DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439

Trends in S3D-Movie Quality Evaluated on 105 Films Using 10 Metrics

Dmitriy Vatolin, Alexander Bokov, Mikhail Erofeev, Vyacheslav Napadovsky;
Lomonosov Moscow State University, Moscow, Russia

Abstract

In this paper we present a large-scale analysis of S3D-movie technical quality spanning a large portion of stereoscopic-cinema history. We evaluated 105 Blu-ray 3D releases, including titles like the 1954 classic Dial M for Murder*, as well as contemporary stereoscopic productions like* Life of Pi *and* The Great Gatsby*. The analysis is based on objective quality metrics designed to detect specific types of artifacts, including swapped channels, inconsistencies between the stereoscopic views (color, sharpness and geometric as well as temporal asynchrony) and many others. The main challenges we had to overcome were the enormous amount of computational resources and disk space that such analyses require as well as algorithmic difficulties in developing some of the more advanced objective quality metrics. Our study clarifies the quality trends and problems of S3D movie production in general and provides a better understanding of how effectively quality control has been applied to particular movies compared with the overall trend.*

Introduction

Compared with traditional 2D-video production, creation of S3D video is an inherently tougher process that requires precise control of numerous technical quality parameters. Vertical disparities may originate from inconsistencies between camera positions or fields of view and from shooting with converged axes. Mismatches in color, luminance and sharpness between the views regularly occur when capturing S3D video. Most of these issues demonstrably affect the perceived quality of S3D content [1]. Insufficiently accurate depth maps or poor edge processing can lead to annoying artifacts as a result of 2D-to-3D conversion [2]. As our analysis shows, many types of artifacts still appear in recent S3D movies, producing visual discomfort for some viewers and impeding the widespread adoption of stereoscopic content. Decreasing the number of viewers who experience discomfort while viewing S3D movies is a major objective in advancing the industry.

Among the less common issues is channel mismatch—i.e., the result of either a complete swap of the left and right views or a partial swap involving only particular objects in a scene (Figure 2). It usually occurs because of postproduction errors and in some cases can lead to noticeable viewer discomfort. Although it is straightforward to fix, this artifact presents a major challenge with regard to automatic detection. Another relatively rare problem is temporal asynchrony between the stereoscopic views. It can be caused either by hardware defects (such as when the cameras fail to start recording in a synchronized manner), an unstable frame rate during capture or postproduction mistakes. This artifact can be especially annoying in scenes with fast-moving objects, owing to either vertical parallax if the object is moving vertically or to conflicting depth cues if the object is moving horizontally. Our analysis shows that in practice, the duration of most temporal shifts is significantly less than one frame (Figure 1).

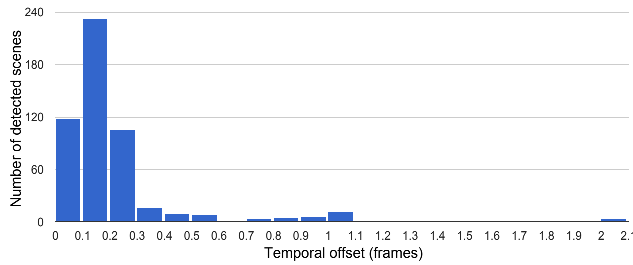


Figure 1: Distribution of temporal offsets between the left and right views for all detected temporal-asynchrony cases in 105 S3D movies.

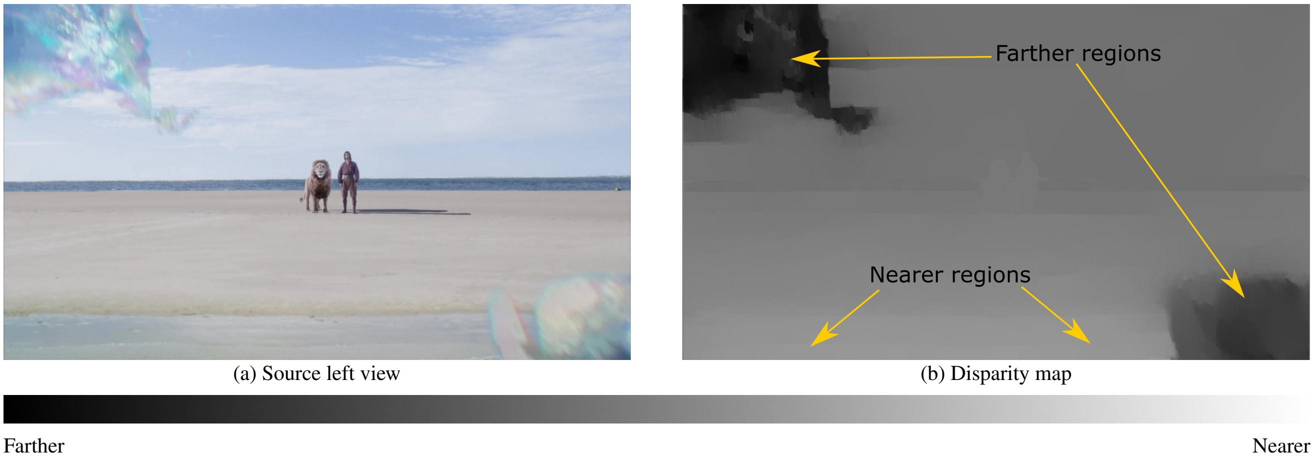
S3D cinematography also introduces a number of guiding principles that are absent from the traditional 2D medium. Following these principles improves the quality of the viewing experience according to subjective user studies [3]. Although most of these principles allow some variation in accordance with the artistic license and vision of the producer, we propose several simple objective metrics designed to roughly measure compliance with these principles. In particular we introduce metrics for estimating the quality of stereo-window-violation handling and of depth continuity, measured as the accumulated magnitude of depth jump cuts (a description of these principles appears in [4]).

Given that the history of stereoscopic cinema goes back almost a century, a natural assumption is that the technical quality of movies and compliance with respective guiding principles has improved as the field matured. The data we collected for our large-scale analysis of 105 S3D movies using 10 quality metrics enables us to test this hypothesis and extract the major trends in S3D-content quality. Moreover, it provides a reference point for evaluating particular movies according to their release date and budget.

Related Work

Several recent studies considered the problem of predicting visual discomfort caused by specific S3D artifacts. In [1], Khaustova et al. investigated how viewer annoyance depends on various technical parameters such as vertical disparity, rotation and field-of-view mismatches between the views, and color and luminance mismatches. They identified perceptual-acceptability thresholds for each of these parameters. The resulting objective metrics showed high correlation with the subjective ranks, but the authors obtained these results using a small data set of three syn-

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439



(a) Source left view (b) Disparity map

Figure 2: Example of a partial channel mismatch involving only certain objects in a scene (the waves in the top-left and bottom-right regions of the frame). The scene is from *The Chronicles of Narnia: The Voyage of the Dawn Treader* (courtesy of Fox 2000 Pictures).

thetic stereoscopic images. Chen et al. [5] proposed several objective metrics for luminance mismatch and evaluated their correlation with the results of the subjective experiment. They used a more diverse data set of 30 natural stereoscopic images. In [6], researchers analyzed an artifact specific to S3D video by evaluating visual discomfort caused by temporal asynchrony. This analysis only considered integer frame shifts, however, whereas fractional shifts significantly less than one frame predominate in practice (Figure 1).

Some researchers have proposed general methods for measuring stereoscopic-video quality that try to maximize correlation with the mean opinion score (MOS). Most of them, however, orient toward transmission-related issues, while we focus on problems introduced during production of S3D content. Silva et al. [7] proposed a no-reference metric that takes into account overall disparity distribution, a blockiness measure and the motion characteristics of the scene. According to their results, the metric has outperformed traditional 2D full-reference metrics on the COSPAD1 data set, which contains S3D sequences impaired by H.264 and MJPEG compression as well as by simple image-processing operations like downsampling and sharpening. Han et al. in [8, 9] present metrics that predict perceived S3D-video quality solely on the basis of transmission parameters like network packet-loss rate, bit rate and frame rate. The authors of [10] present a full-reference metric, having evaluated a large variety of measures and taking into account 2D-picture quality, binocular rivalry and depth-map degradation. They maximized the correlation with the MOS by using linear regression. An optimal feature combination achieved a Spearman correlation coefficient of 0.92 on a data set containing S3D sequences degraded by blocking and downsampling.

The metrics we employed in this study, however, are purely technical and are not intended for directly predicting perceptual quality. Dong et al. [11] present a wide range of technical-quality metrics, including vertical misalignment, rotation and zoom mismatches, and color mismatch. The authors used sparse SIFT matching to establish correspondence between the views. They identified vertical parallax as the most common vertical misalignment between matched feature pairs, they extracted scale mismatches locally from SIFT correspondence, and they estimated rotation mismatch globally through simple enumeration of possible values that minimize the difference between the rotated left and original right views. Also, [11] introduces a simple global color-mismatch measure that avoids taking into account the disparity map and the presence of half-occlusion areas. A similar approach appears in [12], where the authors compute the Pearson correlation coefficient between the histograms of the left and right views. Proposed in [13, 14] is a sharpness-mismatch metric that relies on estimating edge-width deviation between the left and right views.

Estimating temporal asynchrony between stereoscopic views can be seen as a special case of the more general task of spatio-temporal video-sequence alignment. Most algorithms that perform this general task do not consider subframe alignment [15, 16], which is necessary when working with S3D video. And the accuracy of those that do estimate subframe offsets [17] can be improved by imposing additional constraints specific to S3D video. Several proposed methods specifically aim to detect temporal asynchrony in stereoscopic sequences. Unfortunately, they either detect only integer frame offsets [6] or are simple asynchrony indicators that fail to specify the precise temporal offset between the views [18, 13].

Very few studies investigate the problem of detecting channel mismatch (swapped views) in stereoscopic sequences. A simple algorithm proposed in [19] is based on the fact that objects near the center or bottom of the screen are typically closer to the viewer than objects near the top or sides of the screen. A more advanced algorithm, presented in [20], uses the location of half-occlusion areas to more accurately predict channel mismatch. Another approach that analyzes half-occlusions appears in [21].

Sudden, discontinuous disparity changes can be a source of visual discomfort to viewers watching S3D movies [22]. This issue is especially problematic when depth jump cuts (scene cuts with significant disparity changes) appear repeatedly, forcing the viewer to frequently adjust the vergence response. A proposed computational model for predicting the vergence-adjustment time after jump cuts appears in [23]. The authors constructed a bilinear model based on two main factors: target disparity and disparity change. Several technical metrics proposed in the literature measure the magnitude of depth jump cuts. Winkler [12] uses

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439

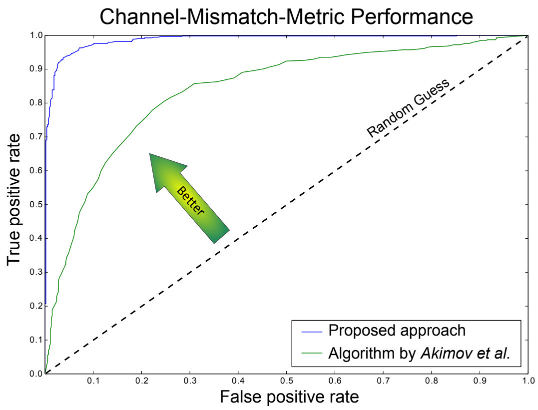


Figure 3: ROC curves of the proposed approach and the algorithm described in [20].

the Pearson correlation coefficient between disparity histograms of consecutive frames. Delis et al. [24] define separate thresholds for the difference in mean positive and mean negative disparity between consecutive frames to detect depth jump cuts.

Finally, the problem of automatically detecting stereo-window violations has received some attention in recent studies. In [13, 25], the authors propose a simple method of detecting objects in negative-disparity space that touch either the left or right boundary of the frame. A more advanced method proposed in [26] also considers whether the violating object is in focus (it does not flag blurred objects as stereo-window violations). But neither of these methods estimates already existing floating windows, which are a common tool for fixing stereo-window violations in stereoscopic productions [4].

Methodology

This study employed 10 objective quality metrics to analyze the following technical parameters of S3D video:

1. Extreme horizontal-disparity values;
2. Vertical parallax;
3. Color mismatch between views;
4. Sharpness mismatch;
5. Field-of-view/scale mismatch;
6. Rotation mismatch;
7. Temporal asynchrony;
8. Channel mismatch (swapped views);
9. Depth continuity (cumulative magnitude of depth jump cuts);
10. Stereo-window violation.

Descriptions of the first three metrics are available in [27]. To estimate sharpness mismatch we use an improved version of the metric described in [27] that is more robust to large half-occlusion areas and color mismatch between views (see Appendix A in [28]). We briefly describe the remaining six metrics below.

Scale- and rotation-mismatch metrics enable more-precise assessment of geometric inconsistencies compared with our vertical-parallax metric. They estimate the parameters of a simplified affine transform model between the left and right views that considers only vertical offsets between the corresponding points (x, y) and (x', y'):

$$y' = ax + by + c, \tag{1}$$

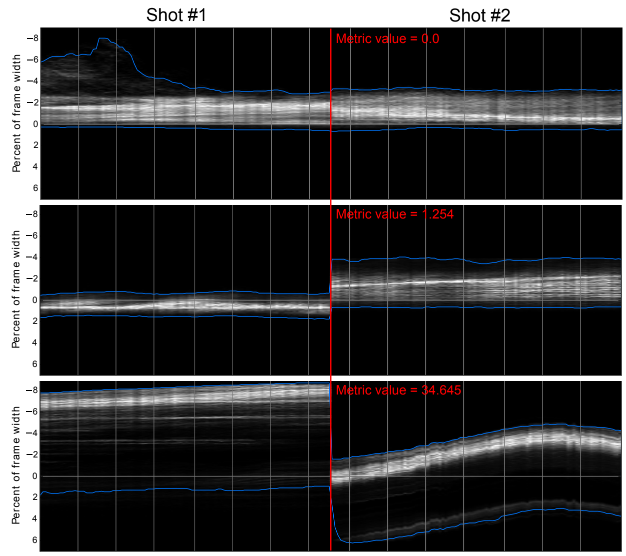


Figure 4: Illustration of the depth-continuity metric. Each chart is a per-frame depiction of disparity distributions, with lighter colors indicating more-frequent disparity values in a given frame. The metric enables detection of cases where the disparity distribution changes too dramatically between subsequent shots (depth jump cuts).

where a, b, and c define the respective amount of rotation, scale and vertical shift between the views. We obtain the initial dense correspondence between the views using a block-based approach [29] and filter out unreliable matches using the LRC criterion [30]. Finally, we fit the resulting model to the filtered matches using RANSAC [31].

The main idea of the proposed **temporal-asynchrony** metric is to analyze the correlation between the vertical projection of an object's speed and the object's vertical disparity. In particular, let v_y be the vertical projection of the vector matching a point in the current frame to the previous frame and let d_y be the vertical disparity of the same point. Then we propose the following model:

$$d_y = \Delta t v_y + d, \tag{2}$$

where d is the global vertical offset between the views and Δt is the temporal offset. We use RANSAC [31] to fit this model to the data collected over the whole scene (we assume each scene has a constant temporal offset).

The **channel-mismatch** metric is a notably improved version of the algorithm proposed in [20]. It produces significantly fewer false-positive and false-negative errors (Figure 3) and enabled us to analyze all 105 movies with a moderate amount of manual effort. We compared the algorithms using a data set containing 1,000 scenes, each 30 frames long, taken from five S3D movies; the views in half of the scenes are swapped. Using the algorithm from [20] would be completely impractical owing to the enormous amount of false positives.

To estimate **depth continuity** we must first detect depth jump cuts and estimate their magnitude (Figure 4). We do so by estimating the earth mover's distance (EMD) between the disparity histograms of consecutive frames, which we can compute very efficiently in the case of 1D histograms [32]. Moreover, we intro-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439

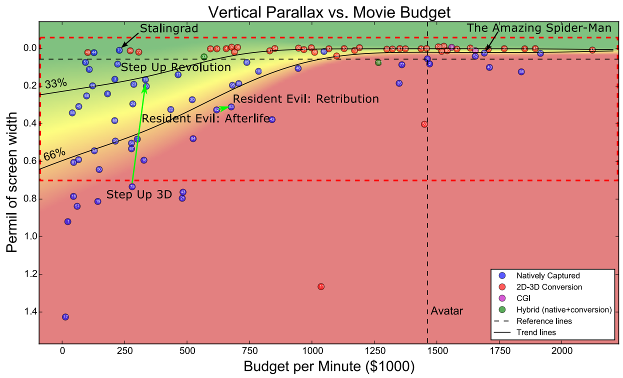


Figure 5: Chart illustrating how technical quality (measured using the vertical-parallax metric) depends on the movie budget. A clear trend of improving quality with increasing budget is apparent.

duce a simple visual model that uses two priors: a center prior (viewers are more likely to look at the center of the screen) and a sharpness prior (viewers are more likely to converge on objects that are in focus). So we compute the magnitude of a jump cut in the following way:

$$d_i = max(\text{EMD}(h^S_{i-1}, h^S_i) - d_0, 0), \tag{3}$$

where h^S_i and h^S_{i-1} are disparity histograms of the current and previous frames weighted by their respective saliency maps and d_0 is a constant that prevents accumulation of small disparity variations between consecutive frames (we use $d_0 = 0.75$, assuming disparity values are measured in percent of screen width).

To detect **stereo-window violations** we analyze disparity distributions in narrow stripes along the left and right borders of the frame. We also evaluate the width of the floating window. A stereo-window violation is detected when the disparity of the closest object near the frame edge is lower (closer to the viewer) than the disparity of the floating window. To compute the magnitude (noticeability) of the detected stereo-window violation, we take into account the size of the area occluded by the frame edge, as well as its brightness and texture.

Using these metrics, we analyzed 105 S3D Blu-ray releases. Some of the problems we identified may be present only in the Blu-ray release, with the cinema versions being unaffected. The nature of many artifacts, however, makes it unlikely that they can be introduced during an ostensive postproduction stage done exclusively for a Blu-ray release. Our analysis addresses all types of S3D movies (natively captured, post-converted, hybrid and fully rendered in 3D), but the comparison of converted and captured S3D movies is generally unfair, as it fails to take into account problems that are specific to converted movies. It does clearly show how 2D-to-3D conversion helps eliminate some artifacts that commonly appear in captured S3D movies, however.

The main technical obstacle that we had to overcome was the enormous amount of computational resources and disk space that such an analysis requires. Evaluation of one movie can take up to four weeks and consume over 40GB just for the source Blu-ray. To maximize efficiency, we developed a system that efficiently distributes the computations across a cluster of up to 17 computers working in parallel. All of the metrics allow independent per-scene processing of the input video, so parallelizing them is trivial. This system enabled us to finish the evaluation of 105 movies in six months.

Title	Release date	Budget ($M)	Budget per minute ($K/min)
Step Up 3D	August 2010	30	280
Step Up Revolution	July 2012	33	333
Resident Evil: Afterlife	September 2010	60	618
Resident Evil: Retribution	September 2012	65	677
The Amazing Spider-Man	June 2012	230	1,691
Stalingrad	October 2013	30	229
Avatar	December 2009	237	1,463

Table 1: Selected movies illustrating overall trends in the charts.

Results

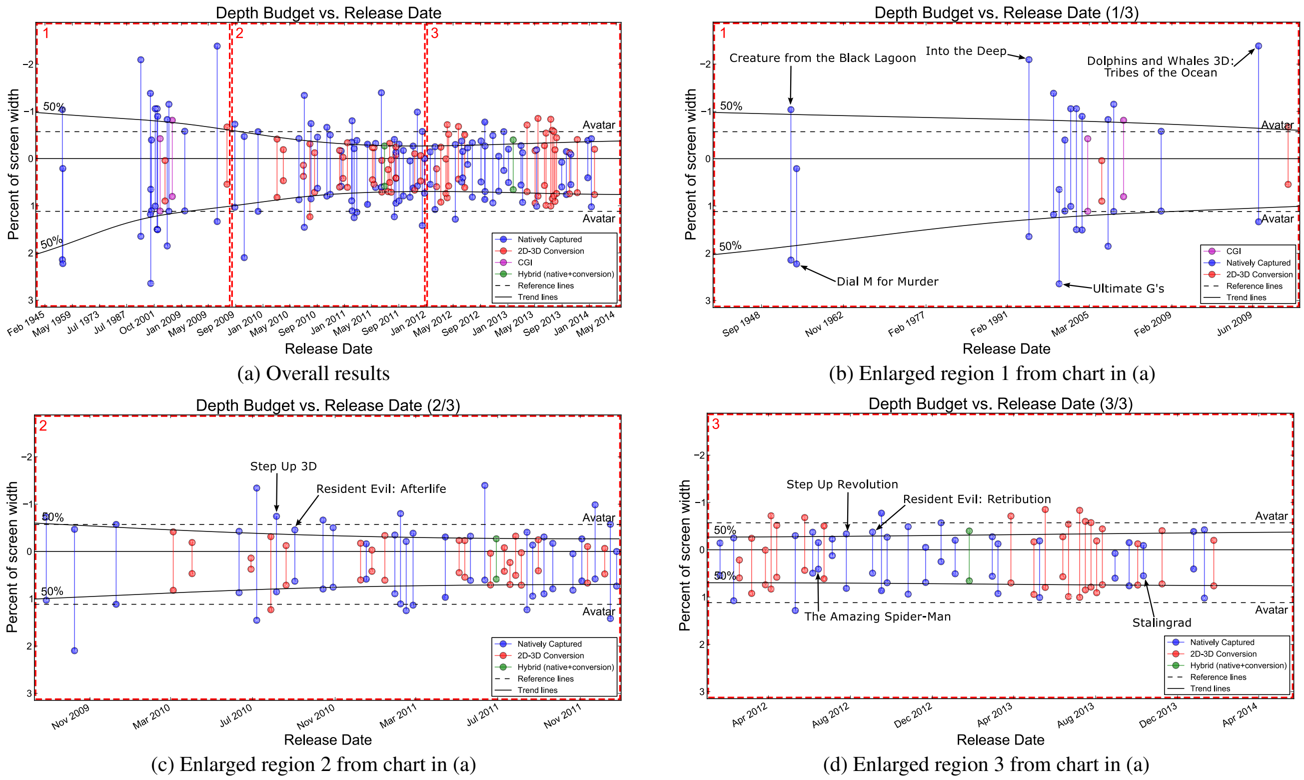
In this section we present a number of overall comparison charts for different movies according to their release date and quality as measured by one of our technical-quality metrics. We selected several movies to illustrate the general trends (Table 1). For the *Step Up* and *Resident Evil* franchises we can assess how quality changes with time at the same budget (approximately \$300.000 and \$600.000 per minute, respectively). *The Amazing Spider-Man* is an example of a high-budget blockbuster, *Stalingrad* is intended to represent a good-quality low-budget movie and *Avatar* serves as a reference point for the whole analysis. Figure 5 illustrates the relationship of these movies to others in budget per minute. Many more charts and ratings, as well as full disambiguation of all the movies in these charts, is available in a separate technical report [33].

Depth-Budget Trends

Figure 6a depicts the average disparities of the closest/farthest objects in the movies we evaluated. Negative values correspond to objects in front of the screen, and positive values correspond to objects behind the screen. We measure disparity in percent of screen width. A 1% positive disparity on a 6.5m-wide screen corresponds to a distance of 6.5cm between projections of the same object in the left and right views. Such disparity already represents binocular infinity for most viewers, since interpupillary distance seldom exceeds 6.5cm. Some of the movies, however, consistently place objects beyond this threshold, effectively forcing viewers to diverge their eyes in order to fixate on these far-away objects in the theater. As a reference in all our charts we present the evaluation results for the film *Avatar* along with trend lines computed as a certain percentile of all the results in a sliding window.

As Figure 6a demonstrates, the average depth budget slowly decreased until the spring of 2012 and has remained relatively constant since. In a theater, the audience usually favors more-conservative depth budgets, as excessively large disparity values

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439



(a) Overall results

(b) Enlarged region 1 from chart in (a)

(c) Enlarged region 2 from chart in (a)

(d) Enlarged region 3 from chart in (a)

Figure 6: Results of depth-budget analysis. The charts depict two values for each movie: the average disparity of the closest point in a scene and the average disparity of the farthest point.

can lead to visual discomfort. Moreover, viewers perceive movies with small depth budgets as lacking in 3D effect when displayed on home 3D screens—another possible reason for the observed trend, as small depth budgets can serve as a tool for compelling people to prefer the theater over 3DTV systems, and they allow producers to use the same depth budget for the cinema and Blu-ray versions. Note that 2D-to-3D conversion usually results in lesser depth budgets according to our results (Figure 6a). This situation may be due to the fact that increasing the depth budget makes the 2D-to-3D conversion process more labor intensive owing to the bigger occlusion areas that require filling.

Figure 6b depicts an enlarged version of the first chart region. It shows that among the 105 movies we examined, those with the greatest average negative disparity are *Into the Deep* and *Dolphins and Whales 3D: Tribes of the Ocean*. These films predominantly consist of underwater shots, which are very challenging when it comes to proper stereoscopic capturing. Figure 6c shows that *Avatar*, which earned over $2 billion at the box office, spawned a whole wave of movies with lesser depth budgets and lower S3D quality in general, as the following sections of this discussion illustrate. The depth budgets of *Step Up 3D* and *Resident Evil: Afterlife*, however, are only slightly lower than that of *Avatar*. As Figure 6d shows, the 2012 sequels to these movies became approximately 15–20% "flatter" than the originals, following the overall trend. *Stalingrad* is over two times "flatter" than *Avatar*. One of the highest-budget movies in the group, *The Amazing Spider-Man* is 2.5 times "flatter" than *Avatar*.

We should emphasize that we evaluated Blu-ray releases of these films, which are intended for home viewing on a relatively small screen. The above-mentioned trend of decreasing depth budgets likely played a major part in overall disappointment with S3D movies, largely thanks to an unimpressive 3D effect when viewed on a home 3DTV system.

Common S3D-Artifact Trends

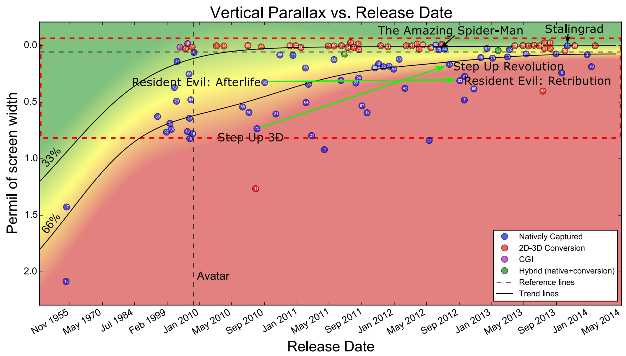
As briefly mentioned above, the $2.7 billion box-office earnings of *Avatar* led to number of negative consequences:

- A serious lack of high-quality equipment for shooting in S3D;
- More importantly, a dearth of S3D professionals—both highly specialized individuals trained to perform certain tasks in the S3D-shooting pipeline and individuals with a more general understanding of the whole pipeline. Consequently, some people could have claimed more stereography experience than they actually had;
- Presumably, numerous S3D movies released in 2010 and, partially, 2011 were produced in a hurry to cash in on *Avatar's* success but payed little attention to quality control, resulting in relatively low standards.

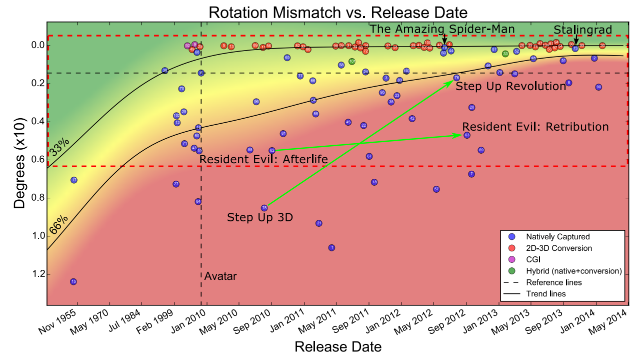
Undoubtedly, some positive effects accrued as well:

- Several companies invested in development of next-generation digital S3D cameras that offer previously unavailable technical characteristics;
- Many studios and professionals gained experience in shooting S3D. Novel, previously impractical production pipelines emerged, involving 24/7 processing of shots on several different continents thanks to broadband Internet connections;
- New S3D postproduction tools appeared, providing capa-

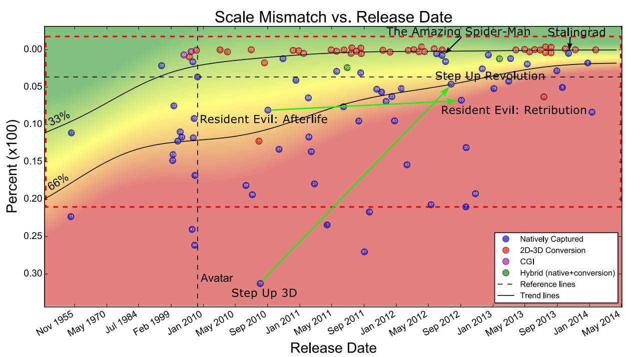
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439



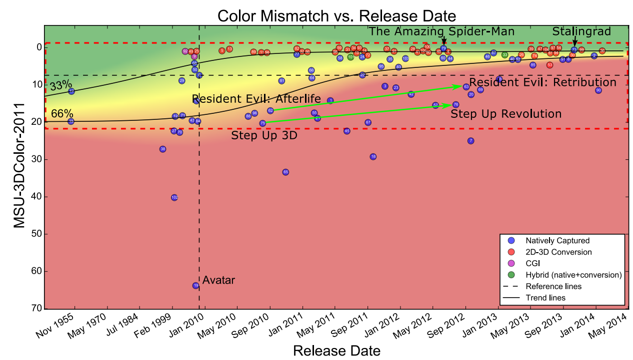
(a) Average vertical parallax



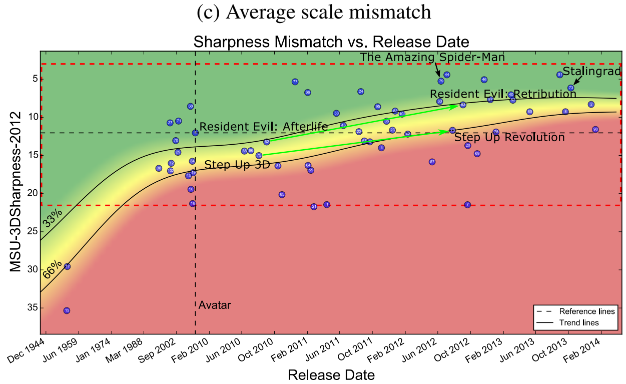
(b) Average rotation mismatch



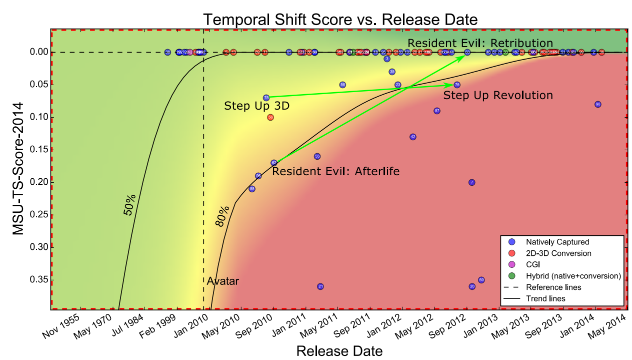
(c) Average scale mismatch



(d) Average color mismatch



(e) Average sharpness mismatch



(f) Cumulative temporal-asynchrony score (scene duration$\times$offset)

Figure 7: Results of common-S3D-artifact analysis depicting vertical parallax (a), rotation mismatch (b), scale mismatch (c), color mismatch (d), sharpness mismatch (e) and temporal asynchrony (f).

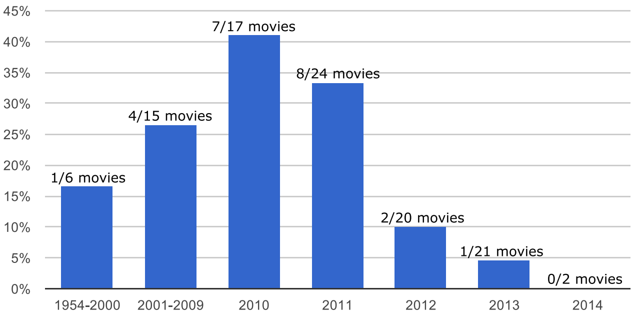
bilities that were previously considered impossible and enabling both real-time control of certain quality parameters and automatic correction of some stereoscopic artifacts.

We analyzed how these factors affected S3D movies as measured by our technical-quality metrics. First, we assessed the geometric inconsistencies between the views, which depend on both the capture and postproduction pipelines. Figure 7a demonstrates how the average vertical parallax of S3D movies has changed with time. We make several observations:

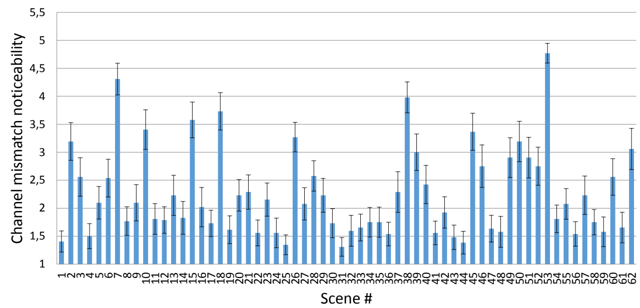
- *Resident Evil: Retribution* improved on its predecessor, but not enough relative to general trends. The film moved beyond the 66th-percentile line, indicating it is among the 34% worst S3D movies of its time period according to this technical parameter;
- *Step Up Revolution* improved significantly compared with *Step Up 3D*—even more than the general trend;
- *The Amazing Spider-Man* has excellent technical quality according to our vertical-parallax measurements, which is unsurprising given that it is one of the biggest-budget entries among the 105 S3D films that we evaluated;
- Surprisingly enough, *Stalingrad* demonstrated technical quality on par with that of *The Amazing Spider-Man* despite having a lower budget than all the above-mentioned movies;
- *Avatar* had excellent technical quality when it was released, surpassing all the natively captured S3D movies that preceded it according to our vertical-parallax measurements. Today, however, producers can achieve higher quality in much lower budgets, with *Stalingrad* being a prime example.

Figure 7b illustrates the average rotation mismatch (in degrees) between the views. According to this parameter, *Step Up Revolution* improved much more than *Resident Evil: Retribution*

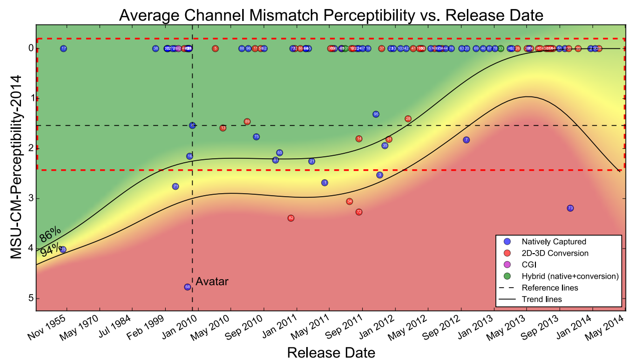
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439



(a) Percent of movies containing at least one detected scene with channel mismatch



(b) Subjective noticeability of channel mismatch in detected scenes. Each value is the average of subjective ranks from 49 people.



(c) Average channel-mismatch noticeability

Figure 8: Statistics of scenes with detected channel mismatch from 105 S3D movies.

compared with their respective predecessors. By 2014, more than half of captured S3D movies were better than *Avatar* according to our rotation-mismatch metric. Figure 7c illustrates scale mismatch, measured in percent. *Step Up 3D* earned the worst result among all 105 movies by this metric. Notably, *Step Up Revolution* managed to decrease the scale mismatch by a factor of six relative to its predecessor.

Color mismatch between stereoscopic views seldom causes immediate viewer discomfort, allowing the anaglyph format to exist. But prolonged exposure to stereoscopic content that has color mismatch, coupled with other issues, can lead to accumulation of visual discomfort. Figure 7d shows the results of our analysis. Both *Resident Evil: Retribution* and *Step Up Revolution* demonstrate a moderate improvement in the number of color inconsistencies. Other movies have improved more, however, effectively moving the mentioned sequels to the "red" zone of movies with the worst technical quality according to this parameter. Worth noting is that *Stalingrad* consistently ranks among the movies with the best technical quality despite its relatively small budget.

Moderate sharpness mismatch is often acceptable, as it has little effect on binocular fusion. But beyond a certain threshold, which may vary from person to person, sharpness mismatch can be a source of visual discomfort. It is also substantially harder to fix in postproduction compared with previously discussed problems, and it leads to qualitatively different trend behavior (Figure 7e). We make several additional observations:

- Both *Resident Evil: Retribution* and *Step Up Revolution* demonstrate improvements precisely along the general trend line;
- *Stalingrad* is very close to *The Amazing Spider-Man* in sharpness mismatch despite having an 87% lower budget;
- Despite having many CGI scenes, *Avatar* demonstrates a mediocre result; by 2014, nearly all films have better technical quality according to this parameter.

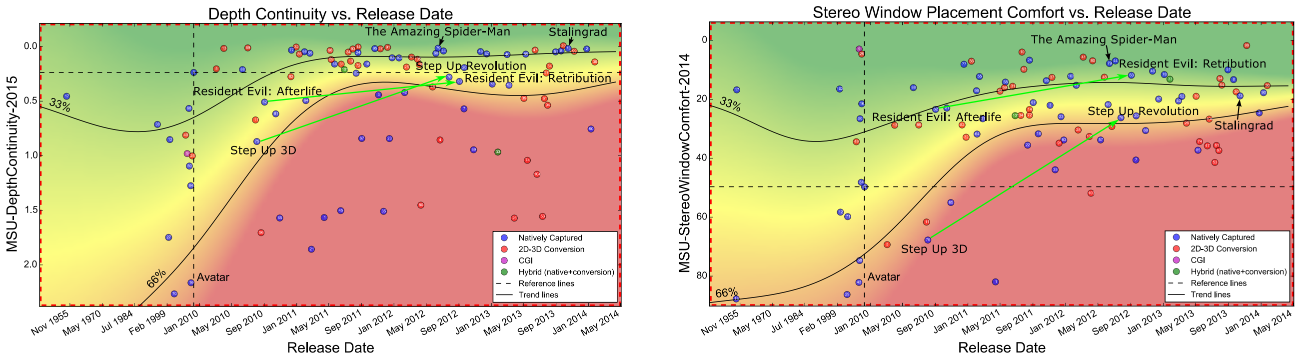
Our temporal-asynchrony metric enabled us to detect a total of 515 scenes in 27 movies with confirmed temporal offsets of 0.1 to 2.0 frames between the views. The situation has substantially improved since the 1950s, and temporal asynchronies are relatively rare in contemporary movies (Figure 7f). To rank the films we use a temporal-asynchrony score computed as the sum of temporal offsets multiplied by the respective scene length for all scenes in a movie. *Resident Evil: Afterlife* had some temporal-asynchrony problems, but the sequel completely addressed them. *Step Up Revolution* has shown a minor improvement compared with its predecessor. Both *The Amazing Spider-Man* and *Stalingrad* lacked any detected cases of temporal asynchrony.

Channel-Mismatch Trends

Using our channel-mismatch metric, we detected 65 manually verified scenes with channel mismatch in 23 different movies. So the probability of a movie having at least one scene demonstrating this artifact is 22%. Figure 8a provides more-detailed statistics. Interestingly, the peak value of 41% occurred in 2010—another testament to the negative effect that *Avatar* had on the industry. But the situation improved very quickly, and in 2013 only 1 of the 21 evaluated movies contains a scene with channel mismatch. The emergence of automatic quality-control systems will likely solve the problem entirely.

Different cases of channel mismatch may have drastically different noticeability among viewers, depending on a range of factors including scene length, depth budget and brightness levels.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439



(a) Depth continuity
(b) Average quality of stereo-window-violation handling

Figure 9: Results of analysis by additional quality metrics.

To estimate the discomfort caused by different scenes with channel mismatch, we conducted an experiment, involving 59 human subjects. We composed a video sequence of 56 detected scenes exhibiting channel mismatch (we excluded some detected scenes that were highly similar to others in the group). Moreover, each scene in the sequence was preceded and succeeded by scenes with the correct channel order to better simulate real viewing conditions and to provide an additional reference point for viewers.

We repeated each of these three-scene fragments three times before giving the subjects some time to rest and fill out the questionnaire. We asked them to rank each scene on a scale of 1 (imperceptible) to 5 (severe discomfort). The inclusion of several additional scenes with no channel mismatch ensured that the subjects ranked scenes correctly. We showed half of the subjects the sequence in reverse order to suppress the possible influence of a given scene on the subjective mark for the subsequent one. We excluded the results of 10 subjects who demonstrated maximal deviation from statistical average or who gave high marks to the control scenes containing no channel mismatch. Figure 8b presents our results.

As expected, detected scenes are very different in terms of perceived discomfort. Dark and "flat" scenes are virtually indistinguishable from scenes containing no channel mismatch (for instance, scene #31, from *The Three Musketeers*), whereas channel mismatch in some scenes (scene #53, from *Sharks 3D*) are unanimously considered to be very annoying. Some trends can be extracted from Figure 8c, which depicts the average noticeability of scenes with channel mismatch for different movies. With very few exceptions, the issue is practically nonexistent in contemporary S3D movies.

Trends in Depth Continuity and Stereo-Window-Violation Handling

We also evaluated how well different movies comply with common guiding principles of stereography. Figure 9a illustrates trends in depth continuity measured as the cumulative magnitude of all depth jump cuts in a movie. We make several observations:

- A clearly visible trend line indicates that recent S3D movies, on average, pay more attention to maintaining depth continuity between subsequent shots;
- Both *Resident Evil: Retribution* and *Step Up Revolution* demonstrate similar improvements roughly along the general trend line;
- As in virtually all of our other measurements, *Stalingrad* and *The Amazing Spider-Man* place among the best movies according to our depth-continuity metric.

To measure the quality of stereo-window-violation handling, we assess the average noticeability of stereo-window violations throughout a movie using our metric. Figure 9b presents the results. The following points are noteworthy:

- Both *Resident Evil: Retribution* and *Step Up Revolution* have improved much compared with their predecessors—more than most of the other movies we evaluated;
- *Avatar's* relatively mediocre result, even despite its 2009 release date, can be explained by the fact that the film used no floating windows, making stereo-window violations practically inevitable;
- Despite enabling significantly better performance on many other parameters, 2D-to-3D conversion is similar to native capturing when it comes to stereo-window-violation handling. This result is unsurprising, as the issue relates more to the guiding principles of stereography, which may be subjective and therefore at the mercy of artistic license and the vision of the producer. Thus, the quality of stereo-window-violation handling does not depend on the production method and is mostly defined by the conscious decision (or absence of such) to address this problem in one way or the other. Such a decision often must be made as early as the preproduction stage.

Conclusion

In this paper we presented the results of a large-scale technical-quality analysis of 105 S3D movies using 10 objective metrics. To summarize, we observe positive trends in all of the measured technical-quality parameters, indicating that S3D movie-production pipelines have considerably improved over the past six years. In particular, we point out the following:

- A new generation of S3D cameras has emerged with previously impossible technical characteristics;
- Accessible broadband Internet connections have transformed the organization of production pipelines, dramatically increasing their efficiency;
- New real-time quality-control systems can roughly estimate some technical parameters and perform corrections on the fly;
- Fundamentally new tools for semiautomatic 2D-to-3D con-

version have made converted S3D movies much cheaper and faster to produce while providing competitive quality, considering the numerous problems of native S3D capturing;

- Nearly all the common problems can now be fixed in postproduction—including sharpness mismatch and temporal asynchrony between the views, which until just recently were considered unfixable;
- Finally and perhaps most importantly, many new people joined the industry, working professionally with the new cameras, on-set quality-control systems, 2D-to-3D conversion software and new tools for correcting problems in postproduction.

The observed trends suggest that the technical quality of S3D movies will continue to improve, gradually eliminating the visual discomfort that some people experience owing to these kinds of technical issues. The quality of S3D projection systems, however, can also be a major contributor to the overall quality of the experience, but such factors are beyond the scope of this study.

Acknowledgments

This study was funded by the RFBR under research project 15-01-08632 A.

References

[1] D. Khaustova, J. Fournier, E. Wyckens, and O. Le Meur, An objective method for 3D quality prediction using visual annoyance and acceptability level, IS&T/SPIE Electronic Imaging, pp. 93910P–93910P. (2015).

[2] A. Bokov, D. Vatolin, A. Zachesov, A. Belous, and M. Erofeev, Automatic detection of artifacts in converted S3D video, IS&T/SPIE Electronic Imaging, pp. 901112–901112. (2014).

[3] C. W. Liu, T. H. Huang, M. H. Chang, K. Y. Lee, C. K. Liang, and Y. Y. Chuang, 3D cinematography principles and their applications to stereoscopic media processing, Proceedings of the 19th ACM International Conference on Multimedia, pp. 253–262. (2011).

[4] B. Mendiburu, 3D movie making: stereoscopic digital cinema from script to screen, CRC Press, 2009.

[5] J. Chen, J. Zhou, J. Sun, and A. C. Bovik, Binocular mismatch induced by luminance discrepancies on stereoscopic images, IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. (2014).

[6] L. Goldmann, J. S. Lee, and T. Ebrahimi, Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection, IEEE International Conference on Image Processing (ICIP), pp. 3241–3244. (2010).

[7] A. R. Silva, M. E. V. Melgar, and M. C. Farias, A no-reference stereoscopic quality metric, IS&T/SPIE Electronic Imaging, pp. 93930B–93930B. (2015).

[8] Y. Han, Z. Yuan, and G. M. Muntean, No reference objective quality metric for stereoscopic 3D video, IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–6. (2014).

[9] Y. Han, Z. Yuan, and G. M. Muntean, Extended no reference objective quality metric for stereoscopic 3D video, IEEE International Conference on Communication Workshop (ICCW), pp. 1729–1734. (2015).

[10] F. Battisti, M. Carli, A. Stramacci, A. Boev, and A. Gotchev, A perceptual quality metric for high-definition stereoscopic 3D video, IS&T/SPIE Electronic Imaging, pp. 939916–939916. (2015).

[11] Q. Dong, T. Zhou, Z. Guo, and J. Xiao, A stereo camera distortion detecting method for 3DTV video quality assessment, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–4. (2013).

[12] S. Winkler, Efficient measurement of stereoscopic 3D video content issues, IS&T/SPIE Electronic Imaging, pp. 90160Q–90160Q. (2014).

[13] M. Liu, I. Mademlis, P. Ndjiki-Nya, J. C. Le Quintrec, N. Nikolaidis, and I. Pitas, Efficient automatic detection of 3D video artifacts, IEEE International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6. (2014).

[14] M. Liu and K. Muller, Automatic analysis of sharpness mismatch between stereoscopic views for stereo 3D videos, International Conference on 3D Imaging (IC3D), pp. 1–6. (2014).

[15] Y. Caspi and M. Irani, Spatio-temporal alignment of sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(11), 1409–1424 (2002).

[16] F. Diego, J. Serrat, and A. M. Lopez, Joint spatio-temporal alignment of sequences, IEEE Transactions on Multimedia, 15(6), 1377–1387 (2013).

[17] B. Meyer, T. Stich, M. A. Magnor, and M. Pollefeys, Subframe temporal alignment of non-stationary cameras, British Machine Vision Conference (BMVC), pp. 1–10. (2008).

[18] M. Liu, and P. Ndjiki-Nya, Automatic detection of temporal synchronization mismatches between the stereoscopic channels for stereo 3D videos, IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–6. (2014).

[19] M. Knee, Getting machines to watch 3D for you, SMPTE Motion Imaging Journal, 121(3), 52–58 (2012).

[20] D. Akimov, A. Shestov, A. Voronov, and D. Vatolin, Automatic left-right channel swap detection, International Conference on 3D Imaging (IC3D), pp. 1–6. (2012).

[21] J. Bouchard, Y. Nazzar, and J. J. Clark, Half-occluded regions and detection of pseudoscopy, International Conference on 3D Vision (3DV), pp. 215–223. (2015).

[22] S. Poulakos, G. Roethlin, A. Schwaninger, A. Smolic, and M. Gross, Alternating attention in continuous stereoscopic depth, Proceedings of the ACM Symposium on Applied Perception, pp. 59–66. (2014).

[23] T. J. Mu, J. J. Sun, R. R. Martin, and S. M. Hu, A response time model for abrupt changes in binocular disparity, The Visual Computer, 31(5), 675–687 (2014).

[24] S. Delis, N. Nikolaidis, and I. Pitas, Automatic detection of depth jump cuts and bent window effects in stereoscopic videos, IEEE 3D Image/Video Technologies and Applications Workshop (IVMSP), pp. 1–4. (2013).

[25] S. Delis, N. Nikolaidis, and I. Pitas, Automatic 3D defects identification in stereoscopic videos, IEEE International Conference on Image Processing (ICIP), pp. 2227–2231. (2013).

[26] Y. Nazzar, J. Bouchard, and J. J. Clark, Detection of stereo window violation in 3D movies, ACM SIGGRAPH 2014 Posters, p. 61. (2014).

[27] A. Voronov, D. Vatolin, D. Sumin, V. Napadovsky, and A. Borisov, Methodology for stereoscopic motion-picture quality assessment, IS&T/SPIE Electronic Imaging, pp. 864810–864810. (2013).

[28] D. Vatolin, A. Voronov, D. Sumin, G. Rozhkova, V. Napadovsky, A. Borisov, and M. Arsaev, VQMT3D project stereo-film-quality analysis report 2, 2013. [Online]. Available: http://compression.ru/video/vqmt3d/second-report.

[29] K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, Fast video super-

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-439

resolution via classification, IEEE International Conference on Image Processing (ICIP), pp. 349–352. (2008).

[30] G. Egnal, M. Mintz, and R. P. Wildes, A stereo confidence metric using single view imagery with comparison to five alternative approaches, Image and Vision Computing, 22(12), 943–957 (2004).

[31] M. A. Fischler and R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, 24(6), 381–395 (1981).

[32] S. H. Cha and S. N. Srihari, On measuring the distance between histograms, Pattern Recognition, 35(6), 1355–1370 (2002).

[33] D. Vatolin, V. Napadovsky, A. Bokov, A. Shalpegin, V. Yanushkovsky and S. Lavrushkin, VQMT3D project stereo-film-quality analysis report 10, 2016. [Online]. Available: http://compression.ru/video/vqmt3d/report10.

Author Biography

Dmitriy Vatolin received his M.S. degree in 1996 and his Ph.D. in 2000, both from Moscow State University. Currently he is head of the Video Group at the CS MSU Graphics & Media Lab. His research interests include compression methods, video processing, 3D video techniques (depth from motion, focus and other cues, video matting, background restoration, high quality stereo generation), as well as 3D video quality assessment (metrics for 2D-to-3D conversion artifacts, temporal asynchrony, swapped channels and many others). His contact email is dmitriy@graphics.cs.msu.ru.

Alexander Bokov received his M.S. degree in computer science from the Moscow State University (2015), where he is currently a Ph.D. student. His research interests include S3D video processing and quality assessment. His contact email is abokov@graphics.cs.msu.ru.

Mikhail Erofeev received his M.S. degree in computer science from the Moscow State University (2013), where he is currently a Ph.D. student. His research interests include various image and video processing techniques, machine learning and video compression. He is co-founder of a public benchmark for video matting algorithms: http://videomatting.com/. His contact email is merofeev@graphics.cs.msu.ru.

Vyacheslav Napadovsky received his M.S. degree in computer science from the Moscow State University (2014). His research interests include S3D quality estimation and artifact correction. His contact email is vnapadovsky@graphics.cs.msu.ru.