Visible Difference Predictors: A Class of Perception-Based Metrics

Alexandre Chapiro, Param Hanji**, Maliha Ashraf**, Yuta Asano*, Rafal Mantiuk***

***Meta, USA**

****University of Cambridge, UK**

Abstract

The Visible Difference Predictors are a class of data driven, white box, efficiently implemented image or video difference metrics. They model important aspects of perception like spatial and temporal vision, foveation, and more, and are calibrated on datasets relevant for display and graphics applications. In this paper, we present a historic retrospective of VDPs, and a high-level technical overview and comparison to other metrics in the literature. Finally, we put forward a practical guide for selecting the appropriate metric for a given engineering problem and discuss how metrics can be effectively combined with subjective testing for high-confidence assessments.

Author Keywords

Image metrics; Display; Psychophysics; Perceptual Evaluation.

1. Introduction

Automatic evaluation of image and video quality is an important task in visual computing. It helps reduce the need for costly subjective evaluation and allows finer-grain searching of parameter spaces for studied techniques. This manuscript focuses on a series of metrics called *visual difference predictors* (VDPs), which are geared for use in display engineering.

Visual metrics: Designing display systems with good image quality is an important problem for scientists and engineers. Practical display systems are not perfect and contend with ever present visual artifacts (Fig. 1). It is often necessary to maximize the visual image quality of a given display while maintaining desirable tradeoffs with ancillary aspects of the system, such as weight, cost, or power consumption requirements. General-purpose visual difference metrics are often employed in this context. VDPs belong to the popular full-reference difference metrics class: these methods predict the perceived difference between two pieces of content, typically a pristine *reference* and a *test* which has been distorted by the artifact being studied.

Accurate perceptual predictions of this type are difficult due to the complexity of the human visual system, which is not easily modeled in its entirety. Most modern metrics ignore aspects of human vision that are difficult to model, such as color, foveation, or temporal vision, but which may be crucial for display engineering applications. Most popular metrics also do not model details of the display but choose to operate on standardized pixel values instead. This can be undesirable for specialized applications, as critical aspects of a display like resolution, size, brightness, color space, frame rate, and so on are left unaccounted. Consider the task of estimating distortions introduced by the optical stack in head mounted displays. These may include blur (Fig. 1, top-left), color fringes (top-middle), contrast loss (center-left), waveguide nonuniformity (WGNU, bottom-right), or the associated dynamic correction error (DCE, center-right). These errors are difficult to model in pixel space as they are dependent on the physical characteristics of light emitted by the display. The same is true for light source artifacts, like nonuniformity of individual elements in LED displays (LSNU, center-bottom).

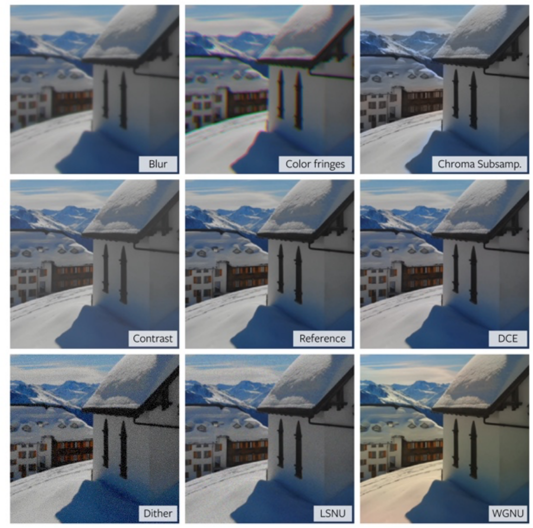

Figure 1. Images surrounding the reference show examples of common display artifacts, particularly those introduced by optical elements used in modern head-mounted displays.

2. Related Work – Popular Metrics

In this section, we will present an overview of VDP goals and present a qualitative comparison to a selection of metrics highly popular in display engineering. For a detailed technical discussion of modern perceptual difference metrics, including quantitative performance evaluations on several image quality datasets, please see our latest publications [1,2].

Visual Difference Predictors: VDP metrics are united by a standard approach. In the past, the philosophy of VDPs has been described by a ranked list of priorities:

1) Match data from human studies;
2) Good computational usability;
3) Plausible modeling of visual system mechanisms.

Previous VDPs [4,3] became popular for applications in industry and academia. This paper is focused on two recent successors [1,2] which target novel applications in foveation and color.

Popular metrics: While a great number of metrics is available in the literature, we will focus this discussion on a few methods that are especially popular for display applications. The PSNR formula [5] provides a way to scale mean-squared-error (MSE) in easier to interpret dB units. The structural similarity index measures (SSIM and MS-SSIM) [6] account for the spatial surrounding of a given pixel by analyzing its neighborhood. The CIE DE00 [7] formula is popular for color differences, developed for work on uniform patches. LPIPS [8] is a deep-learning algorithm, using network features to predict visible differences. Notably, while network-based approaches can be powerful, they struggle to extrapolate from training data, unlike VDPs' explicit modeling of vision, which allows for stable results even in cases not previously faced by the metric.

Table 1. This qualitative comparison of select popular metrics shows which aspects of vision and display are modeled

Metrics:	Approach	Display Geometry	Display Photometry	Spatial Features	Video	Color	Foveation
PSNR [5]	Signal quality	No	No	No	No	No	No
SSIM [6]	Correlation	No	No	Yes	No	No	No
CIE DE00 [7]	Equation	No	Yes	No	No	Yes	No
LPIPS [8]	Machine learning	No	No	Yes	No	Yes	No
FovVideoVDP [1]	VDP	Yes	Yes	Yes	Yes	No	Yes
ColorVideoVDP [2]	VDP	Yes	Yes	Yes	Yes	Yes	No

Table 1 above highlights features modeled by each metric:

- *Display geometry* describes the physical placement of the display's pixels in relation to the viewer and is not modeled by the non-VDP metrics.
- *Display photometry* describes modeling of the light being emitted by the display (as opposed to operating purely on pixel values), and except for CIE DE00, is rarely modeled by other metrics.
- *Spatial features* refer to whether a pixel's value is modeled independently or in context of its neighbors.
- *Video* metrics consider the temporal context of the pixel. None of the non-VDP metrics discussed here have this treatment, and most metrics in the literature include at most a window of 2 subsequent frames.
- *Color* refers to whether pixels are modeled in terms of chroma, or luma only.
- *Foveation* refers to modeling the loss of sensitivity for parts of the image outside the fovea. As human vision generally degrades in this case, this information may be important for methods like foveated rendering.

Because display artifacts often present mixes of several of these aspects of display and vision, it is important for a metric to accurately cover as many of these as possible.

3. VDP pipeline

In this work, we will focus on two modern VDPs, both of which are image and video difference metrics, and account for the viewing conditions, geometric, and photometric characteristics of the display.

Each metric opens the door to new, previously impossible, use cases. For both methods, extensive testing on novel and existing subjective image quality datasets also shows a significant gain in prediction performance compared to other methods for standard tasks.

FovVideoVDP [1] models the spatial, temporal, and peripheral aspects of perception. It is the first work that simultaneously treats these three central aspects of vision. It is especially useful for work on displays that cover a large field-of-view, such as Virtual and Augmented Reality displays, and associated methods, such as foveated rendering.

ColorVideoVDP [2] models spatial and temporal aspects of vision, for both luminance and color. It is built on novel psychophysical models of chromatic spatiotemporal contrast sensitivity and cross-channel contrast masking. ColorVideoVDP opens the doors to many novel applications which require the joint automated spatiotemporal assessment of luminance and color distortions, including video streaming, display specification and design, visual comparison of results, and perceptually guided quality optimization.

An outline of the workflow of VDP metrics is shown in Figure 2. For a detailed explanation of the techniques used in these metrics, we refer the reader to the paper manuscripts. The actively maintained implementation of each method is available on GitHub*. Comprehensive documentation and several practical examples are also provided.

Finally, for more context on the contrast sensitivity component, we point the readers to our recent work introducing a data-driven multidimensional contrast sensitivity function, stelaCSF [9].

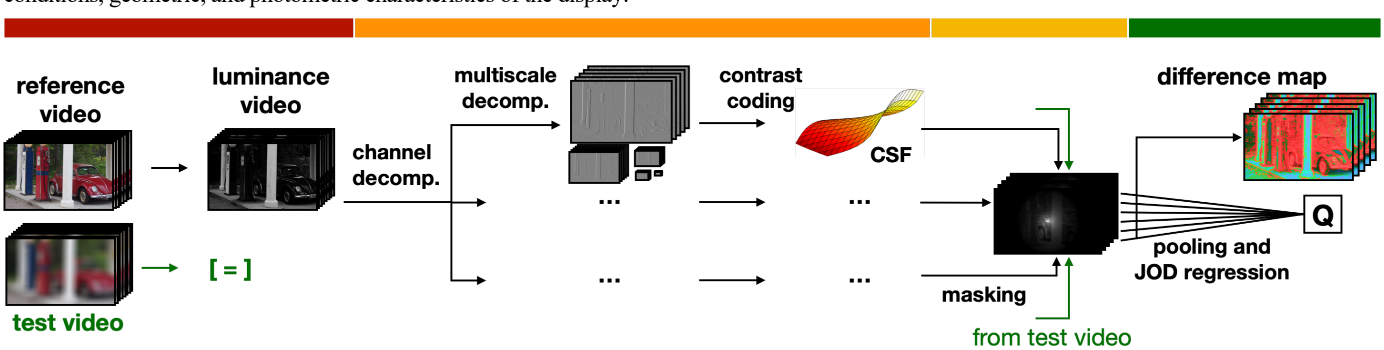

Figure 2. This figure shows an outline of the VDP pipeline. The red section represents video characterization, where the reference and test inputs are modeled in terms of display geometry and photometry. In orange, perceptual modeling is done using visual channels, including temporal and spatial multiscale decomposition (as well as additional color channels for ColorVideoVDP), and filtered through a CSF model. In yellow, the masking model is used to predict the visible difference between test and reference while considering the context of each pixel location. Finally, in green, resulting values are pooled and either rendered in a visual representation of the visible difference map or regressed to a single unified quality score.

* FovVideoVDP: https://github.com/gfxdisp/FovVideoVDP, and ColorVideoVDP: https://github.com/gfxdisp/ColorVideoVDP.

4. Metric Usage

The main advantage of employing metrics in an engineering workflow is their ease of use, which only requires simulating the conditions being examined. However, metrics are imperfect representations of the human visual system. In consequence, to obtain a high confidence assessment of display artifact visibility, it is often desirable to perform some subjective evaluation when physical prototyping becomes feasible. This evaluation can also be used to verify the results of the metric, e.g. by testing a subset of conditions simulated in an earlier design phase.

Most metrics in the literature output results in arbitrarily scaled units – for example, PSNR values typically span ranges of 30-50 dB. However, there is no clear pipeline to relate the metric output to subjective study results, or a straightforward way to select perceptually acceptable tradeoffs by setting an appropriate metric output threshold value. VDP metrics solve this issue by scaling the output on a perceptually meaningful absolute scale of Just-Objectionable-Differences (JODs, shown in Fig. 3). All VDP metrics' outputs are scaled in JODs using the same rule:

$$\boldsymbol{Q(R,T) = 10 - JOD(R,T)}$$

where **Q(R,T)** is the metric's output, and **JOD(R,T)** represents the expected JOD-scaled probability that the reference content is chosen over the test content in a 2-alternate-force-choice (2AFC) procedure. In other words, a metric score of 10 represents a case where the test image or video is 0 JODs away from the reference, i.e. it has quality perfectly identical to the reference and would be selected at random (50/50) with it in a subjective study. A VDP score of 9 represents 1 JOD, or a 75/25 split, etc.

For users interested in learning more about designing simple but effective pairwise comparison studies, the authors recommend the work of Perez-Ortiz et al. [10], as well as the use of the associated GitHub library, Pwcmp*. Note that results scaled in this way can be directly compared to VDP outputs in absolute terms for the purpose of validation.

Given that multiple VDP metrics exist, a natural question relates to which metric should be selected for a given task. Currently, we recommend using FovVideoVDP for tasks that require foveation, and ColorVideoVDP for all other tasks.

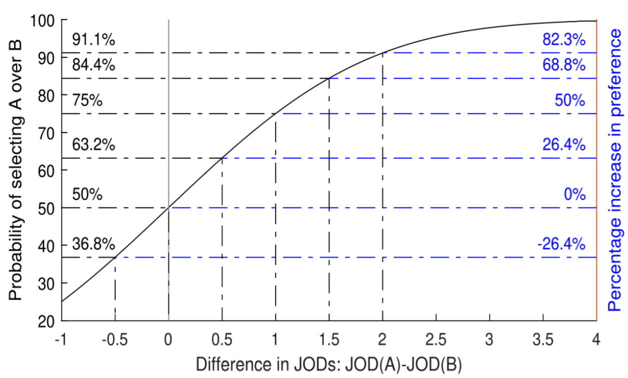

Figure 3. This figure depicts the JOD scale, relating the probability of the test image being preferred over the reference on the y axis to the JOD value on the x axis.

* Pwcmp: https://github.com/gfxdisp/pwcmp_rating_unified

5. Conclusions

In this manuscript, we presented a brief overview of VDP visual difference metrics and how they relate to other methods popular in display engineering. A brief outline of how metrics can be used for practical tasks is provided. We encourage readers to explore the full technical treatment of each VDP [1,2], and to try out these methods using the provided implementations.

6. Acknowledgements

We thank our colleagues at the Applied Perception Science team at Meta.

7. References

1. Mantiuk RK, Denes G, Chapiro A, Kaplanyan A, Rufo G, Bachy R, Lian T, Patney A. FovVideoVDP: A visible difference predictor for wide field-of-view video. ACM Transactions on Graphics (TOG). 2021 Jul 19;40(4):1-9.
2. Mantiuk RK, Hanji P, Ashraf M, Asano Y, Chapiro A. ColorVideoVDP: A visual difference predictor for image, video and display distortions. ACM Transactions on Graphics (TOG). 2024 Jul 28.
3. Mantiuk R, Kim KJ, Rempel AG, Heidrich W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Transactions on graphics (TOG). 2011 Jul 25;30(4):1-4.
4. Daly SJ. Visible differences predictor: an algorithm for the assessment of image fidelity. Human Vision, Visual Processing, and Digital Display III 1992 Aug 27 (Vol. 1666, pp. 2-15). SPIE.
5. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. 20th international conference on pattern recognition 2010 Aug 23 (pp. 2366-2369). IEEE.
6. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. The 37th Asilomar Conference on Signals, Systems & Computers, 2003 2003 Nov 9 (Vol. 2, pp. 1398-1402). IEEE.
7. Johnson GM, Fairchild MD. A top-down description of S-CIELAB and CIEDE2000. Color Research & Application. 2003 Dec;28(6):425-35.
8. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 586-595).
9. Mantiuk RK, Ashraf M, Chapiro A. stelaCSF: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. ACM Transactions on Graphics (TOG). 2022 Jul 22;41(4):1-6
10. Perez-Ortiz M, Mikhailiuk A, Zerman E, Hulusic V, Valenzise G, Mantiuk RK. From pairwise comparisons and rating to a unified quality scale. IEEE Transactions on Image Processing. 2019 Aug 28;29:1139-51