

Structured Spatial Domain Image and Data Comparison Metrics

Nivedita Sahasrabudhe⁴, John E. West^{1,4}, Raghu Machiraju^{2,4}, Mark Janus^{3,4}

¹DoD High Performance Computing Center, Information Technology Laboratory, USAE Waterways Experiment Station

²Department of Computer Science, ³Department of Aerospace Engineering, ⁴NSF Engineering Research Center, Mississippi State University

Abstract

Often, images or datasets have to be compared, to facilitate choices of visualization and simulation parameters respectively. Common comparison techniques include side-by-side viewing and juxtaposition, in order to facilitate visual verification of verisimilitude. In this paper, we propose quantitative techniques which accentuate differences in images and datasets. The comparison is enabled through a collection of partial metrics which, essentially, measure the lack of correlation between the datasets or images being compared. That is, they attempt to expose and measure the extent of the inherent structures in the difference between images or datasets. Besides yielding numerical attributes, the metrics also produce images, which can visually highlight differences. Our metrics are simple to compute and operate in the spatial domain. We demonstrate the effectiveness of our metrics through examples for comparing images and datasets.

CR Categories and Subject Descriptors: I.2.10 Vision and Scene Understanding; I.3.3 Picture/Image Generation; I.3.6 Methodology and Techniques; I.3.7 Three-Dimensional Graphics and Realism; I.3.8 Applications; I.4.0 General; I.4.3 Enhancement.

Additional Keywords: metrics, steering, rendering, correlation measure.

1 INTRODUCTION

One of the goals of any visualization algorithm is to produce a 2D image that is typically used for a variety of purposes. The image is judged for suitability by visual inspection, and if not deemed satisfactory a new image is obtained. Suitability is influenced by the rigors of the underlying application and the aesthetic leanings of the user. Seemingly appropriate parameters of the dataset and algorithm are then altered to produce yet another image. This refinement loop is often continued a few times. It would be useful to employ an image comparison metric that guides the refinement process. For instance, the number of slices in a volumetric rendering algorithm [20] for unstructured grids can alter the final quality of the image quite easily. It is certainly the case, that an exceptionally large number of slices will result in high-quality images. However, the computational cost can be staggering. Hence, a metric which measures the improvement in quality

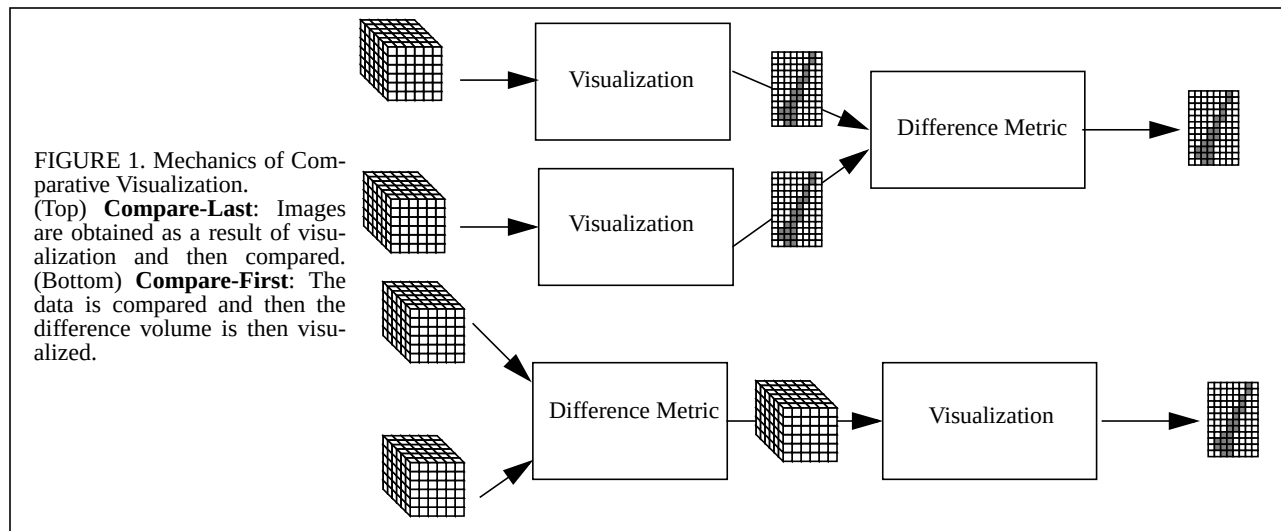
with increasing number of slices can be very useful. The metric, ideally, should identify beneficial or adverse effects of various rendering operations on the final image and facilitate the choice of parameters. In response, we propose structured spatial metrics in this paper. Numerical comparison metrics have been proposed elsewhere [3][15]. The proposed metrics differ from previous ones, in that, they operate mainly in the spatial domain and attempt to reveal structured differences. It is the structured differences that are discerned by the user during the iterative process of parameter selection.

The example described above illustrates the need for metrics to guide visualization algorithms. Essentially, the images generated by two different visualizations are compared and contrasted. Additionally, differences in computed solutions need to be discerned. For example, parallel and distributed implementation of complicated physical phenomenon is a necessity, given the need for fine temporal and spatial resolutions for certain problems. A visual presentation comparing solution differences using associated metrics and thresholds will help in steering parallel applications toward more efficient decomposition strategies. Similarly, one can also seek examples from design optimization [12] to illustrate the need for comparison metrics.

The need for objective comparative techniques is also motivated by the need to analyze and visualize very large quantities of data (*terascale visualization*). Traditionally, plots of quantities are displayed at various points and compared. This simple visual comparison has performed well when the datasets are not large or beset with uncertainty. It is very likely, that some very interesting features in the datasets are not compared in the case of large and perhaps noisy datasets. Better solutions include a side-by-side image-level comparison or a juxtaposition or feature-based comparison of datasets. Visualization techniques have helped the situation to some extent. Side-by-side comparison of images rely on the visualization algorithms to generate images which are actually compared. Juxtaposition requires an exercise in feature detection and further use of visualization techniques. The errors inherent in the visualization algorithms can mislead the comparative effort. In [16] a data-level comparison system was described. It was observed that visualization can in fact obfuscate the differences. Comparative visualization should allow for the detection and identification of the differences in an enhanced way and, if needed, allow for an objective measurement of the differences. In essence, comparative visualization tools should be able to: (a) detect and identify differences, (b) allow the user to quantify the differences easily, and (c) minimize the adverse effects of visualization techniques.

Our approach is to devise metrics that accentuate differences between images and datasets. The emphasis is either on the comparison of images or the data itself. In this paper we suggest that quantitative metrics may be devised to assist the researcher in evaluating the difference between solutions, and suggest some useful

¹ {nive,hatter,raghu,mark}@erc.msstate.edu
NSF Engineering Research Center for Computational Field Simulation
P.O. Box 9627
Mississippi State, MS 39762



features of such metrics. Comparison is achieved through an exercise in determining the lack of spatial correlation between two datasets or images. Our metrics operate in the spatial domain and will complement the appraisals of side-by-side comparisons and juxtapositions. Note that these metrics and methods are not suggested as a replacement for experience and intuition. The proposed work can be seen as an extension of some elements of the research conducted in [14] [16] and [17].

The methods described here can be applied to situations from other domains including medical imaging. In Section 2, we describe previous work. Later, in Section 3, we describe our metrics and in Section 4 we provide results. Finally, we summarize the paper and describe future work.

2 PREVIOUS WORK

It has often been said that images should (a) describe spatial relationships, (b) show the structure of data, (c) allow pattern matching approaches to problem solving, (d) get attention, and (e) invoke aesthetic appreciation. In many areas of visual and optical engineering, issues of quality in images (factors (a)-(c)) have been paid more than adequate attention. In image synthesis and visualization, factors (d) and (e) have dominated the discussion on image quality until lately. In this section we discuss work reported in the area of comparative visualization, objective metrics and associated issues of errors arising from visualization.

Comparative Visualization - Image-level and data-level approaches have been reported in the literature [4][14][16][17]. In the former approach, images are generated from each of the sources by a separate visualization pipeline. The pipelines can be tuned such that similar images are obtained. Image-level comparisons can be especially used to validate experimental techniques. Thus, optical techniques such as Schlieren photographs and physical techniques such as dye, smoke, and powder can be simulated numerically with ray and particle tracing techniques and the resulting images can be compared. Image level comparisons, however, rely on verisimilitude rather than data content. Data-level approaches, on the other hand, advocate that the data be transformed to a common representation and fed into the same pipeline. Silver and Zabusky reported an approach called visionmetrics [17]. The approach is based on detection of specific features such as vortex tubes and shock waves, resulting in explicit representation of feature objects. The common representation is necessary given the variety of grids in existence. Attributes such as size, shape, posi-

tion, local extremes, etc. can be calculated for each feature object. The features are also correlated and are compared qualitatively and quantitatively (juxtaposition) and the exercise is often referred to as visionmetrics. Helman and Hesselink [4] perform a feature extraction based on the critical point analysis of the vector field. The topological structure also serves as a low-level data representation. Pagendarm and Post provide a good review of the nascent field and also describe the comparative visualization of experimental Schlieren images and simulated ones [14]. They also assess the influence of the visualization techniques on the comparative effort. Although significant, none of these techniques have found their way into viable tools given their computational expense and lack of generality.

Objective Metrics - Image coding (compression) has received much attention from researchers, in terms of characterization of artifacts and errors. Many objective metrics have been proposed to measure the differences in images. A good review of computational image metrics is available in [1]. The earliest metrics: mean square error (MSE), root mean square area (RMSE), and peak signal-to-noise ratio are spatial domain metrics. However, these metrics often arrive at a global measure that can sometimes overstate the distortion a user perceives in the image [3]. This lack of correlation between the type of error in an image and the response of the human visual system to varying types of errors led researchers to begin developing image metrics. One of the first efforts in the application of perceptual metrics to image synthesis is discussed in Rushmeier [15]. After transforming and weighting by a HVS model, the metric computes the MSE between the two images as an overall quality measure. A variety of approaches using wavelets in various capacities as image metrics have been reported [3][5]. Gaddipatti et al. employed a metric to guide the visualization process [3]. Both [3] and [15] employed the Contrast Sensitivity Function (CSF) [8] to model the HVS. Williams and Usselton provide a good set of guidelines for measuring image quality in volume rendering [19]. In [16] histograms were used to allow comparison. However, the notion of the spatial origin is lost when histograms are used.

Researchers in medical imaging have long been concerned with the efficacy and fidelity of reconstruction algorithms. The fidelity of tomographic reconstruction algorithms have been studied through the creation of phantoms, designing figures of merits (FOM), and providing a statistical testing of hypothesis. A later effort by Miyahara et al. [9] discusses the application of the Picture Quality Scale (PQS) as a metric for the quality of JPEG-encoded images. This metric attempts to classify and quantify distortions in images by producing a set of partial distortion measures, which

may be combined in a non-redundant manner to produce an overall quality measure. The PQS is a spatial domain model, which attempts to measure the amount, location, and structure of the error perceived by the HVS. There are five partial distortion measures, each of which quantifies a different aspect of the image error. Most of the factors, in turn, are based upon the first factor, the luminance coding error (essentially a per-pixel difference weighted for contrast sensitivity). The remaining factors weigh the luminance coding error by a spatial frequency response model, end of coding block discontinuities, errors with strong spatial correlation, and errors in the vicinity of high contrast transitions. The factors are then assigned weights and combined to yield an overall quality measure. The proposed metrics are similar in spirit to the PQS.

Errors from Visualization - Very little work has been expended on characterizing the error from visualization algorithms. Some efforts in visualization of uncertainty have been conducted [19]. Much of the effort there, went into the use of visualization as a means to display the uncertainty that arises from transformations (sampling, smoothing, modeling) or rendering algorithms. Although laudable, it is perhaps equally important to understand how parameters of visualization influence operations and how the final quality is effected. Theoretical frameworks described in [7][10][13] are useful. However, more understanding is accrued from efforts similar to those described in [3] and [15]. Therein, the metric is used to explore the effects of different operators as they change in spatial or frequency extent. In the following section we explain our metrics.

3 COMPARATIVE METRICS

Comparative visualization can take different forms as shown in FIGURE 1. Moreover, the metrics themselves can be either evaluated in the spatial or frequency domains. In this section we describe our metrics. We first explain the choices we make in developing comparison metrics; the choices essentially explore some basic issues we think are important to the developing field of comparative visualization.

Scalar vs. Vector Fields - Questions about the statistical properties of vector fields are still being sought and answers are far from satisfactory. For scalar fields, answers to some of the same questions are more sound and are amenable to further numerical and computational use. We therefore restrict ourselves to scalar fields.

Spatial vs. Frequency Based Metrics - We are interested in spatial correlation (or lack of) between datasets. The difference metric should be a measure of the lack of correlation. It is for this reason we do not seek frequency based measures. However, frequency-based and wavelet-based methods can still be profitably employed to locate regions of high transition. For instance, the wavelet-based metric in [3] measures the difference in images in regions of higher transition.

Noise - Image quality metrics are often designed to be robust in the presence of noise. Experimental data is usually corrupted with noise. In computed numerical datasets, noise will appear as perturbations in solution values and can be introduced by changing numerical characteristics of new solution algorithms or by differences in numerical precision between different computer architectures. A Gaussian model of noise can be employed to characterize the noise that arises from numerical simulations.

Compare First or Last? - As FIGURE 1. shows it is possible to conduct comparative efforts in either of two ways. The first method requires that the individual volumes are visualized and the images are compared, while the second requires that a difference dataset be computed which is later visualized. The first method is efficient since images are compared; however, the visualization

algorithms can introduce uncertainties in the images and it is likely that final difference image is beset with errors. The errors from visualization algorithms are ill-understood and a computational scientist will not be able to make a viable comparison. Given the advantages of both methods, it is useful for any comparative metric to be applicable in either manner. The classification of methods described here is not dissimilar from that described in [14]. However, both the image and data level comparisons in our pipelines foster the use quantitative methods for actual comparison.

Multi-part Metric - It is rare that a single-faceted numerical measure can capture all the subtle differences between two datasets and images. We would like to be able to produce a set of partial metrics, M_i , which identify the detailed character of different types of local error in the data sets. Therefore, to obtain a versatile metric, it is useful to design multi-part metrics. This was the guiding rationale behind the PQS metric and has been successfully applied to characterize the fidelity of image compression algorithms [9].

Gridding Techniques - The metrics should work across all grids and exploit the underlying interpolative nature of the given grids. This assumption should hold trivially true when the two grids to be compared have the same topology and sampling resolution. Thus, for unstructured and curvilinear grids the measures can be cell-based and employ linear interpolation, if valid.

Role of Perception - Since we are interested in metrics that can be applied to compare both images and datasets, we do not include perceptual considerations in our metrics. However, they can be easily incorporated for image comparisons.

Metric Properties - How do we design a difference metric? First, we need a set of properties which hold over the metric M , measuring the distance between two data sets A and B as $M(A,B)$. First, $M(A,A) = 0$. Second, $M(A,B) > 0$, if A and B are even slightly dissimilar. Third, $M(A,B)$ is small if the differences between A and B are small; likewise, $M(A,B)$ is large if A and B are different. Fourth, if datasets A and B are different, and A and C are similar, $M(A,B)/M(A,C) > 1$. Another property which comes to mind as possibly desirable is symmetry; that is, $M(A,B) = M(B,A)$. The first four properties are somewhat self-evident. The fifth property provides a mechanism for the evaluation of the solution in terms of different types of differences separately. We propose partial difference metrics which satisfy the first four properties completely, although they are not completely symmetric. This is not an undesirable trait in a metric and can be used to advantage. Similar properties have been enumerated in [3] and [15].

3.1 Proposed Comparative Metric

In this section, we present our metrics. We propose that four partial measures of differences be considered: (a) an overall or global point measure, (b) a measure of the localization of differences or a lack of correlation, (c) differences in areas of high transition or singularities and, (d) areas of low transition or smoothness. This approach to metric design allows the user to find not only the magnitude of the difference between two datasets, but also provides a means of discerning the types of difference and where those differences are located. The proposed metric, M , is composed of four partial metrics: M_1 , M_2 , M_3 and M_4 .

The importance of a particular component of the difference metric will vary depending upon intended use of the simulation. For example, a researcher attempting to model the region around a shock very accurately, will be most interested in minimizing the distance between a computed solution and some benchmark in areas of high transition, while differences in other areas may not be of great concern. We now describe each of the partial metrics.

Global Difference M_1 - The first element of the metric, M_1 , is the L^2 norm. This metric can broadly distinguish between, for example, two subsonic flows computed over airfoils at low angles of attack, where the solutions are essentially similar, but air speed is increased in one simulation. By itself, however, it suffers from problems similar to those which afflict the MSE in image quality metrics [3]. In two dimensions the partial metric M_1 is computed as:

$$M_1 = \sqrt{\sum_{i,j} [A(i,j) - B(i,j)]^2}$$

where A and B are the images or datasets being compared.

Lack of Spatial Correlation and Zone of Error M_2 - The second partial metric, M_2 , can be thought of as a measure of the degree to which differences between two datasets are localized to specific regions. It bears a strong resemblance to the statistical measure of correlation between two series. Fixed size regions can be used to determine correlations. In data analysis, one would like to be able to weigh the relevance of an absolute aggregate measure of differences, such as that found in M_1 , with a measure of the size of the region over which these differences are spread. This partial metric provides a means to accomplish this, yielding a measure between 0 and 1, which grows larger as the portion of the data set over which differences exist grows larger. It is computed locally and then aggregated to form the comprehensive metric; the computation proceeds as a series of steps. First, the simple difference is computed at all sample locations:

$$d(i, j) = A(i, j) - B(i, j)$$

This difference is then used to compute the following quantity

$$m_2(m, n) = \sum_{k,l} |r(m, n, k, l)|^\tau$$

where τ is a parameter used to emphasize the presence of errors while de-emphasizing their magnitude. The results in this paper are reported using $\tau = 0.25$. The quantity $r(m, n, k, l)$ is the local correlation which is computed as:

$$r(m, n, k, l) = \frac{[A_1 - B_1]}{\theta}$$

where,

$$A_1 = \sum_{i,j} d(i, j) d(i + k, j + l)$$

$$B_1 = \frac{\sum_{i,j} d(i, j) \sum_{i,j} d(i + k, j + l)}{\theta - 1}$$

The sums are computed over the region of pixels where (i, j) and $(i + k, j + l)$ both lie in a $\theta \times \theta$ window centered at (m, n) . Computing correlation coefficients over the range $[1...k]$ and $[1...l]$ and then aggregating them provides a simple method by which features that persist over several distance scales can be given more weight than small scale, or highly localized, features.

The range $[-k...-1]$ and $[-l...-1]$ is not taken into account, since it will not yield any new distance scale which has not been included so far. In this case the window over which the correlation is computed is symmetric; this restriction may be relaxed in situations where correlations in one dimension are more relevant than those in another. Finally, the partial metric M_2 is computed as:

$$M_2 = \frac{1}{N} \sum_{m,n} \frac{m_2(m, n)}{\max(m_2(m, n))}$$

where N is the total number of pixels or grid samples. The metric is normalized to de-emphasize the magnitude of errors in favor of emphasizing their correlation. It may be useful to include the visualization of the partial metric itself in assessing the difference between two data sets. This may be accomplished by converting the range of partial metrics into $\langle r, g, b \rangle$ triples to produce a correlation image for visual inspection. These partial metric images are very effective in conveying information about the location of differences, as will be shown in Section 4.

Differences Near Areas of High Transition M_3 - The third partial metric, M_3 , is a measure of errors in areas of sharp transitions. It is of interest as a metric since we are concerned with measuring the distance between datasets which may contain solution discontinuities, and it is often the case that numerical schemes in computational field simulation are differentiated by their ability to capture discontinuities. This measure is computed by creating an activity mask from the reference dataset (A here) in the horizontal direction

$$V_h(i, j) = \frac{|A(i - k_h, j) - A(i + k_h, j)|}{\max[A(i, j)]}$$

where k_h is the extent of the activity window. A similar computation is performed in the vertical direction (k_v need not be the same as k_h), and the local partial metric is defined as

$$m_3(i, j) = |d(i, j)| [T_h(V_h(i, j)) + T_v(V_v(i, j))]$$

where the function $T_h()$ thresholds the activity mask, producing a binary mask which is only nonzero in locations where the activity level is above the threshold value. Selection of this threshold value will vary with both the application and the information a researcher desires from the metric. Note that if $T_h()$ differs from $T_v()$ the metric may be biased in favor of features in the direction of most interest to the researcher. Note also, that it is in the determination of the activity masks that this metric becomes asymmetric, $M(A, B) \neq M(B, A)$. The asymmetry arises because areas of high activity are first identified in the reference dataset, A , producing a mask which is used to extract and highlight differences between A and B . The third partial metric is computed as

$$M_3 = \sum_{i,j} m_3(i, j)$$

As with the second partial metric, both the set of partial metrics and the activity mask can be converted to an image for post-processing analysis, providing the researcher with more detailed information about specific regions of the data. Note that a suitable choice of a mask allows for directional features to be detected by designing suitable masks. Automatic methods can also be employed to create the masks. We can also employ the wavelet-based techniques described in [6] to create masks. In fact the metric in [3] can be used instead, for metric M_3 .

Differences in Areas of Smooth Transition M_4 - The fourth partial metric, M_4 , can be computed using the logical inverse of the thresholded activity masks computed for M_3 . It captures the differences in regions of small activity. This metric is computed in the same way as the partial metric M_3 .

To reiterate, our metric has the following salient features:

- It is spatial in nature; however, it can employ wavelet and frequency based methods to compute partial metrics M_3 and M_4 . This allows for a versatile metric that can be applied to several datasets from different domains.
- It is easy to add perceptual considerations into the metric. For

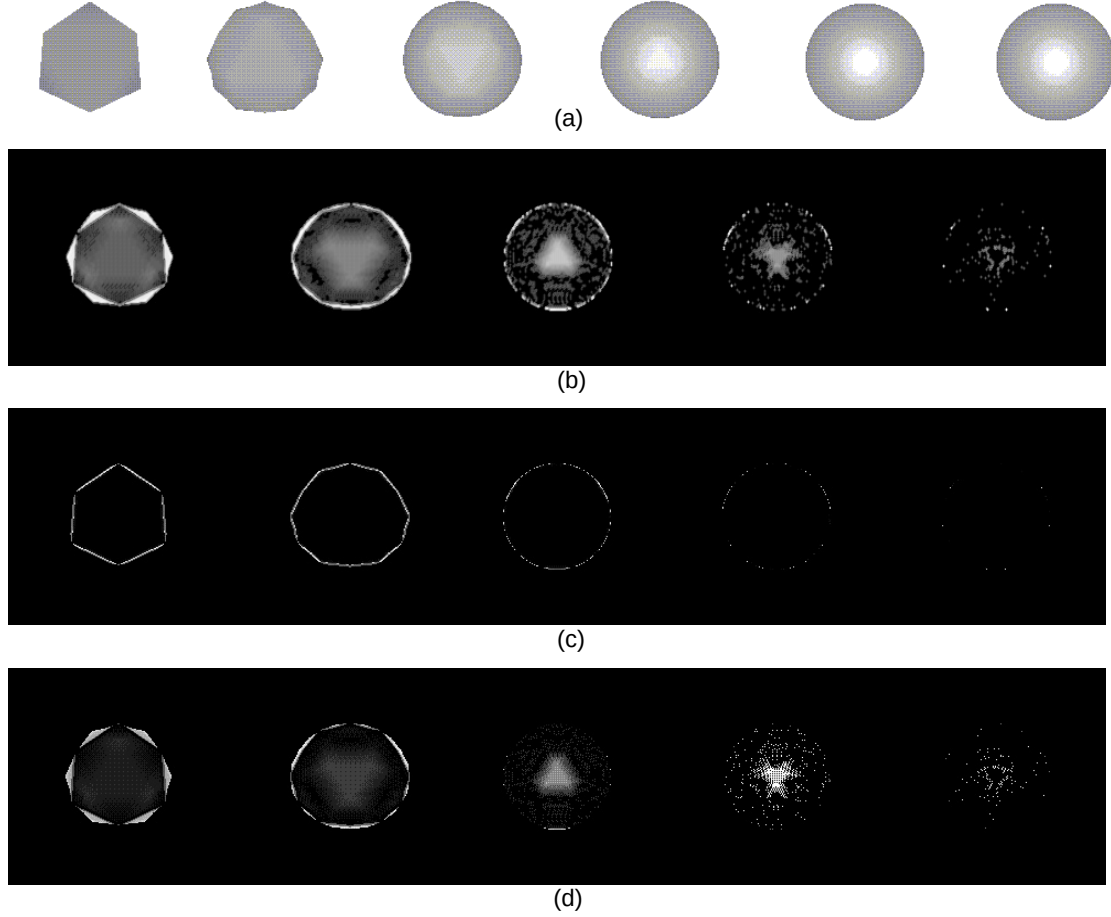


FIGURE 2. Example 1; Comparison of synthetic images with varying LOD: (a) Spheres with LOD 1 through 6 (left to right). (b) Second partial metric images for comparison of successive images from (a). (c) Third partial metric images. (d) Fourth partial metric images.

instance, both partial metrics M_3 and M_4 can include a computational model of the Human Visual System since they operate in regions of high and low transitions.

- Again, by a suitable choice of domain, the metric can be made very robust to noise. For datasets from CFS noise may not be major consideration; however for datasets from medical domain, effective denoising is required. Once again, by suitable choice of thresholds and masks one can discriminate between noise and useful information.
- The resulting metric, however, will not be commutative in nature, especially given the use of a reference mask employed in the computation of partial metrics M_3 , and M_4 . However, the metric can respond suitably to differences of varying magnitudes, satisfying the first four properties listed in Section 3.2.
- The metric not only quantifies the differences but can also depict the spatial spread of errors.
- The metric allows for a user-driven choice of region of interest. Otherwise, an automatic method based on edge-detection or wavelet coefficient magnitudes can be employed [6].

4 RESULTS

We illustrate the effectiveness of the proposed metrics through some examples. The first two deal with image comparisons while the last two examples examine comparison of datasets.

The images referred to in the latter, are visualizations of the datasets that were compared.

Example 1: Response of Metrics to Data and Visualization/Rendering Parameters - In an attempt to subject the proposed metrics to rigorous validation, we test their performance on synthetically generated images.

The basic test image we employ is a simple sphere, obtained by recursive subdivision and subsequent normalization of a regular tetrahedron, in the center of a window of 200×200 pixels. The discretized tetrahedron is projected onto a 2D image using the Z-buffer algorithm. The intent of the experiment was to measure the efficacy of the metrics in capturing changes in the projected detail. If effective, such metrics can be used to perhaps dictate choices between models with varying level-of-detail (LOD). In order to obtain various images, we changed the coarseness of the sphere or the level-of-detail, by changing the number of times the tetrahedron was subdivided. All the other parameters which could affect the rendering of the sphere - such as the material properties, the position and type of light, etc., were constant. Images were obtained by varying the LOD from 1 to 6, and successive images were compared to evaluate M_1 , M_2 , M_3 and M_4 . FIGURE 2.a shows the test images for LOD increasing from 1 to 6. We set our high activity mask such that it would pick out only a significant difference in intensity levels of neighboring pixels, for example - a change from white to gray. For our test images, such changes occur

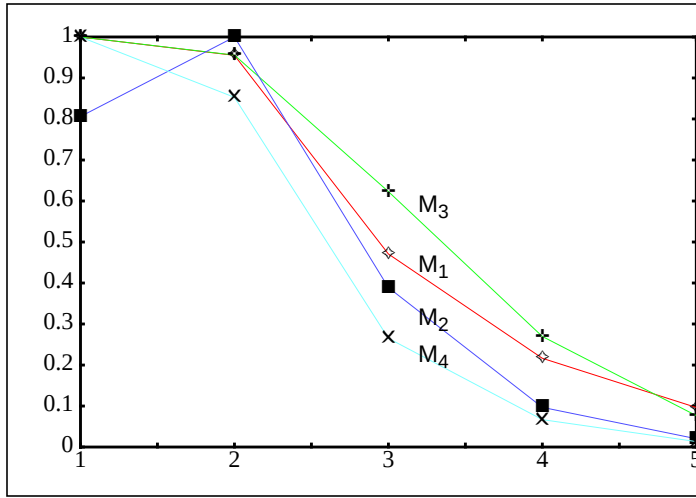


FIGURE 3. LOD Example. Plots of normalized partial metrics for successive comparisons. Each metric value is divided by its maximum.

only at the periphery. Inside the periphery, we have only smooth changes.

FIGURE 2.b, shows the partial metric images for comparisons of successive test images from FIGURE 2.a. The first row of images show the spread of differences and indicate that as LOD increases, the error becomes more and more localized near the center of the sphere. The images in FIGURE 2.c, show the differences in the areas of high transition. They describe changes near the periphery just as expected. Finally, the last row of images show differences in areas of smooth transition. It is very clear that the metrics relate the changes in spatial detail with the changes in detail of projected images. FIGURE 3. shows a graph of the normalized partial metric values. The behavior of the plots reflects the observations made above. The partial metric M_1 does not fall rapidly in numerical value; the images also show significant differences in the earlier trials. M_2 on the other hand shows an increase in the differences and falls only after the third and fourth images are compared. Lastly M_3 decreases monotonically as M_4 does.

We performed similar experiments with shading parameters. The metrics yielded results concurrent with the images obtained.

Example 2: Volume Rendering - We tested our metric on images obtained from a slice-based volume rendering algorithm for unstructured grids, generated from tetrahedral cells [20]. The algorithm slices the unstructured grid parallel to the screen and then composites the resulting triangles from each slice. The images suffer primarily from aliasing when the number of slices is small. However, large number of slices will result in blurring of features. This can be seen from PLATE 1. We employ our metrics to gauge the effect of a varying number of slices on the *blunt fin* dataset. FIGURE 4. shows the M_2 partial images for comparisons between images rendered with different number of slices. The persistent nature of the aliasing artifacts is not only visible from all the second partial images, but evident from Table 1 also.

Another noteworthy aspect is that the metric brings out very structured differences. The errors between images emanate from the absence or presence of features at all spatial locations. The partial images of M_3 (high transition differences) do not reveal much information in this case; they only reveal differences along the boundaries of the flow domain. Finally, M_4 (low transition differences) registers not much change between successive partial images. This example was also tested against a wavelet based metric proposed in [3]. The wavelet-based metric performed well to capture transition regions. However, it fared poorly to capture structured differences and did not really indicate the presence of aliasing for very large number of slices.

Example 3: Computational Flow Simulations - PLATE 2.a shows the visualization of the results of simulations of the flow around a cylinder for one of the computed quantities: the scalar density value of the flow. The flow enters the domain from the left, exiting on the right. The solutions compute the steady flow using both first and second order (Warming-Beam) upwind methods. A stagnation point exists to the left of the cylinder and separation points exist at either end of the cylinders. The color maps in the images use a continuous color spectrum proceeding from black through blue, green, yellow, and red for low through high data values. The images were produced with the Flow Analysis Software Toolkit (FAST) [2].

Table 1: Normalized partial metric values

Comparison	M_2	M_3	M_4
20 vs. 10	0.88	1.00	1.00
30 vs. 20	0.91	0.53	0.55
40 vs. 30	0.91	0.42	0.39
50 vs. 40	0.83	0.35	0.30
60 vs. 50	0.82	0.30	0.25
70 vs. 60	0.94	0.24	0.21
80 vs. 70	0.93	0.19	0.18
90 vs. 80	1.00	0.19	0.17
100 vs. 90	0.99	0.18	0.16

The first-order method characterizes the physical phenomenon less accurately; the separation points are shown to be askew from the vertical axis. A side-by-side comparison yields insight to an extent of the gross features including the stagnation point and the separation points. The spread and the extent of the difference between the two datasets is however, not easily discernible. Table 2 displays the results of applying the metric to the datasets, using the first order solution as the reference dataset. The second partial metric shown in Table 2 indicates that the differences are concentrated in a reasonably small region of the solution domain. The values for the third and fourth partial metrics show that these errors are con-

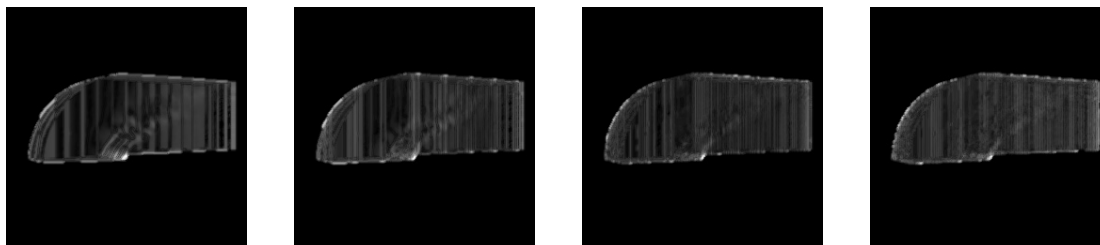


FIGURE 4. Example 2; Comparison of volume rendered images with varying number of slices: Second partial metric images for comparison of images for $n=20, 30, 50$ and 90 slices and $n-10$.

centrated almost exclusively in areas of low activity. A relatively small threshold value was used in developing the activity masks, so both regions of high and low activity were identified even though the solution does not contain any shocks.

The partial metric images are most useful when used in conjunction with the aggregate figures presented in Table 1. PLATE 2.b shows the three partial metric images. The image for the second partial metric shows that differences are concentrated in the region surrounding the cylinder, and are highest in the flow immediately adjacent to the cylinder. The image for the third partial metric shows that the region of high activity is located to the rear of the cylinder, and differences in this area are small. The final partial metric image reveals that differences in regions of low activity are highest around the stagnation point, and coincide with differences evident in solution quality as shown in PLATE 2.a. In the case of datasets, the quantitative numerical values can be more useful. However, images can still provide a wealth of information.

Table 2: Metric values for the first and second order cylinder solutions

Metric	Value
M_1	1.704
M_2	0.188
M_3	5.112
M_4	95.00

Example 4: Multidisciplinary Design Optimization - As a final example, we investigate the potential use of this metric in the field of Multidisciplinary Design Optimization (MDO). The flow through a turbine cascade was chosen as a test case. For comparison to traditional means of computing sensitivity derivatives, first we compute the x -component of the fluid momentum. We then repeat the computation using a forward perturbation (an increase) and a backward perturbation (a decrease) of the incoming Mach number. A comparison is then made through the use of the partial metric M_2 , to determine if comparative visualization can provide researchers and design engineers with information to qualitatively evaluate design sensitivity. In the MDO arena, a similar procedure is used to evaluate design sensitivity derivatives and is referred to as a finite difference approach. An additional approach utilizes the analytic derivatives of the function in question and is quite a laborious task. PLATE 3.b shows the finite difference image and the image for partial metric M_2 . In the latter, the band on the lower right edge of the blade (the location where the computational grid merges with itself) appears because in the comparative visualization process, data is presently not shared across this domain "boundary". A design cycle requires the designation of a particular objective function(s) which the software is instructed to improve or

optimize. In general, the use of surface measurables and/or integrated performance criteria are used as objective functions. With visualization as a vehicle, the analytical concepts being developed regarding field comparison open an additional avenue to the design engineer using MDO.

5 SUMMARY AND FUTURE WORK

In this paper, objective metrics for comparison were suggested, and results of their application to a variety of visualization scenarios were examined. The proposed structured multi-part metric provides a set of measures which effectively describe not only the magnitude of differences, but also their type. In addition, the images produced from the local version of each partial metric, yielded detailed information about the distribution of each type of difference throughout the solution domain in an intuitive and engaging format. We demonstrated our methods on four examples. These metrics can be employed to compare images and datasets alike and facilitates the comparison of datasets in a variety of ways. Finally, the framework allows the user to also specify activity masks and examine in depth the regions of his or her choice.

Overall, the results of this early investigation are encouraging. However, much interesting work remains to be done in the following areas:

- **Generation of masks** - The effectiveness of the activity masks in isolating regions of activity, is sensitively dependent upon the user's understanding of the data being investigated and his intelligent choice of a threshold value. There are recent results which suggest that this procedure can be effectively automated [6]; incorporating this approach may yield significantly better results than the current, simpler approach.
- **Errors of visualization**: Any display technique will corrupt the difference image; studies are needed to understand the uncertainty introduced.
- **Combination of Partial Metrics**: For more quantitative insights it may be necessary to combine the various partial metrics. Since, some of the partial metrics measure different aspects of the same underlying structural differences of the images being compared, a decorrelation effort is perhaps useful. The decorrelation effort can be achieved through a Principal Component Analysis or a KL-Transform.

6 ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under Grant No. ACI-9734483 (CAREER); Hearin Research Foundation Grant, Mississippi State University. Thanks also go to Dr. James Newman III, Mississippi State University and Dr. Kwan-Liu Ma, University of California-Davis, for useful comments.

REFERENCES

- [1] Ahumada A. J., "Computational Image Quality Metrics: A Review," *Society for Information Display International Symposium Digest of Technical Papers*, Vol. 24, Playa del Rey, CA, pp. 305-308.
- [2] Bancroft G., Merritt F., Plessel T., Kelaita P., McCabe R., Globus A., "FAST: A Multi-Processed Environment for Visualization of Computational Fluid Dynamics," AIAA Paper 91-0793, *Proceedings of the 29th Aerospace Sciences Meeting*, Reno, NV, January 1991.
- [3] Gaddipati A., Machiraju R., Yagel R., "Steering Image generation using Wavelet Based Perceptual Metric," *Computer Graphics Forum* (Proceedings of Eurographics '97), September 1997.
- [4] Helman J. L., Hesselink L., "Visualizing Vector Field Topology in Fluid Flows," *IEEE Computer Graphics and Applications* 11(3), May 1991, pp. 36-46.
- [5] Jacobs C. E., Finkelstein A., Salesin D. H., "Fast Multiresolution Image Querying," *Computer Graphics*, Proceedings of SIGGRAPH 95, pp. 277-286, 1995.
- [6] Machiraju R., Gaddipati A., Yagel R., "Detection and Enhancement of Scale Coherent Structures Using Wavelet Transform Products," *Proceedings of the Technical Conference on Wavelets in Signal and Image Processing*, SPIE Annual Meeting, San Diego, CA, July 27 - August 2, 1997.
- [7] Machiraju R., Yagel R., "Reconstruction Error and Control: A Sampling Theory Approach," *IEEE Transactions on Visualization and Graphics*, vol. 2, No. 3, December 1996.
- [8] Mannos J. L., Sakrison D. J., "The Effects of a Visual Fidelity Criterion on the Encoding of Images," *IEEE Transactions on Information Theory*, 20(3):525-536, July 1974.
- [9] Miyahara M., Kotani K., Algazi V., "Objective Picture Quality Scale (PQS) for Image Coding," *Technical Research Report*, Department of Electrical Engineering, University of California at Davis, May 1997.
- [10] Moeller T., Machiraju R., Muller K., Yagel R., "Evaluation and Design of Optimal Filters Using a Taylor Series Expansion," *IEEE Transactions of Visualization and Graphics*, Volume 3, Number 2, June, 1997.
- [11] Moeller T., Machiraju R., Muller K., Yagel R., "A Comparison of Normal Estimation Schemes," *Proceedings of Visualization '97*, Phoenix, AZ.
- [12] Newman III, J.C., Taylor III, A.C., and Barnwell, R.W., "Aerodynamic Shape Sensitivity Analysis and Design Optimization of Complex Configurations Using Unstructured Grids", *AIAA Paper* 97-2275, June, 1997.
- [13] Novins K., Arvo J., "A Power Series Algorithm for Highly Accurate Volume Rendering," *Proceedings of 1992 Workshop on Volume Visualization*, Boston, October 1992, pp.83-90.
- [14] Pagendarm H-G., Posts F. H., "Comparative Visualization - Approaches and Examples," *Fifth Eurographics Workshop on Visualization in Scientific Computing*, Rostock, Germany, May 30 - June 1, 1994.
- [15] Rushmeier H., Ward G., Piatko C., Sanders P., Rust B., "Comparing Real and Synthetic Images: Some Ideas About Metrics," *Proceedings of Sixth Eurographics Workshop on Rendering*, Dublin, Ireland, 1995, pp. 82-91.
- [16] Shen Q., Pang A., Uselton S., "Data Level Comparison of Wind Tunnel and Computational Fluid Dynamics Data," *Proceedings of Visualization '98*, Research Triangle Park, NC, pp. 415-418.
- [17] Silver D., Zabusky N. J., "Quantifying Visualizations for Reduced Modeling in Nonlinear Science: Extracting Structures from Datasets," *Journal of Visual Communication and Image Representation*, 4(1), March 1993, pp. 44-61.
- [18] Williams P. L., Uselton S. P., "Foundations of Measuring Volume Rendering Quality," *NAS Technical Report*, NAS-96-021, NASA Ames, CA.
- [19] Wittenbrink C. M., Pang A. T., Lodha S. K., "Glyphs for Visualizing Uncertainty in Vector Fields," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 2, No. 3, September 1996.
- [20] Yagel R., Reed D. M., Law A., Shih P., Shareef N., "Hardware Assisted Volume Rendering of Unstructured Grids by Incremental Slicing", *Proceedings of 1996 Symposium on Volume Visualization*, San Francisco, CA, September 1996, pp. 55-62.

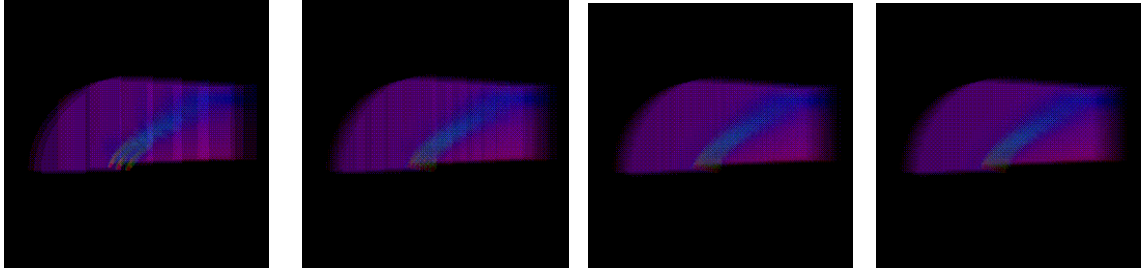


PLATE 1. Volume rendered images of the blunt fin dataset: The number of slices varying as 10, 20, 40 and 80 slices from left to right.

PLATE 2. Example 3; Flow around a cylinder. The density variation is shown with a spectral color map: (a) Density around a right circular cylinder computed using first (left) and second order (right) upwind methods. (b) Partial metric values M_2 , M_3 and M_4 , respectively (left to right), on comparing first and second order methods.

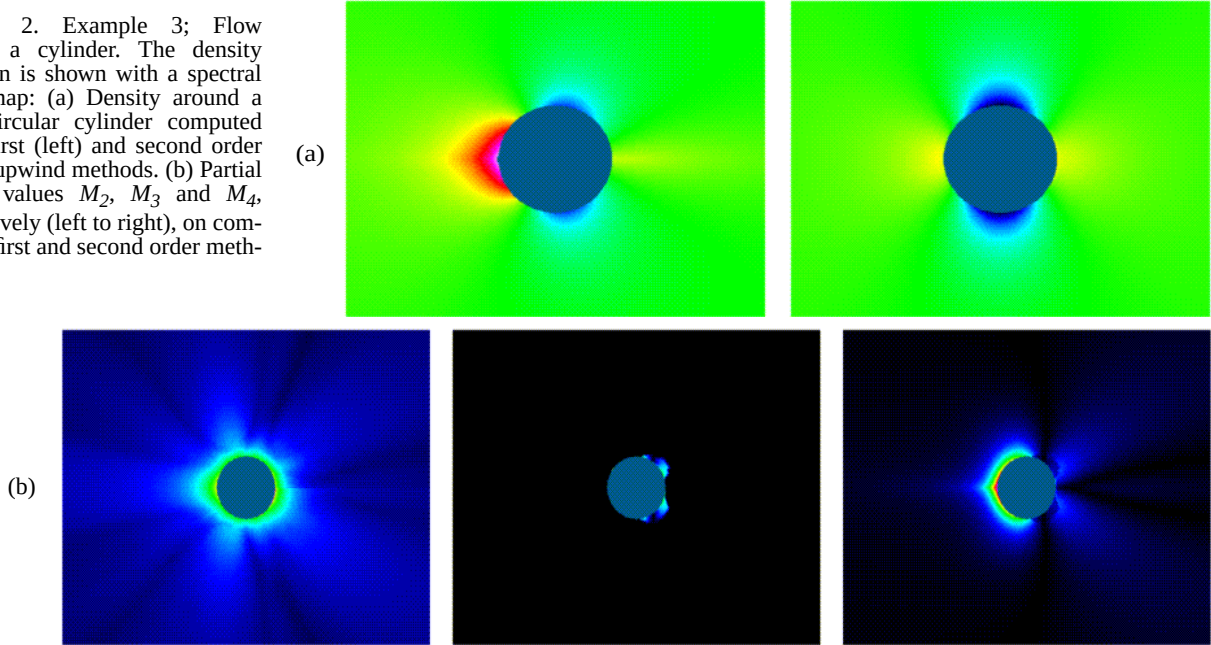


PLATE 3. Example 4; MDO: (a) X -momentum for a negative(left) and a positive(right) perturbation in Mach number. (b) The finite-difference of the X -momentum with respect to the Mach number(left) and the M_2 image(right) for comparison of datasets corresponding to images in (a).

