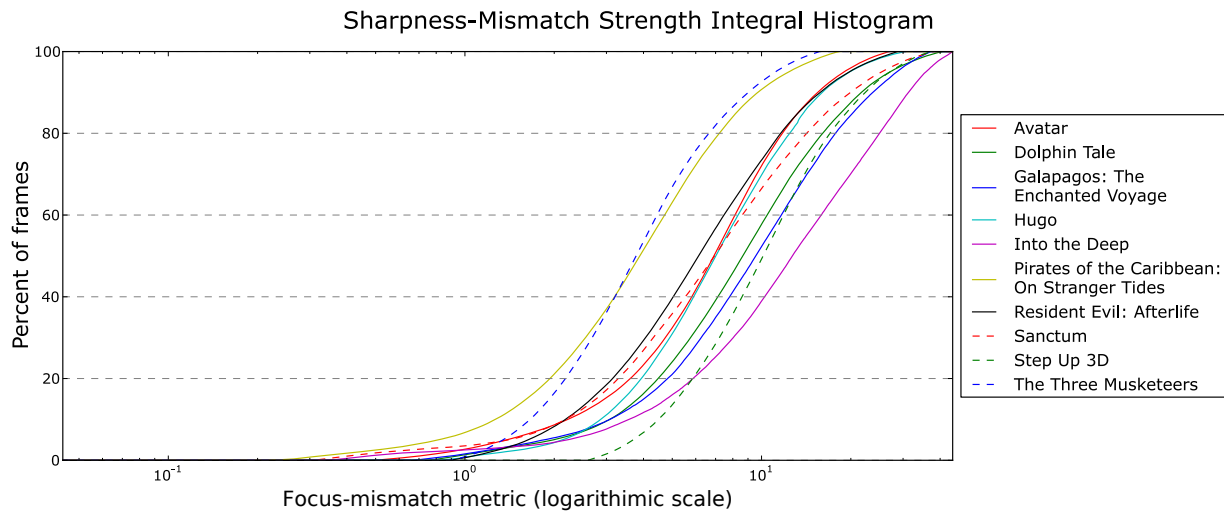Figure 12: Diagram illustrating distribution of vertical disparity for the evaluated films.

Figure 13: Diagram illustrating distribution of sharpness-mismatch metric value for the evaluated films.