

Quality Metric for Shadow Rendering

Krzysztof Kluczek

Gdańsk University of Technology
Department of Intelligent Interactive Systems,
Narutowicza 11/12, 80-233 Gdańsk
Email: krzysiek@devkk.net

Abstract—Shadow rendering is one of the most important aspects of rendering 3D environments, yet, the problem is far from trivial. A number of shadow rendering algorithms exist, with various degrees of rendering quality, fidelity and performance. Additionally, many of such algorithms offer high degrees of flexibility when it comes to fine tuning. This paper proposes a new method of measurement of quality of shadows produced by rendering algorithms, which method can be used for automation of algorithm choice and fine-tuning of such algorithms to specific data sets and use cases.

I. INTRODUCTION

ONE of the most important aspects of rendering virtual objects and environments is rendering of shadows [1]. While a great number of shadow rendering algorithms was developed, all of them require some degree of compromise between image quality, fidelity and rendering performance. With ray-tracing allowing very high shadow fidelity at the cost of rendering performance, and shadow map techniques offering [2] real-time performance even on low-end devices at the cost of image quality, the choice of shadow rendering algorithms is often one of key decision taken during development of rendering applications. Many of such algorithms offer further parametrization, which allows fine-tuning to given purpose. To simplify the task of such choices and enable some degree of automation in this matter an automated shadow rendering quality measurement algorithm is required. This paper presents a proposal of such an algorithm.

II. RELATED WORK

While a lot of research has already been done in the area of image quality assessment, it is surprising that we couldn't find any proposal of a complete metric fit to measure and compare quality of rendered shadows. The proposed NoRM No-Reference Image Quality Metric [3] was capable of detecting certain type of shadow rendering artifacts (shadow map aliasing), but because the method was heavily based on machine learning, it was impossible to be easily reproduced in our environment. Another study [4] evaluates existing full-reference image quality metrics in the context of detection and measurement of two certain types of shadow artifacts (acne and peter panning) showing that SSIM [5] and MSSIM [6] metrics outperform other evaluated methods. Still, in our research of existing works we were unable to find a quality metric that would focus on all aspects of evaluating quality of shadow rendering algorithms.

III. PROPOSED MODEL

The proposed quality metric is capable of quality assessment of shadow rendered for a single light source (for any type of such light source) as seen from a single view point. The model is based on weighted average of several submetrics, each focusing on a key aspect of rendered shadow quality:

- shadow fidelity
- aesthetics
- detail
- rendering performance

The weighted average of the above aspects allows tuning the metric itself to given purpose and expectations. For example, while rendering performance will be much more important in the real-time applications than absolute shadow fidelity, the exact opposite will be true for offline rendering in cinematography.

Because the proposed metric was developed mostly for use in comparing shadow rendering algorithms, the metric computation itself does not need to be real-time. Therefore, we allowed ourselves to use full-reference, white-box methods for simplicity. The reference images were generated using ray-tracing algorithm with accurate soft-shadow simulation and high sample count (1024 samples per pixel). White-box approach allows easy separation of the shadow information from the rest of the scene by use of shadow masks. The shadow mask is an additional output from a renderer, which represents only the geometry term (the visibility factor) for a given light. The example of shadow mask image generated for our test scene can be seen on Figure 1.

IV. SHADOW FIDELITY MEASUREMENT

The purpose of the shadow fidelity metric component is measurement of how closely the shadow under evaluation matches the expected physical shadow. This component includes assessment of correct shadow placement and topology, as well as correct representation of fully shadowed and fully unshadowed area, but does not include noise errors, which will be included in the shadow aesthetics component of the metric. Such separation was implemented in order to enable fine-tuning the final metric by specifying the weights for each submetric separately to adjust the importance factors to specific needs.

The shadow fidelity submetric is, just like the main metric, a weighted average of several components:

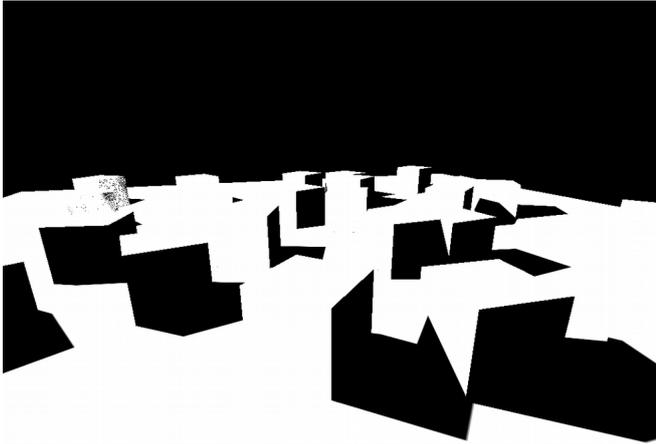


Fig 1. An example of shadow mask image for our test scene

- 10% shadow contour match
- 50% shadow contour match
- 90% shadow contour match
- dark area coverage
- lit area coverage

The first three components assure that the soft shadow boundary in test image is located where it ought to be, while the last two components assure that there are no extra features present in regions, which are empty in the reference image.

A. 10%, 50% and 90% Shadow Contour Matching

The shadow contour line is defined as isoline, for which the shadow factor (the geometry factor from the lighting equation) is equal to specified percentage value. Three contour lines are used in our method: the 10% line, 50% line and 90% line. These factors were chosen experimentally to provide 10% margin from fully lit and fully shadowed areas. Such margin is needed, as the submetric should minimize impact of noise errors, which are to be detected with different submetric. Additional line, the 50% contour line, was introduced to test the fidelity of shadow cross-section – the distribution of shadow mask values across the penumbra region.

The method for extracting shadow contour lines is identical for all contour line values. First, the shadow mask image is blurred to reduce impact of noise on the shadow line. In our tests we used 11x11 blur kernel for images with 1024x768 pixel resolution. Then, the isoline is extracted by comparing each pixel of the blurred mask with its right, lower and lower-right neighbors. The pixel is marked if and only if a region consisting of the pixel and the neighbors contains shadow mask value transition across the contour value. As a corner case, if the region has all pixel values equal to the contour line threshold, the pixel is still marked in order to guarantee continuity of the contour line. The above operation (blurring and contour line extraction) is carried on both the test and the reference images. In the final step, for each contour pixel in reference image a contour

pixel in test image is found, and quadratic mean of distances to such pixels is computed. The final score of the given contour line match submetric is computed as:

$$CM = 1 - \frac{D_{RMS}}{D_{MAX}} \quad (1)$$

where:

CM – is the contour line submetric score,

D_{RMS} – is the quadratic mean of contour distances,

D_{MAX} – is the worst-case maximum contour distance

Ideally, the mean would equal zero, which means perfect match, yielding a perfect score of 1. A worst-case scenario, yielding a score of 0, is assumed to be a scenario in which the quadratic mean of contour distance is equal to D_{MAX} – an arbitrarily chosen maximum contour distance. In our implementation, that distance was set as a 20% of image size (a geometric mean of image width and height in pixels). The choice of that parameter can be arbitrary, because the metric is intended to be a comparative one, and any fine-tuning is meant to be done by choosing the weights of the submetrics.

B. Dark and Light Area Coverage

The dark and light area coverage submetrics are computed in a manner similar to each other. First, the shadow mask image is blurred to reduce the impact of noise and shadow acne on the submetric score. In our implementation, we use the same blurred image as in shadow contour matching submetrics. Then, depending on whether we compute the dark or the light area coverage submetric, we mark all pixels which contain blurred mask values below 10%, or above 90%, for dark and light areas correspondingly. This process is carried on both the test image as well as on reference image. In the final step, both images are compared. Only the region of pixels marked on the reference image is considered. Within that region all unmarked pixels on test image are counted, giving final score of the submetric as:

$$AM = 1 - \frac{N_{unmarked}}{N_{region}} \quad (2)$$

where:

AM – is the area coverage submetric score,

$N_{unmarked}$ – is the count of unmarked pixels in the test image (within considered region),

N_{region} – is the count of all pixels of the considered region

In the event of a perfect match, all pixels marked on the reference image will also be marked on the test image, resulting in ideal score of 1. In worst case, none of the considered pixels will be marked on the test image, giving the score of 0.

V. SHADOW AESTHETICS MEASUREMENT

With the aesthetic submetric score we aim to detect two types of shadow rendering artifacts: the shadow acne and shadow noise in the shadowed region. While the aesthetic quality of shadow can be hard to measure, we decided to focus on shadow noise in this metric, as image noise is one of the most frequent and typical image rendering errors. The shadow aesthetics submetric is, again just like the main met-

ric, a weighted average of several (in this case: two) components:

- shadow ACNE value
- shadow noise value

With both components scaled to yield a value of 1 in case of perfect shadow image and a worst-case value of 0.

A. Shadow Acne Detection

The shadow acne rendering artifacts are typical to shadow map rendering algorithms and are characterized by appearance of high frequency noise within the fully lit regions of the shadow mask. Therefore, to detect such errors we need to measure the noise level within the lit portion of the shadow mask. To compute that level, we first process the shadow mask image using the Roberts cross operator to isolate high-frequency image noise [7]. Then we compute the quadratic mean of computed noise values within the lit region of the shadow mask. The lit region is considered to be the light region computed during measurement of the light area coverage in one of the earlier submetrics. Both the test shadow mask and the reference shadow mask images are processed in the above manner and the final acne detection score is computed as:

$$AV = \min\left(\frac{1 - T_{RMS}}{1 - R_{RMS}}, 1\right) \quad (3)$$

where:

AV – is the score of acne detection,

T_{RMS} – is the quadratic mean of noise detected within the lit region on the test image,

R_{RMS} – is the quadratic mean of noise detected within the lit region on the reference image

The above equation normalizes the acne score of the test image to the acne score of the reference image because some features of the reference image (such as penumbra edges) can be falsely detected as acne, therefore making ideal score of 1 impossible to reach. The final value is limited to 1 to prevent the score from going above that value if test image lacks features, present on the reference image, that would otherwise be falsely detected as acne.

B. Shadow Noise Detection

The shadow noise is detected in a manner identical to the acne detection method outlined above, with the exception that now only the dark region is considered. The dark region is the region detected as a dark portion of the shadow mask during dark region coverage measurement. The rest of the noise detection algorithm remains the same as in acne detection algorithm.

VI. SHADOW DETAIL MEASUREMENT

The insufficient shadow detail comes in most cases from the low shadow map resolution, as compared to its projection on the screen. To measure the quality of shadow detail as a shadow resolution compared to image resolution, we have to utilize white-box testing method and produce yet another image output from our shadow rendering algorithms: the shadow map texels projection on the scene. Such texel

projection image is prepared first by assigning each shadow map texel an unique color index, then projecting such shadow map directly on the scene. To limit further computations to region of interest where shadow map detail can possibly affect rendering, in such projection we ignore any geometry that is not facing the light for which shadow quality is measured, as well as any regions that don't contain any geometry (e.g. sky). The example of texel projection image generated for low resolution shadow map (for the purpose of readability) can be seen on Figure 2.

To compute the shadow detail submetric score the following method is used. For each pixel of interest within the shadow map texel projection image we count the number of pixels, that share the same texel index. Then this count is averaged using quadratic mean. Then the final score of shadow detail submetric is computed as:

$$SD = \frac{1}{S_{ratio}} \quad (4)$$

where:

SD – is shadow detail submetric score,

S_{ratio} – is share ratio computed using the quadratic mean

Because we consider only pixels that have any shadow texels projected on them, the share ratio can never drop below 1. Therefore, the perfect score, when each pixel has its own shadow map texel assigned, is 1. Note that the shadow map resolution exceeding the pixel resolution across any

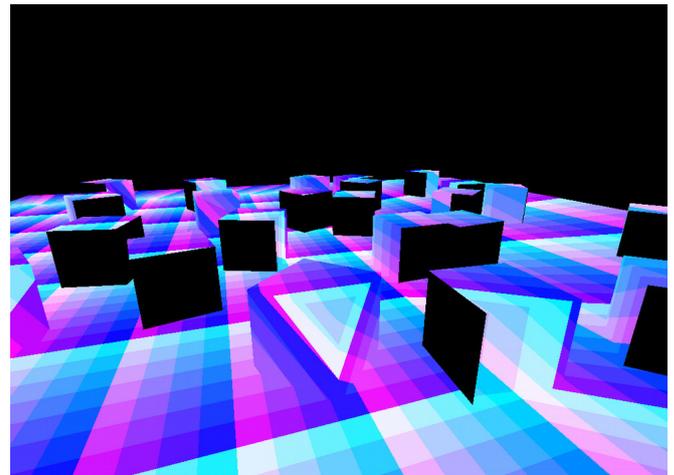


Fig 2. An example of a shadow texel projection image

portion of the scene does not further increase the score.

In the event the given shadow rendering technique does not rely on intermediate projected texture maps and has infinite detail resolution (like raytracing or shadow volume techniques), the shadow detail metric is skipped entirely assuming perfect score of 1.

VII. RENDERING PERFORMANCE SCORE

The rendering performance submetric score is computed basing on shadow rendering time and is computed as:

$$RP = \frac{1}{1 + \frac{T_{\text{shadowed}} - T_{\text{unshadowed}}}{T_{\text{ref}}}} \quad (5)$$

where:

RP – is the rendering performance score,

T_{shadowed} – is the rendering time of the scene with the shadow rendering algorithm enabled,

$T_{\text{unshadowed}}$ – is the rendering time of the scene with the shadow rendering algorithm disabled,

T_{ref} – is the reference time value, scaling the score distribution

We suggest using value of T_{ref} equal to the inverse of the minimum final target frame rate (maximum time of rendering allowed for a complete single frame). With rendering performance score specified this way, the score assumes value of 1 in ideal case of shadow rendering not taking any time. When shadow rendering time rises to maximum time allowed for rendering the whole frame (making the shadow algorithm impossible to use), the performance score drops to value of 0.5. Further increase in shadow rendering time causes the score to drop further towards 0 as the rendering time rises towards infinity.

During our tests we did not use performance measurement submetric, as our implementations of shadowing algorithms were not fully optimized. Therefore, any performance comparisons would not produce any meaningful results.

VIII. EXAMPLE RESULTS

To test our quality metric on real data sets we have implemented four different shadowing techniques, in addition to raytracing, which was used as reference. All of the implemented test techniques were based on the shadow mapping technique for point light shadows. The most widely used for such cases Cube Shadow Map technique (CSM) was implemented, as well as slightly modified Tetrahedral Shadow Map technique introduced by [8]. Additionally, both Cube Shadow Map and Tetrahedral Shadow Map techniques were implemented with Variance Shadow Map variants (CVSM and TVSM, respectively). The four techniques were tested with various shadow map texture resolutions using a simple synthetic test scene with a single point light. The test cases are named after the technique being tested and shadow map resolution being used (e.g. CSM 1024). For cube shadow maps, the resolution is specified as edge length of the shadow map cube, in texels. For tetrahedral shadow maps the resolution is specified as edge length of a cube map, which would have a number of map texels closest to the number of texels in the tetrahedral map, thus making the resolution specification comparable to the cube shadow maps.

The reference shadow mask image for our data set can be seen on Figure 3. The shadow mask images used for each test variant are presented on Figure 4. The quality measure-

ment results are presented in Table I, separately for each submetric component. As expected, the lower resolution shadow map variants achieve lower scores, with score changes most noticeable for really low resolution shadow maps. The tetrahedral shadow mapping techniques score on average worse than cube shadow map techniques, as the map shape distortions are more prominent in the tetrahedral maps, resulting in uneven distribution of shadow map texels over the lit scene. The Variance Shadow Mapping variants perform comparably to the non-VSM variants, with noticeable improvement within the lit region at the cost of degradation of shadow region coverage, resulting from softening the shadow edge across the region transitions. It can be seen, that this also slightly influences acne and noise measurements, as this influence was limited, but not completely removed from the metric.

IX. CONCLUSION

In this article we presented a quality metric proposal for quality measurements of shadow rendering algorithms. We presented several components of such metric, which were designed to measure quality of various characteristics of the shadow mask images. While there is still room for improvements when it comes to the methods and algorithms used, the quality measurement metric presented can already be useful in comparison of shadow rendering algorithms, detecting common rendering artifacts and yielding stable and comparable results.

REFERENCES

- [1] P. Boulenguez, B. Airieau, M.-C. Larabi, D. Meneveau, "Towards a perceptual quality metric for computer-generated images," in *Proc. SPIE 8293, Image Quality and System Performance IX*, Burlingame, CA 2012, doi:10.1117/12.908067.
- [2] L. Williams, "Casting curved shadows on curved surfaces", *Proc. of the 5th annual conference on Computer graphics and interactive techniques - SIGGRAPH '78*, 1978.
- [3] R. Herzog, M. Čadik, T. O. Ayđin, K. I. Kim, K. Myszkowski, H. P. Seidel, H. (2012). NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis. *Computer Graphics Forum*, 31(2pt3), pp. 545-554.
- [4] R. Piórkowski, "Automatic Detection of Shadow Acne and Peter Panning Artefacts in Computer Games." in *Central European Seminar on Computer Graphics for students 2015*, Smolenice, Slovakia, pp 117-123.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.
- [6] Z. Wang, E. P. Simoncelli, A. C. Bovik, "Multiscale structural similarity for image quality assessment," *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003, pp. 1398-1402 Vol.2. doi: 10.1109/ACSSC.2003.1292216
- [7] L. S. Davis, "A survey of edge detection techniques", *Computer Graphics and Image Processing*, vol 4, no. 3, pp 248-260, 1975
- [8] H.-C. Liao, "Shadow mapping for omni-directional light using tetrahedron mapping", in *GPU Pro: Advanced Rendering Techniques*, Boca Raton, CRC Press, 2010, pp. 455-475.

TABLE I.
QUALITY SUBMETRICS VALUES FOR DIFFERENT SHADOW RENDERING TECHNIQUES FOR THE TEST SCENE

Shadow rendering method	10% contour score	50% contour score	90% contour score	Dark area coverage	Lit area coverage	ACNE score	Shadow noise	Detail
Ray-tracing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CSM 128	0.909	0.909	0.907	0.920	0.764	0.995	0.994	0.000
CSM 256	0.935	0.931	0.925	0.952	0.776	0.996	0.992	0.002
CSM 512	0.945	0.940	0.937	0.973	0.803	0.995	0.991	0.006
CSM 1024	0.952	0.948	0.944	0.985	0.858	0.990	0.994	0.021
CVSM 128	0.848	0.890	0.901	0.851	0.790	0.986	0.992	0.000
CVSM 256	0.911	0.913	0.925	0.913	0.802	0.985	0.993	0.002
CVSM 512	0.942	0.943	0.940	0.953	0.830	0.988	0.994	0.006
CVSM 1024	0.948	0.948	0.945	0.975	0.865	0.994	0.995	0.021
TSM 128	0.871	0.866	0.864	0.953	0.596	0.989	0.987	0.000
TSM 256	0.882	0.879	0.862	0.969	0.630	0.991	0.988	0.000
TSM 512	0.925	0.908	0.894	0.975	0.730	0.995	0.986	0.001
TSM 1024	0.950	0.937	0.932	0.985	0.829	0.993	0.988	0.005
TVSM 128	0.858	0.894	0.901	0.842	0.713	0.985	0.993	0.000
TVSM 256	0.902	0.921	0.915	0.887	0.761	0.982	0.992	0.000
TVSM 512	0.926	0.936	0.935	0.934	0.837	0.987	0.993	0.001
TVSM 1024	0.949	0.948	0.945	0.965	0.862	0.991	0.994	0.005

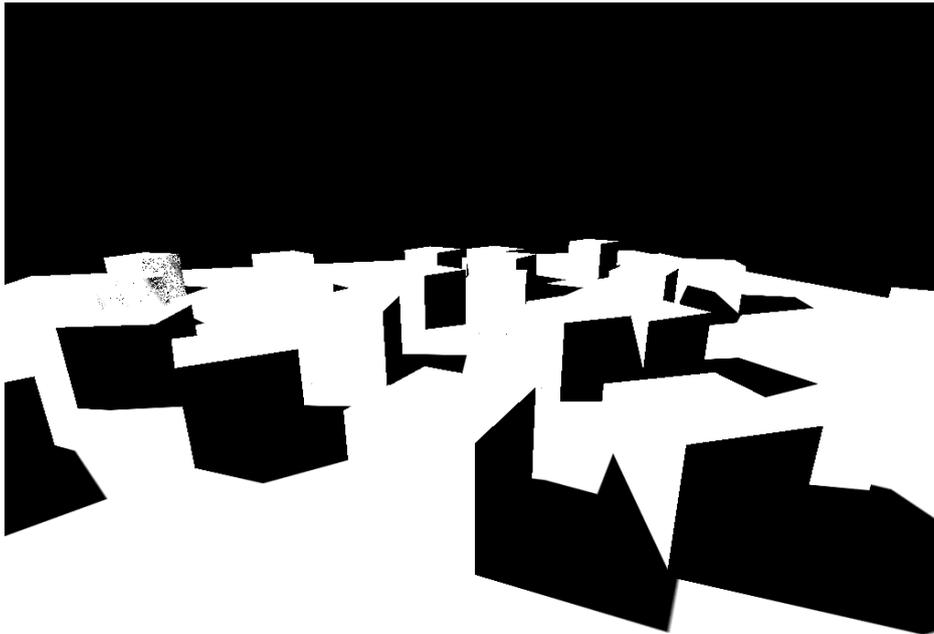


Fig 3. The reference shadow mask image generated using a commercial ray-tracer

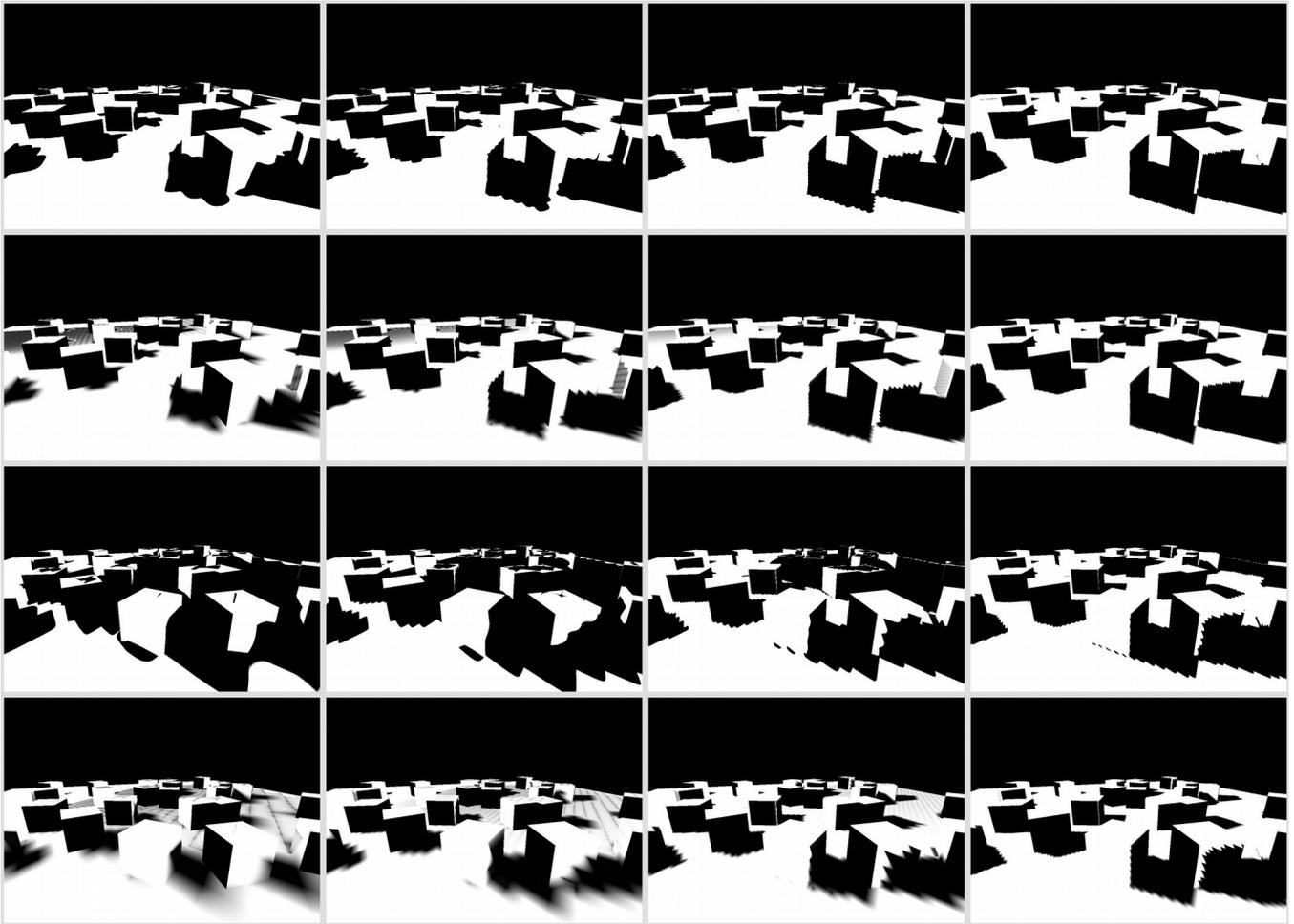


Fig 4. Shadow mask images used in our tests; techniques in rows: CSM, CVSM, TSM and TVSM; resolutions in columns: 128, 256, 512, 1024