

Perceptually-motivated Stereoscopic Film Grain

Krzysztof Templin¹ Piotr Didyk² Karol Myszkowski¹ Hans-Peter Seidel¹

¹MPI Informatik

²MIT CSAIL

Abstract

Independent management of film grain in each view of a stereoscopic video can lead to visual discomfort. The existing alternative is to project the grain onto the scene geometry. Such grain, however, looks unnatural, changes object perception, and emphasizes inaccuracies in depth arising during 2D-to-3D conversion. We propose an advanced method of grain positioning that scatters the grain in the scene space. In a series of perceptual experiments, we estimate the optimal parameter values for the proposed method, analyze the user preference distribution among the proposed and the two existing methods, and show influence of the method on the object perception.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image generation—Display algorithms, Viewing algorithms

1. Introduction

Analog photographs and films often feature a random high-frequency texture, commonly called *film grain*. It is a by-product of the photographic process, in which crystals of silver salts that were exposed to light are transformed into larger groups of metallic silver or dye clouds, creating an image of visible granularity. Film grain is often considered an artifact and is removed in post-production. This, however, is not always an easy task, since there is no simple way of discriminating between random noise and fine details of the photographed objects. Moreover, grain is sometimes intentionally preserved or even added to evoke certain mood, stylize, or imitate the look of old movies. The fundamental requirement in such cases is to retain a uniform look of grain between various regions of the film. For example, when computer-generated elements are inserted into a scene, matching film grain has to be added. This allows to seamlessly integrate different types of content, without creating a clear distinction between them, which would be perceived as an artifact. For similar reasons, grain has to be added also to fully synthetic shots, because of a possible mismatch with the previous, real shot. Even if the objects in the scene are real and are merely to be processed (e. g., resized), the grain has to be removed, and added back afterwards [Sey11a]. Thus, grain management, i. e., the set of techniques for removing, adding, and matching the grain, is a significant part of the film post-production process.

A noisy pattern similar to grain can also appear in digital photography, however it is often recognized as less appealing

than analog grain. On the other hand, pictures taken with the sensor set to lower sensitivities can look too clean. Therefore, film grain can be added to mask digital noise or compensate for “too synthetic” look [KMAK08]. The idea of adding grain is not limited to photography or film, but also appears in computer games [Gia]; e. g., the best-selling game *Limbo*, which uses strong film grain as a means of stylization.

Grain in stereoscopic 3D Grain application has been identified as a significant problem in the 3D film post-production process: for example, the VFX company Pixomondo spent weeks on R&D just to address the issue of grain in the Oscar-winning film *Hugo* [Sey11b]. The difficulty is due to the interplay between the left- and the right-eye image. If the same grain pattern is added to both channels, it is fused by the observer, and has the depth of the screen plane. This leads to an unpleasant shower-door effect, and causes double vision if the distance between the screen plane and the scene is large. Another option is to add two uncorrelated grain patterns, in agreement with what happens when two cameras are used. However, only limited amounts of uncorrelated grain can be tolerated [LL96], because presence of many unmatched features impedes fusion and leads to visual discomfort. Binocular rivalry may occur, and cause characteristic “shiny look” (Fig. 1, top row). The last option is to project grain on the surface of the objects, i. e., display it at the same depth as the object it occludes. This technique does not have disadvantages of the two previous ones, and is a natural choice especially in imagery created in the process

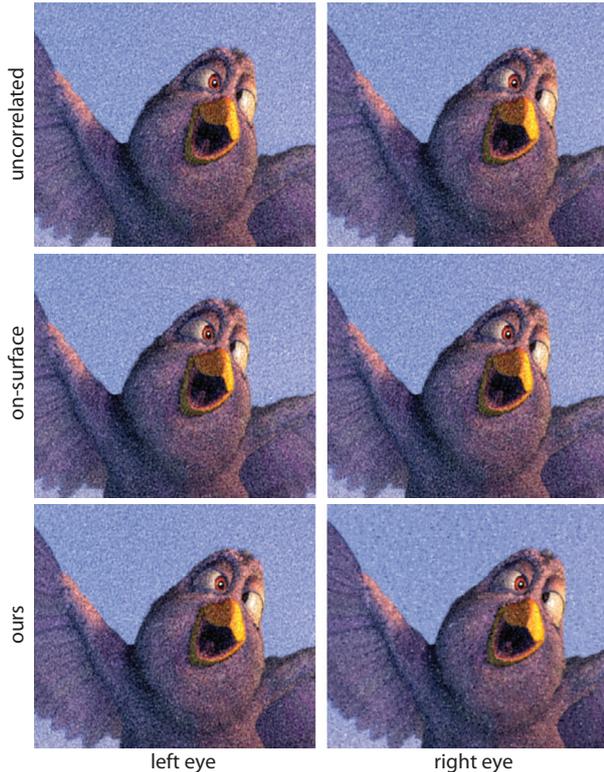


Figure 1: Film grain overlay in stereoscopic 3D. Grain that is added independently in each eye is hard to fuse and causes discomfort. In extreme cases, binocular rivalry appears, and the image looks “shiny” (top row). Projecting grain on the surface does not ensure medium-scene separation (middle row). Our technique ensures that grain is separated from the scene, but is easy on eyes (bottom row). Use uncrossed (parallel) free fusion to see the examples.

Bird scene copyright: Blender Foundation (www.bigbuckbunny.org).

of 2D-to-3D conversion, since the grain can be displaced together with the objects and does not require much additional attention. However, this approach creates impression that the grain belongs to the objects’ texture, and emphasizes any imperfections of depth (Fig. 1, middle row). Conversion from 2D does not preclude usage of uncorrelated grain, however it is not an easy task to remove all existing grain, and thus some portion of it may remain on the surface. The industry standard is to use uncorrelated grain, projected grain, or combination of both [Sey11b, Sey12, Rid11]. Winter and Gandolph [WG13] build on the idea of projected grain, and propose how to handle grain in the case of uncertain depth values in the stereoscopic content.

Our contribution We propose a new approach to adding grain, in which the input grain pattern is decomposed into particles and distributed in depth (Fig. 1, bottom row). We

draw inspiration from the way other film artifacts are treated during 2D-to-3D conversion: lens flares, or scratches and bigger dust particles found in old films are usually placed somewhere between the objects and the observer. To our knowledge, however, it has not been proposed so far to treat film grain in the same way.

We motivate our choice by the need of medium-scene separation: There is a distinction between the mental image of a depicted object and its depiction, and one cannot see both at the same time [Gom00]. Projecting grain on the surface of objects violates this distinction in a certain way – instead of a stereoscopic *grainy depiction* of an object we obtain a stereoscopic depiction of a *grainy object*. By detaching the grain from the objects, we make an effort to restore, at least partially, the medium-scene separation which is disrupted when moving from two-dimensional imagery to stereoscopic 3D. Additional benefit of our approach is that we avoid emphasizing potential S3D artifacts, such as unnatural flatness or depth map errors. Lastly, our approach can improve the stereoscopic composition of a scene: In traditional films, one should avoid “visual clutter”, as it leads to a feeling of uneasiness in the audience. In S3D this rule is reversed – if there are too few objects in the scene, stereoscopic depth cues will be too sparse, and the overall look will be less intense [Men09]. Since our grain introduces additional details in depth, it can help to avoid this problem.

2. Related work

Adding noise can help hide banding artifacts [DF03] or enhance perceived sharpness [JF00, KAK09] of the image. The human visual system (HVS) tends to naturally mask repetitive signals through adaptation processes that lead to increasing contrast detection threshold for such signals. This way the effective noise visibility is reduced while the salience of novel image content is enhanced [FJ05].

Stereoscopically displayed volumetric point clouds are common data representation in immersive virtual reality systems developed for medical imaging, scientific visualization, and volumetric rendering applications. Wang et al. [WPF10] observe that by increasing the point density or size the ability to explore 3D environment might be deteriorated due to occlusions. Our goals are quite different as stereoscopic grain is not intended as a means to convey any specific information, but rather to accentuate rich stereoscopic appearance of the scene.

Procedural noise is an important tool to add visually rich appearance to synthetic images [LLC*10]. Geigel and Musgrave [GM97] presented a model for simulating the photographic process on digital images. Stephenson and Saunders [SS07] described the synthesis of film grain based upon its noise-power spectrum. De Stefano et al. [DSCW06] proposed a method based on a causal auto-regressive model to generate plausible-looking grain patterns given input samples

of existing grain. In our work we focus on *adding* grain, and thus, we assume that the grain pattern is already given.

Adding film grain to images can be seen as an NPAR-style operation. There are a number of papers dealing with the problem of stylizing stereoscopic imagery [NAK12, SG05, SG04, KLKL13]; however, they focus on minimizing conflicts between the left and the right eye, and no effort is made to separate the stylization and the objects in depth. In the context of grain application, these algorithms are therefore analogous to on-surface grain. Lee et al. [LKKL13] found that in stereo line drawing brush stroke texture stylization enhances the depth impression with respect to plain lines.

3. Background

In this section, we provide perceptual background on binocular vision of stimuli, which show structural similarity to film grain. This way we are able to justify our design choices in Sec. 4 concerning the grain representation, which enables its comfortable viewing as a volumetric structure that features distinct depth properties with respect to scene surfaces.

Grain perception as 3D structure shows a number of analogies to depth perception in random-dot stereograms (RDSs) [Jul64], where binocular correspondence between dots is found without any explicit prior reference to a specific object recognition. In both cases, such correspondence can be found only through local pooling over the dot patterns, as each dot, when considered independently, could be matched to a large number of its counterparts in the other eye. Lankheet and Lennie [LL96] investigated various factors that can affect the HVS sensitivity to binocular correlation detection, which is required for depth recovery in the stereoscopic dot structure. They considered the dot life time as short as 26 ms and did not observe any improvement in the correlation sensitivity when the dots have been displayed for longer times. This suggests that binocular correlation processing well integrates location-varying information in successive frames for dynamic RDSs. Moreover, such time-varying fresh patterns of dots, which represent consistently the same disparity relationships, reduce a chance for a false disparity match in the neuronal receptive field, as it is unlikely that at the next frame the new dot pattern will support again the same false match [CD01, p. 217]. Also, the overall dot density does not seem to affect in any significant way the correlation performance, at least when the dot density is beyond 40 dots/deg² [LL96, Fig. 5]. All these observations apply to our film grain approach, where a new dot pattern is generated for each frame with the dot life time of at least 20 ms (assuming the framerate 50 fps or less), and typical dot density falling into the range 75–550 dots/deg² (estimated by counting the local extrema of the grain pattern).

The problem of stereo-transparency perceived in surfaces defined solely by disparity in RDSs has been investigated [AT88, TAW08, TWA10], where one of the key issues is the visibility of distinct transparent layers. Tsirlin et al. [TAW08,

Fig. 9] found that even three layers cannot be visually separated for the dot density higher than 8–10 dots/deg² per layer. Moreover, the visual separability of the layers is significantly deteriorated when the number of layers increases or dot patterns overlap between layers [TWA10], and most importantly when the inter-layer disparity drops below 1.9 arcmin [TAW08]. Since the density of grain dots is relatively high, the layered grain representation composed of several layers becomes a simple alternative to a full volumetric structure. We pursue this design option in Sec. 4, as the layering approach enables simple real-time GPU implementation, which is important in the context of computer games.

Relatively little is known on the perception of stereoscopic volumes of dots. Recently, Goutcher et al. [GOW12] investigated the HVS sensitivity to changes in the range and distribution of disparity-defined volumes of dots, and observed that for many ranges dots drawn from the Gaussian distribution could not be distinguished from an entirely uniform distribution. They concluded that the HVS uses an impoverished representation of the structure of stereoscopic volumes. This means that using more sophisticated distributions is not likely to have much visual impact. Therefore, in this work we always assume the uniform dot density allocation, and all our efforts to improve the appearance of stereoscopic grain are focused on modulating the thickness of its volumetric structure (Sec. 4).

4. Stereoscopic grain

The input to our algorithm are the left- and right-eye images L , R , together with the film grain pattern G to be applied, and the dense correspondence map $d: \mathbb{N}^2 \rightarrow \mathbb{R}$ between L and R . For any pixel position \mathbf{p} in R , $[\mathbf{p}_x + d(\mathbf{p}), \mathbf{p}_y]$ is the corresponding position in L . The grain is applied to the image using an application operator \oplus , which is typically a weighted addition, with the weights dependent on the pixel intensities in the input image.

The output is a modified grain pattern G' , such that the stereo pair $(L \oplus G, R \oplus G')$, gives impression of grain floating in space. To achieve this goal, the grain pattern needs to be re-interpreted as a collection of shapes in 3D space, that appears exactly as G when seen by the left eye. Based on that interpretation G' is determined. We proceed in two steps: (1) the grain pattern is segmented into individual grains, that are afterwards assigned to n different layers; (2) these layers are then appropriately stacked in depth, with increasing distance from the surface of the objects.

Grain segmentation In this step every pixel of the grain pattern G is assigned to one of the layers G_1, G_2, \dots, G_n . For any pixel \mathbf{p} , $G_i(\mathbf{p})$ equals $G(\mathbf{p})$ if \mathbf{p} has been assigned to layer i , and 0 otherwise. The assignments are made using a similar approach to *watershed by flooding* introduced by Baucher et al. [BL79]. First, we detect local luminance minima and maxima in G using a 3×3 min- and max-filter, and

assign them to layers by random. Next, the assignments are propagated iteratively. In each iteration, pixels that have been already assigned to layers propagate their assignments to their immediate unassigned neighbors. If at any iteration two or more pixels try to propagate to the same pixel, the one with the closest luminance value has the precedence. Since the spread between grains is usually in the order of several pixels, only few iterations are needed to assign all pixels to layers. An example result of this algorithm is shown in Fig. 2.

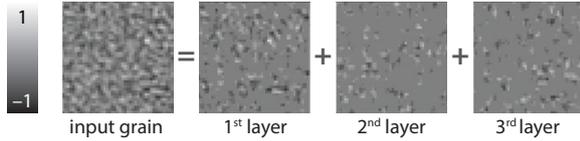


Figure 2: Each pixel of the grain pattern is assigned to one of the layers.

Layer stacking Now, the layers can be distributed in depth. The baseline distribution is obtained by putting

$$G'_b(\mathbf{p}) = \bigoplus_{i=1}^n G_i(\mathbf{p}_x + d(\mathbf{p}) - \frac{i}{n} \cdot \alpha, \mathbf{p}_y),$$

where n is the number of layers, and α is a parameter defining the thickness of the “grain cloud”: When $\alpha = 0$, grain is placed on the surface of the objects, for $\alpha < 0$ it appears to be inside, and for $\alpha > 0$ it surrounds them. The greater α the thicker the cloud around the objects.

Using constant α results in a regular distribution of film grain. In some cases, however, this may be not the best solution, therefore, we allow replacing α with a smoothly varying activity map $A: \mathbb{N}^2 \rightarrow [\alpha_{\min}, \alpha_{\max}]$, which maps a position in the image to a desired grain-cloud thickness. The α_{\min} and α_{\max} parameters are derived experimentally in Sec. 6 A thick grain cloud can obscure small depth details in the original scene, due to the *disparity masking* phenomenon [HR02, Chapter 19.6.3d], where the perception of a disparity corrugation is affected by another, superimposed signal. Therefore, it is necessary to modulate α value taking into account the scene geometry, and use a smaller value in regions with a high disparity variation. Bigger values of α may be used in flat regions to maximize depth impression, and counteract objectionable flatness (e. g., cardboarding effect or lack of details). An important observation is that masking affects mostly signals of similar spatial frequencies. As grain adds mostly high frequency disparity corrugations, A needs to account only for those. Additionally, A does not need to account for very high spatial frequencies (above 5 cpd) because those have a negligible effect on disparity perception [Ty175]. As a result, we first need to separate the signal that should be considered by the function A . We do it using a simplified version of the binocular disparity model presented by Didyk et al. [DRE*11]. The vergence angles are computed

separately for each location in the scene assuming that the observer verges on it perfectly. Thus, the correspondence map d is converted to a vergence map v , operating in visual angles instead of pixel shifts. Here, we follow terminology from perception literature, where *disparity* is defined as difference of *vergence angles* [HR02, Fig. 19.1]. Then, the relevant disparity signal is separated by a band-pass filter with cut-off frequencies ϕ_L and ϕ_H . One could consider a full frequency decomposition to multiple, narrow frequency bands, as it was done in the original disparity model. However, we found that our solution is sufficient and more practical. It is also motivated by the fact that the HVS has only a limited number of visual channels that are tuned to different disparity frequencies. Although the individual channel bandwidth has not been clearly established, the existing estimate suggest the range of 2–3 octaves [HR02, Chapter 19.6.3d]. We found $\phi_L = 0.625$ and $\phi_H = 5$ cpd (3 octaves) to give good results.

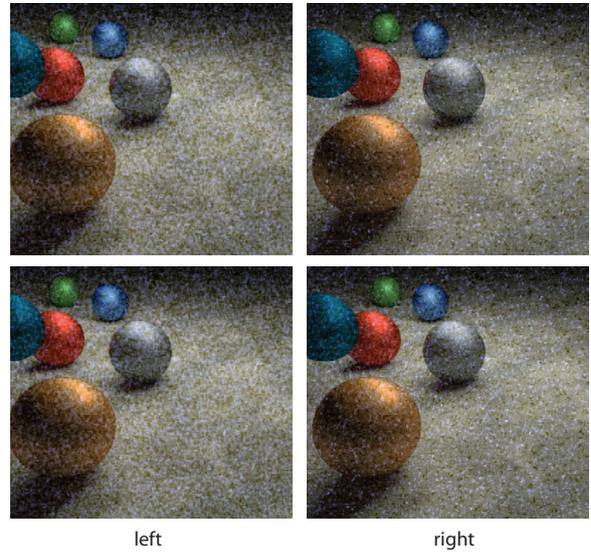


Figure 3: In the regions with high disparity variation, a thick grain cloud may attenuate perceived distances in depth (top). The activity map detects such regions and reduces the thickness accordingly (bottom). Note, how the distances between the balls are better preserved. The thickness of the grain cloud on the right side remains unchanged.

The band-limited vergence map contains signal whose perception may be affected by the additional grain disparity. At this point we are not interested in exact disparity values, but rather in regions where the thickness of the grain cloud needs to be attenuated due to high vergence variations in the original image. Therefore, we apply thresholding at the amplitude θ (we used 2 arc min), and apply low-pass filter with a cut-off frequency $0.5 \cdot \phi_L$. We denote the result as \hat{v} , and use it to modulate the activity map:

$$A = \alpha_{\max} - \hat{v} \cdot (\alpha_{\max} - \alpha_{\min}).$$

Finally, the resulting grain distribution is defined as

$$G'(\mathbf{p}) = \bigoplus_{i=1}^n G_i(\mathbf{p}_x + d(\mathbf{p}) - \frac{i}{n} \cdot A(\mathbf{p}), \mathbf{p}_y),$$

and is illustrated in Fig. 4. See Fig. 3 for a comparison of pictures with and without the activity map.

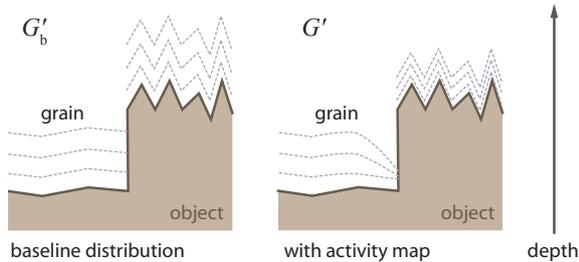


Figure 4: A thick grain cloud suspended above the object may mask small depth details of the geometry (left). Using activity map (right) the thickness of the grain cloud is attenuated in the regions with high disparity variance; hence, the small depth details stay visible.

Compositing In order to add our grain layer to existing footage, they are both combined using addition in gamma-corrected space as the grain application operator. This guarantees that the grain is approximately equally visible everywhere in the picture.

5. Results

We applied our method to two rendered sequences – BIRD and SINTEL (Fig. 1, third row, and Fig. 6, left side) and one video sequence – BALLET (Fig. 6, right side). Each sequence simulates different type of grain: large and clearly visible grain of an old, black-and-white film (SINTEL), less pronounced grain of a more recent film (BIRD), and fine grain of a modern (yet grainy) film (BALLET). In the SINTEL and BIRD sequences, we used freely available scans of 35mm film (<http://7dblue.wordpress.com/tools-downloads/>). To mimic different sizes of film stock, two differently sized crops of the grain images were used. For the BALLET sequence, the grain was generated in Adobe After Effects CS4. We believe that with this variability, we exhausted the range of useful sizes of grain. For example, it is unlikely to find bigger grain in films or games than the one used in the SINTEL sequence. On the other hand, the grain in BALLET sequence is barely visible. Additionally, we also generated an image of a face, where the depth map was artificially compressed to mimic a typical artifact of 2D-to-3D compression, and compared on-surface grain to ours (Fig. 5).

We used $n = 5$ layers and we set the volume parameters to $\alpha_{\min} = 5.3$ px (ca. 8 arc min) and $\alpha_{\max} = 9.6$ px (ca.

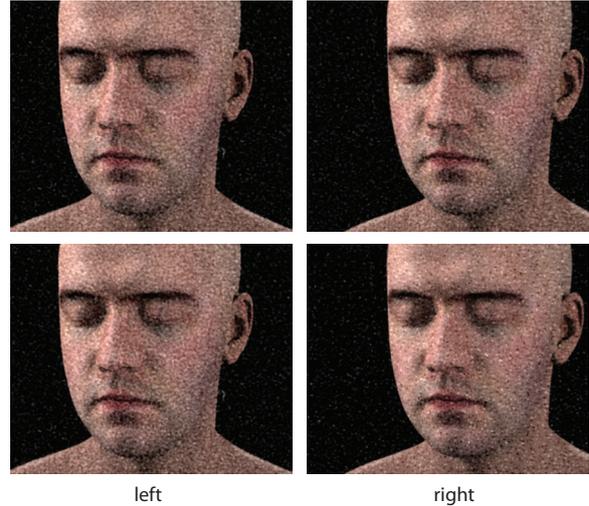


Figure 5: When the depth buffer is not detailed enough, on-surface grain contributes to undesired cardboarding effect (top). Our floating grain solves this problem (bottom). See study in Sec. 8 for details. Note, that the grain in this figure was exaggerated for the purpose of illustration. Refer to the supplemental materials for real-sized grain.

Model copyright: Lee Perry-Smith (www.ir-ltd.net).

14.4 arc min). The α -parameters were derived in a perceptual study described in Sec. 6. The figures in the paper serve as an illustration only (in particular Fig. 5 features exaggerated grain). We refer user to the supplemental materials, where the resulting videos are provided. Please note, that it is very important to use a stereo system with minimal cross-talk levels, because the stereoscopic grain effect can be easily destroyed by ghosting. For similar reasons, videos should be watched at full resolution (no subsampling). Therefore, we discourage use of anaglyph glasses or systems that reduce resolution, e. g., row-interleaved displays, and recommend shutter glasses or dual-projector systems.

6. Parameters Estimation

Our method for stereoscopic grain has two free parameters α_{\min} and α_{\max} , which are responsible for controlling thickness of the grain volume. Although both of them could be set by a skillful artist, in this section, we present a procedure that was used to obtain good values that can be used independently of the content.

Subjects Thirteen subjects (7 F, 6 M) took part in the experiment. All had basic background in computer graphics or computer vision, however, they were naïve with respect to the goal of the study, and their knowledge in stereoscopic 3D graphics was limited. They had normal or corrected-to-normal vision, and were screened for stereo-blindness.

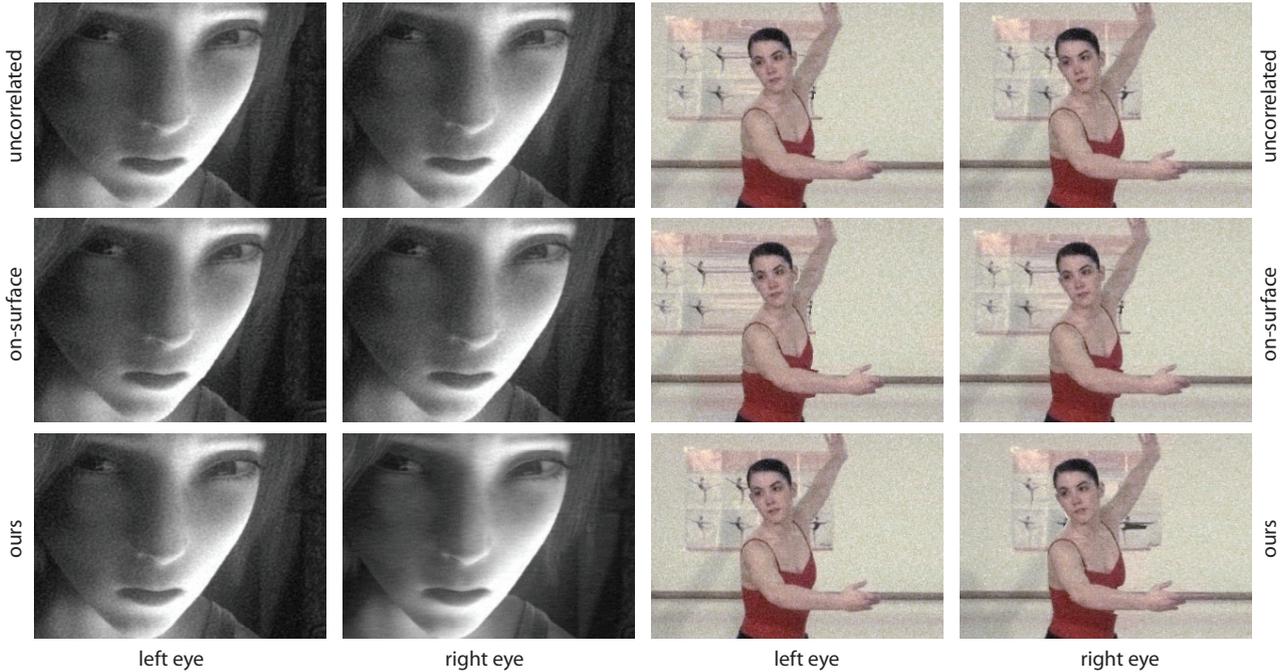


Figure 6: Results of our algorithm: sequence SINTEL (left side) and BALLETT (right side). We compare uncorrelated grain (first row) with on-surface grain (second row), and our floating grain (third row). The images are supposed to be viewed using uncrossed (parallel) free fusion, and are provided as an illustration only. Please refer to the supplemental materials for the full video sequences.

Sintel scene copyright: Blender Foundation (www.sintel.org), Ballet sequence copyright: Microsoft Research.

Equipment We used an Asus VG278HE 27-inch display (1920×1080 pixels), along with NVIDIA 3D Vision 2 active shutter glasses. The screen was observed from a distance of 50 cm. Measurements were performed in controlled, office-lighting conditions. The stimuli were displayed on a neutral grey background.

Task Because the participants were not familiar with different solutions for stereoscopic grain, the first part of the experiment was a training part. The subjects were shown the BIRD sequence, and they could switch between different kinds of grain (i. e., on-surface, uncorrelated, and our volumetric grain). They were also free to manipulate thickness of the volumetric grain using a slider and pause the sequence. Pausing was allowed only in the training session, in other experiments this option was disabled. Afterwards, they were asked to adjust the thickness of the volumetric grain so that the volume appearance is clear. In order to check whether they can distinguish among different kinds of grain after this short introductory session, they were shown the three different methods in random order (volumetric grain with their own settings), and were asked to assign them to their names. Ten participants did not have problems with identifying the methods, and they took part in the main experiment.

In order to estimate the two parameters (α_{min} and α_{max})

scene	avg. α_{min}	avg. α_{max}	on-surf.	ours
BIRD	5.2 ± 0.8 px	8.0 ± 1.4 px	7/10	8/10
SINTEL	4.4 ± 0.4 px	9.0 ± 1.9 px	9/10	9/10
BALLETT	6.3 ± 1.1 px	12.1 ± 2.2 px	5/10	7/10
AVG	5.3 px	9.6 px	21/30	24/30

Table 1: The results of the parameter-estimation study and the preference study. The second and the third columns show average values of α -parameters by scene. The indicated errors are standard errors of the mean. The last two columns show how many times the given method was preferred over uncorrelated grain. Both results are significant, with p -values in one-sided sign test 0.02 and 0.0007, respectively. The result for ours vs. on-surface (16/30, not shown) is not significant.

we designed a two-step process. To estimate α_{min} , the participants were asked to adjust the thickness of our grain in the sequences BIRD, SINTEL, and BALLETT (presented in random order), so that it had a just noticeable volume. At this point the attenuation map was disabled. Next, the map was enabled and the participants could adjust α_{max} to their liking. Table 1 (second and third column) presents the total and by-scene averages of the two parameters.

7. Preference study

In order to evaluate our technique, we conducted a preference study the day after the parameter estimation study (Sec. 6), in which the same 10 subjects participated. The apparatus and viewing conditions were the same as in the parameter estimation experiment.

Stimuli The sequences BIRD, SINTEL, and BALLET were used as the stimuli. Each sequence was processed using the three grain application methods, i. e., uncorrelated, on-surface, and ours. The grand average values of α_{\min} and α_{\max} obtained in the parameter estimation study were used for our method.

Task In a single trial, the subject was presented one of the sequences, and could freely switch between three versions (labeled A, B, and C) corresponding to different grain application methods. The subject was asked by the experimenter to indicate the version he/she preferred the most, and confirm the choice by pressing the Enter key. Then, the indicated version was removed, and the same question was repeated for the remaining two versions. Order of sequences, and order of methods for each sequence was randomized. The results of this study are presented in Table. 1 (fourth and fifth column).

8. Shape naturalness

In the third study we analyzed the influence of our technique on shape perception, and its ability to mask artifacts of the depth map. The subjects and viewing condition were the same as in the two other studies.

Stimulus In this experiment we used the FACE sequence in two versions: with on-surface grain and our grain. The depth buffer in this sequence had been remapped to enforce insufficient variation in depth, that often arises in the process of 2D-to-3D conversion.

Task This experiment consisted of a single trial. In it, the subject was shown the two versions of the sequence side-by-side (labeled A and B) in a randomized order. Next, the subject was asked by the experimenter to indicate in which version the face appeared more natural in terms of the 3D shape. Eight out of ten subjects found the face in the sequence processed using our method, as having more natural shape. The result is significant with $p < 0.055$ in the one-sided sign test.

9. Additional Results

An interesting case are “surfaces” with ill-defined depth, such as sky, participating media, or out-of-focus areas. To compare the performance of the two depth-dependent methods, i. e., on-surface and our grain, we generated additional four sequences: SQUIRREL, where the sky constitutes a large portion of the image, CANDLE, containing a semi-transparent

smoke volume, and STONES and FLOOR with depth-of-field effects. In the case of the sky we assumed an arbitrary constant depth at some distance behind the character. To determine the depth for the smoke, we used a stereo correspondence algorithm [HRB*13]. In the out-of-focus areas we used the depth of the corresponding non-blurred sequence. Using the *Match Grain* effect in Adobe After Effects CS6 (at the default settings and 16 px sample size) we closely matched selected frames from the feature films *Saving Private Ryan*, *300*, and *Planet Terror*. The results are presented in Fig. 7. See the supplemental material for the reference frames from the films and the resulting full-resolution animations.

10. Discussion

On-surface grain adds to luminance patterns on the objects, thus influencing their depth perception [DRE*12]. In several cases this may be undesired: First, infinite planes (e. g., sky or very distant backgrounds) and areas of undefined depth (e. g., out-of-focus backgrounds, smoke) look unnatural when their originally fuzzy depth becomes strictly defined (see Fig. 7). Second, when there is not enough depth variation, some objects may seem too flat (see Fig. 5). Last, when the depth buffer is not perfect, the errors are more evident (see Fig. 8). Our solution avoids all these problems: the sky and out-of-focus areas seem to have volume, and unnatural flatness and errors are masked.

We hypothesize that our approach might actually cause some detail hallucination. Stereoscopic grain introduces additional disparity signal to the original scene disparity map, which stimulates disparity-selective neurons that otherwise might not be activated. Additionally, there are a number of effects related to layered RDSs, including attraction between near layers and depth repulsion for layers more distant than 4–6 arc min [SCS91].

The number of layers that we used in the experiments ensured that the grain was perceived as a volume rather than separate layers. We did not explore further the influence of the quantity and relative placement of the layers, since the visual system seems to be insensitive to the specifics of the dot distribution within a volume (Sec. 3). For the same reasons, random distribution of dots within the given volume would yield visually equivalent results. Complete randomization of the grain placement would produce excessive pixel disparities, and would consequently break the binocular fusion.

Since the image-space size of the grain is not modulated, the apparent size of the particles may depend on the distance to the observer, with the more distant particles appearing larger. However, this was a deliberate design choice, as we wanted to modify only one view of the stereoscopic pair in order to maintain backward-compatibility (the other view is identical to the 2D version).

The approach we took in the parameter estimation study, with the estimation preceded by a training part, might have



Figure 8: A fragment of the *BALLET* sequence. With on-surface grain the artifacts of the 2D-to-3D conversion are emphasized, whereas with our grain they are masked.

Ballet sequence copyright: Microsoft Research.

biased the results in favor of our method, because otherwise some subjects would have not noticed the differences between the strategies. However, we feel that it was a justified choice, because at least basic knowledge in stereoscopic 3D and film production, as well as attention to detail is required to appreciate this subtle effect.

The results of the study in Sec. 6 showed that 77% of subjects were able to discern different grain placement methods. The study in Sec. 7 showed that both on-surface and our method are preferred over the uncorrelated grain (p-values in one-sided sign test 0.02 and 0.0007, respectively). Although the difference we found between the on-surface and our grain was not statistically significant, being on par with the industry standard can by no means be considered a failure, because in the end it is a matter of taste which tool to use, and the decision should be left to the artist. Additionally, the study in Sec. 8 demonstrated that there *are* cases, when using our method instead of the on-surface is beneficial: when there are artifacts in the depth map, they are less obvious when our method is used. The main goal of the industry has always been increasing the picture quality. However, despite the technological advances in film-making, grain can be clearly seen even in very recent mainstream stereoscopic 3D films (e. g.,

Transformers: Age of Extinction). Furthermore, intentional lowering of the quality is a very common technique among designers (e. g., “grunge” typefaces) and artists (e. g., “low bit” aesthetic in music). It is unclear if the industry will eventually enforce completely grain-free S3D production in the future, however, we predict that film grain will continue to appear in films at least as a means of stylization (e. g., *Hugo*).

11. Future work

In this paper we considered only film grain, however, our approach could be extended to handle other forms of visual noise. We propose to apply similar methods to 3D images and videos where JPEG/MPEG compression artifacts are clearly visible. As previously, the basic idea is that the medium should be separated from the scene, therefore we do not want the artifacts to be visible on the surface of the objects. Independent processing of the left and right channel accomplishes this goal, but only partially. If the level of compression is considerable, the user may find it hard to fuse the stereo image pair. Moreover, intra-channel alignment of the macro-block grid can appear as the shower-door effect (see Fig. 9). Analogously, the JPEG/MPEG artifacts should

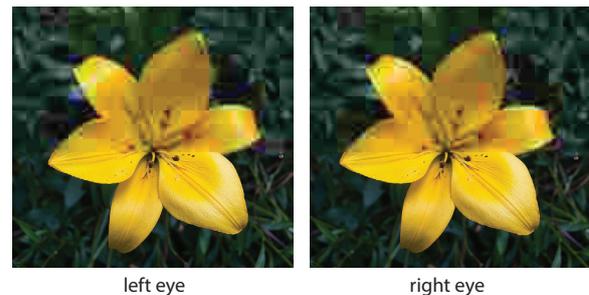


Figure 9: Upper part of the stereoscopic picture has been downsampled and compressed using low-quality settings of the JPEG format. Because the left and right channels have been encoded independently, it is very problematic to fuse them.

Photograph copyright: JJ Harrison. CC-BY-SA 3.0.

be placed somewhere between the objects and the spectator. This way the scene will look “submerged” in the medium rather than simply “cut out” of it.

One can question the need of such techniques: since the network bandwidths and the processing power of the computers are constantly increasing, such artifacts are becoming less of a problem for desktop users. However, 3D-capable hand-held devices are gaining in popularity, and we foresee, that compression artifacts, at least for some time, can still be an issue in this segment of less powerful devices.

Additionally, compression artifacts can be introduced on purpose, as means of stylization or artistic expression. Unlikely as it sounds, the aesthetic value of JPEG artifacts has

been already acknowledged [RS09]. A related idea of intentionally introducing coding errors (or “glitches”) to data stream is a well-established practice in visual arts [Men11]. To our knowledge, the possibilities of applying it in the context of stereoscopic 3D graphics have not been explored so far.

References

- [AT88] AKERSTROM R. A., TODD J. T.: The perception of stereoscopic transparency. *Attention, Perception, and Psychophysics* 44, 5 (1988), 421–432. 3
- [BL79] BEUCHER S., LANTUEJOUL C.: Use of watersheds in contour detection. In *Int. Workshop on Image Processing, Real-time Edge and Motion Detection* (1979). 3
- [CD01] CUMMING B., DEANGELIS G.: The physiology of stereopsis. *Annual Review of Neuroscience* 24 (2001), 203–238. 3
- [DF03] DALY S., FENG X.: Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Color Imaging VIII: Processing, Hardcopy, and Applications* (2003), vol. 5008 of *SPIE*, pp. 455–466. 2
- [DRE*11] DIDYK P., RITSCHER T., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P.: A perceptual model for disparity. *ACM Trans. Graph.* 30 (2011), 96:1–10. 4
- [DRE*12] DIDYK P., RITSCHER T., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P., MATUSIK W.: A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph.* 31, 6 (2012), 184:1–10. 7
- [DSCW06] DE STEFANO A., COLLIS B., WHITE P.: Synthesising and reducing film grain. *Journal of Visual Communication and Image Representation* 17, 1 (2006), 163–182. 2
- [FJ05] FAIRCHILD M. D., JOHNSON G. M.: On the salience of novel stimuli: Adaptation and image noise. In *IS&T/SID 13th Color Imaging Conference* (2005), pp. 333–338. 2
- [Gia] GIANT BOMB: Film grain. www.giantbomb.com/film-grain/92-487/. Accessed 13.06.2014. 1
- [GM97] GEIGEL J., MUSGRAVE F. K.: A model for simulating the photographic development process on digital images. In *Proc. ACM SIGGRAPH* (1997), pp. 135–142. 2
- [Gom00] GOMBRICH E. H.: *Art and illusion*. Princeton University Press, 2000. 2
- [GOW12] GOUTCHER R., O’KANE L., WILCOX L. M.: Representation of stereoscopic volumes. *Journal of Vision* 12, 9 (2012), 221. 3
- [HR02] HOWARD I. P., ROGERS B. J.: *Seeing in Depth*, vol. 2: Depth Perception. I. Porteous, Toronto, 2002. 4
- [HRB*13] HOSNI A., RHEMANN C., BLEYER M., ROTHER C., GELAUTZ M.: Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 35, 2 (2013), 504–511. 7
- [JF00] JOHNSON G. M., FAIRCHILD M. D.: Sharpness rules. In *IS&T/SID 8th Color Imaging Conference* (2000), pp. 24–30. 2
- [Jul64] JULESZ B.: Binocular depth perception without familiarity cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science* 145, 3630 (1964), 356–362. 3
- [KAK09] KURIHARA T., AOKI N., KOBAYASHI H.: Analysis of sharpness increase by image noise. In *Human Vision and electronic Imaging XIV* (2009), vol. 7240 of *SPIE*, pp. 724014:1–9. 2
- [KLL13] KIM Y., LEE Y., KANG H., LEE S.: Stereoscopic 3D line drawing. *ACM Trans. Graph.* 32, 4 (2013), 57:1–13. 3
- [KMAK08] KURIHARA T., MANABE Y., AOKI N., KOBAYASHI H.: Digital image improvement by adding noise: An example by a professional photographer. In *Image Quality and System Performance V* (2008), vol. 6808 of *SPIE*, pp. 1–10. 1
- [LKL13] LEE Y., KIM Y., KANG H., LEE S.: Binocular depth perception of stereoscopic 3D line drawings. In *Proc. ACM Symposium on Applied Perception* (2013), pp. 31–34. 3
- [LL96] LANKHEET M. J., LENNIE P.: Spatio-temporal requirements for binocular correlation in stereopsis. *Vision Research* 36, 4 (1996), 527–538. 1, 3
- [LLC*10] LAGAE A., LEFEBVRE S., COOK R., DE ROSE T., DRETTAKIS G., EBERT D. S., LEWIS J. P., PERLIN K., ZWICKER M.: State of the art in procedural noise functions. In *EG 2010 - State of the Art Reports* (2010). 2
- [Men09] MENDIBURU B.: *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, 2009. 2
- [Men11] MENKMAN R.: *The Glitch Moment(um)*. Institute of Network Cultures, Amsterdam, 2011. 9
- [NAK12] NORTHAM L., ASENTE P., KAPLAN C. S.: Consistent stylization and painterly rendering of stereoscopic 3D images. In *Proc. NPAR* (2012), pp. 47–56. 3
- [Rid11] RIDANOVIC I.: Interview with stereographer Daniele Siragusa. www.hdhead.com/?p=279, 2011. Accessed 13.06.2014. 2
- [RS09] RUFF T., SIMPSON B.: *JPEGS*. Aperture, 2009. 9
- [SCS91] STEVENSON S. B., CORMACK L. K., SCHOR C. M.: Depth attraction and repulsion in random dot stereograms. *Vision Research* 31, 5 (1991), 805–813. 7
- [Sey11a] SEYMOUR M.: Case study: How to make a Captain America wimp. *fxguide* (2011). 1
- [Sey11b] SEYMOUR M.: Hugo: A study of modern inventive visual effects. *fxguide* (2011). 1, 2
- [Sey12] SEYMOUR M.: Art of stereo conversion: 2D to 3D – 2012. *fxguide* (2012). 2
- [SG04] STAVRAKIS E., GELAUTZ M.: Image-based stereoscopic painterly rendering. In *Proc. EGSR* (2004), pp. 53–60. 3
- [SG05] STAVRAKIS E., GELAUTZ M.: Stereoscopic painting with varying levels of detail. In *Stereoscopic Displays and Virtual Reality Systems XII* (2005), vol. 5664 of *SPIE*, pp. 450–459. 3
- [SS07] STEPHENSON I., SAUNDERS A.: Simulating film grain using the noise-power spectrum. In *Theory and Practice of Computer Graphics* (2007), pp. 69–72. 2
- [TAW08] TSIRLIN I., ALLISON R. S., WILCOX L. M.: Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision* 8, 5 (2008). 3
- [TWA10] TSIRLIN I., WILCOX L. M., ALLISON R. S.: Perceptual artifacts in random-dot stereograms. *Perception* 39, 3 (2010), 349–355. 3
- [Ty175] TYLER C. W.: Spatial organization of binocular disparity sensitivity. *Vision Research* 15, 5 (1975), 583–590. 4
- [WG13] WINTER M., GANDOLPH D.: Film grain for stereoscopic or multi-view images, 2013. US Patent App. 13/762,479. 2
- [WPF10] WANG N., PALJIC A., FUCHS P.: A study of perception of volumetric rendering for immersive scientific visualization. In *20th International Conference on Artificial Reality and Teleexistence (ICAT’2010)* (2010), pp. 145–152. 2

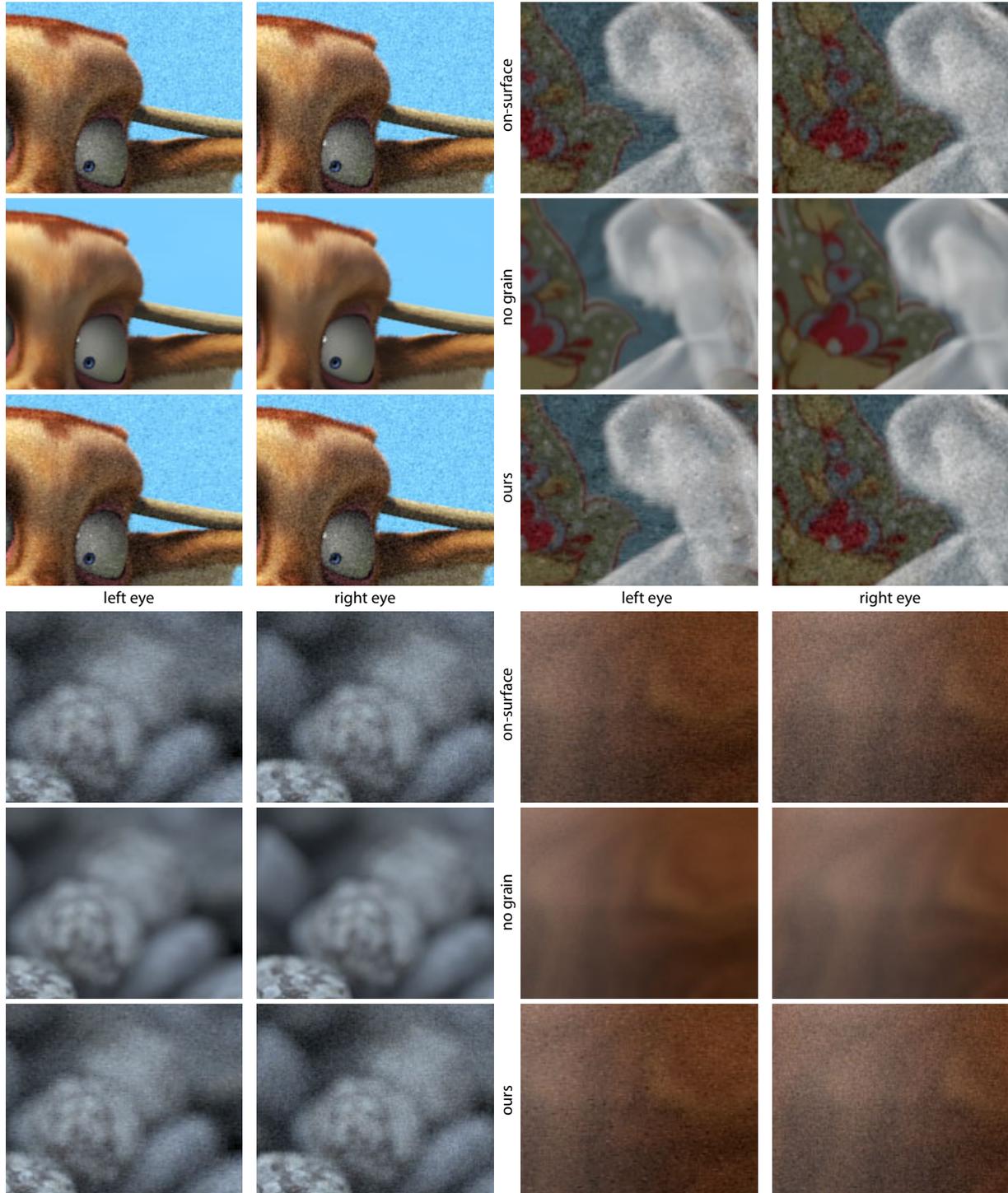


Figure 7: Sequences with ill-defined “surfaces”, such as sky (top-left, grain matching the film 300), smoke (top-right, matching Planet Terror), and out-of-focus areas (bottom, matching Saving Private Ryan). In these cases applying one-layer projected grain changes the fuzzy perception of the objects, which is maintained by our method. See the supplemental material for the full-resolution sequences.

Flying Squirrel scene copyright: Blender Foundation (www.bigbuckbunny.org), Rock Pack model copyright: StevenColemanDesigns

(www.blendswap.com).