

Visual Saliency Metrics for Image Inpainting

Paul Ardis, Amit Singhal

Copyright 2009 Society of Photo-Optical Instrumentation Engineers.

This paper was published in the Proceedings of the SPIE, vol. 7257, and is made available as an electronic reprint with permission of SPIE. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Visual salience metrics for image inpainting

Paul A. Ardis^{a*}, Amit Singhal^b

^a University of Rochester, Rochester, NY, USA 14627

^b Research Laboratories, Eastman Kodak Company, Rochester, NY, USA 14650-2103

ABSTRACT

Quantitative metrics for successful image inpainting currently do not exist, with researchers instead relying upon qualitative human comparisons to evaluate their methodologies and techniques. In an attempt to rectify this situation, we propose two new metrics to capture the notions of noticeability and visual intent in order to evaluate inpainting results. The proposed metrics use a quantitative measure of visual salience based upon a computational model of human visual attention. We demonstrate how these two metrics repeatedly correlate with qualitative opinion in a human observer study, correctly identify the optimum uses for exemplar-based inpainting (as specified in the original publication), and match qualitative opinion in published examples.

Keywords: Metrics, Inpainting, Visual Salience, Attention Modeling, Digital Editing

1. INTRODUCTION

Over the past decade, a number of significant advances in computer vision have been produced, cited, and integrated into other work on the basis of qualitative results. Methods for automatic inpainting of missing image or video data have been at the forefront of this trend, relying on flashy presentation and reader intuition in order to win mainstream approval. Leading publications bemoan the lack of a quantitative measure of inpainting success,¹ yet to date there has been no concerted effort to produce a technique for meaningful numerical comparison of inpainting results. In order to fill this need, we present two novel metrics for inpainting success and visual intent preservation, and demonstrate their use and correctness.

Although there does not exist prior art in the field of metrics for inpainting success, general attention-based metrics of image fidelity²⁻⁵ provide motivation for the proposed metrics. Fidelity metrics, however, are not directly suitable for inpainting as they only seek to quantify the accuracy of reproduction of the original image and thereby do not allow for visually plausible alternatives. Furthermore, these methods have traditionally involved simplistic models of human attention in order to determine visually salient points and regions, ignoring more complete and well-recognized psychophysical models due to computational constraints or author unfamiliarity. We improve upon this by adapting the recently updated iLab Neuromorphic Vision Toolkit (iNVT)^{6,7} to produce metrics for inpainting evaluation based upon relative pre-saccadic salience. In doing so, we indicate the general usefulness of saliency map representations⁸⁻¹² for inpainting analysis.

Throughout this paper, we demonstrate the effectiveness of our metrics with regard to exemplar-based inpainting,^{1,13} the most popular modern variation of isophote-based image inpainting,^{14,16} where image intensity and coloration is treated as fluidic and neighborhood-based search is used to determine likely texture continuity surrounding extended structure. Although a number of other inpainting methods exist, exemplar-based methods have proven popular in many traditional and modern uses (*e.g.*, painting restoration, text removal, object removal). To provide an unbiased ground for applying exemplar-based inpainting, we studied its metric and qualitative scoring in pseudorandomized scenarios, where pseudorandomness is provided by way of the strongly aperiodic Mersenne Twister algorithm,¹⁷ with a comparison against fast raster-order linear interpolation for further intuition of relative scoring.

* ardis@cs.rochester.edu; Dept. of Comp. Sci., Univ. of Rochester, P.O. Box 270226, 734 Comp. Studies Bldg., Rochester, NY 14627

The remainder of this paper is structured as follows: a definition of the two proposed metrics (Section 2), a study of the correlation between metric scores and qualitative opinion (Section 3), an analysis of the metrics in use (Section 4), and our conclusions (Section 5).

2. DEFINITIONS

The first proposed metric, *Average Squared Visual Saliency (ASVS)*, corresponds to the normalized sum of squared saliency of inpainted pixels, computed as

$$(1/|\Omega|) (\sum_{\Omega} (S'(p))^2) \quad (1)$$

where $S'(p)$ is the pre-saccadic saliency map value (a non-negative upper-bounded number), as computed by a computational human attention model, that corresponds to post-inpainting pixel p in the set of inpainted pixels Ω . The intent behind this non-linear metric is to provide a larger score for those regions that are of very high saliency (indicating more likely foveation targets, assuming that saliency is monotonically increasing with real observer visual interest). As inpainting is intended to go without notice by a typical unprepared observer, ASVS provides a score to determine if the inpainted pixels are relatively noticeable as compared to the remainder of the scene. Rather than studying only the likelihood of *any* pixel being the *global* saliency maximum and receiving observer attention (as would be the case with $\max((S'(p))^2)$), ASVS identifies if there are *one or more* pixels that may be *primary or secondary* saliency peaks and therefore receive attention following (less-predictable) post-saccadic saliency update. Normalization is provided to allow for the comparison of ASVS across inpainting situations of varying size.

As saliency computation contains a number of relative components (*i.e.*, feature contrast, center-surround effects), it is possible to produce a low ASVS score by artificially increasing the saliency of unpainted pixels via relative effect. In such cases, the inpainted regions themselves may go without notice yet the resulting image will likely receive a flow of attention different from the original, thereby changing the creator-intended (in paintings) or human-typical (in photographs) flow of attention and disrupting those observers already familiar with the original. The second proposed metric, *Degree of Noticeability (DN)*, is intended to identify noticeable inpainting regions as well as providing an indication of attention change beyond the inpainted region; DN is computed as

$$(1 / (|\Omega| + |\Phi|)) ((\sum_{\Omega} (S'(p))^2) + (\sum_{\Phi} (S'(p) - S(p))^2)) \quad (2)$$

where $S(p)$ is the saliency map value that corresponds to *pre*-inpainting pixel p in the set of non-inpainted pixels Φ . Building on ASVS, DN will additionally identify removed or added primary or secondary saliency peaks within the remainder of the image, and is similarly normalized for comparison across varied-size uses.

Metric precision was set at ANSI/IEEE Standard 754-1985 double precision (64 bits on a 32-bit architecture) for the purposes of this paper. Saliency maps were produced using version 3.1 of iNVT at 1:16 discretization of scale-4 (noiseless) expected visual cortex stimulation with 0.1 ms observation cutoff, 4 orientation channels, 3 center scales (2 to 4), and 2 center-surround channels (3 and 4).

3. OBSERVER STUDY

In order to verify that the proposed metrics matched general human opinion, we performed a series of observer tests, where qualitative scores were gathered for 15% of the metric-scored images pseudorandomly interspersed with copies of the original images (introduced as 10% of viewed samples as a control). Observation was for 5 seconds each, cued by presenting the original (pre-inpainting) image for 5 seconds prior to 3 seconds of pause (to counteract motion effects) and the scoring of corresponding inpainting and control images presented at a fixed distance of approximately 0.75 m and with images approximately 20 cm in height. Scoring was between 1 and 4, using the criteria presented in Table 1. The average quantitative score per qualitative value is shown in Table 2. Control images received correct “1” scores 87% of the time, with the other 13% receiving a “2.” Images were pseudorandomly distributed to observers, with each

image scored exactly twice and a total set of five viewers participating; we hope to perform additional repetitions to confirm our findings, and encourage related results.

Although increased scoring appears to strongly correlate with increased observer notice (as expected, and including subjective evaluation of visual acceptability for DN) a few high-ASVS images received a qualitative “1,” indicating that the inpainted region went unnoticed. As was conjectured earlier, we found that these images were those where high-salience inpainted regions remained non-peak due to increased salience of non-inpainted regions. Although attention was strongly directed, it was directed toward regions that had not been changed, therefore causing the observer to spend the 5 seconds in scrutiny of original texture and believe that there were not any significant alterations. This was quantitatively confirmed by a DN score of at least 3.0000, where “1” images with any lesser DN (averaging less than 1.0000×10^{-3}) resulted in an average ASVS score of 83.84; while still not well-differentiable from “3” scorers, this at least widens the boundary with those receiving a score of “4” (whose identification is most useful in application).

Table 1. Criteria for Observer Scores

QUALITATIVE SCORES	
1	The inpainting region was not identifiable within 5 seconds of continued viewing.
2	The inpainting region was identified, but was visually acceptable (would appear natural if not cued, and was considered consistent with scene content).
3	The inpainting region was identified and was not visually acceptable, but was not noticed (by the scorer) for at least 2 seconds.
4	The inpainting region was identified within 2 seconds and was not visually acceptable

Table 2. Quantitative Values for each Qualitative Score

AVERAGE METRIC PER QUALITATIVE SCORE				
	1	2	3	4
ASVS	97.04	69.50	35.31	334.9
DN	1.247	2.244	3.088	18.62

Example images receiving each score are shown in Fig. 1.

As a result of this study, we conclude that the proposed metrics do correctly (but imperfectly) measure the phenomena that they are intended to (*i.e.*, increased notice of the inpainted region, significant attention change); this was further confirmed by the full results of metric testing.



Fig. 1. Clockwise from top-left (originals appearing at right, with affected region outlined in red): inpainted image receiving a qualitative score of “1” (a), image receiving a “2” (b), image receiving a “3” (c), image receiving a “4” (d).

4. METRICS IN USE

DN and ASVS scores were computed for exemplar-based inpainting and fast raster-order linear interpolation (to provide intuition of relative scoring), where pixels were selected for inpainting pseudorandomly (Fig. 2b), as pseudorandomly placed squares (Fig. 2c), or as pseudorandomly placed and sized “scratches” (Fig. 2d). These applications were chosen to mimic those typically discussed in the literature as well as those selected for publication in scholarly articles on exemplar-based inpainting, and will be further discussed as metric results relate to studied qualitative examples.



Fig. 2. Clockwise from top-left: original image (a), 10% of pixels pseudorandomly selected for inpainting (b), pseudorandomly placed square (48 pixels in width) selected for inpainting (c), 17 pseudorandomly placed scratches selected for inpainting (d). Shown is “A Tale of 1001 Nights” by Gustave Boulanger, with pixels selected for inpainting shown in bright green.

* Each “scratch” is a vertical rectangle of pixels from 1 to 5 pixels in width and 31 to 90 pixels in length (sized pseudorandomly).

In keeping with the inpainting literature, where images are selected for traditional use (*i.e.*, painting restoration) or modern use (*e.g.*, object removal), our dataset consisted of two well-known paintings (“Mona Lisa” by Leonardo da Vinci and “A Tale of 1001 Nights” by Gustave Boulanger) and two novel digital photographs (dubbed *Wolf* and *Woman* to match their main subjects). Experiments were run for each image, using the above-mentioned methods to select increasing numbers of pixels, with this procedure performed repeatedly in order to gather representative average results.

When inpainting pseudorandom pixels, consistently elevated ASVS scores* and elevated-and-increasing DN scores† indicated that exemplar-based inpainting performed poorly (a fact confirmed by qualitative scores that were regularly higher than for scratch- or square-removal of comparable numbers of pixels) and quickly rose to regular “3” or “4” ratings. Exemplar-based methods did not fare significantly better or worse than linear interpolation techniques (Fig. 3), although they were noted to perform slightly worse as the percentage of inpainted pixels increased beyond 7% due to the increasingly small number of feasible “exemplars” (inpainting-free fixed-sized blocks of pixels) that can be used to fill in the holes; continued experimentation identified 13–15% selection as resulting in frequent inpainting failure, where no feasible exemplars are available. These findings were expected, as the large number of inpainting regions and small number of examples to draw upon results in an increased likelihood of at least one obvious mistake. As has been noted by Criminisi *et al.*^{1, 11}, curves are not guaranteed to have correct continuity as isophote extension is a linear process; therefore, pseudorandomly selected pixels falling on curved boundaries are likely to cause artifacting (correctable by precise neighborhood matching if a large number of exemplars is available, although this is not possible in the pseudorandom inpainting scenario). Given that such conditions are likely to arise in general use, we conclude that an exemplar-based approach to random pixel inpainting is a poor choice, and that the original authors were correct in not identifying this scenario as an area of method success.

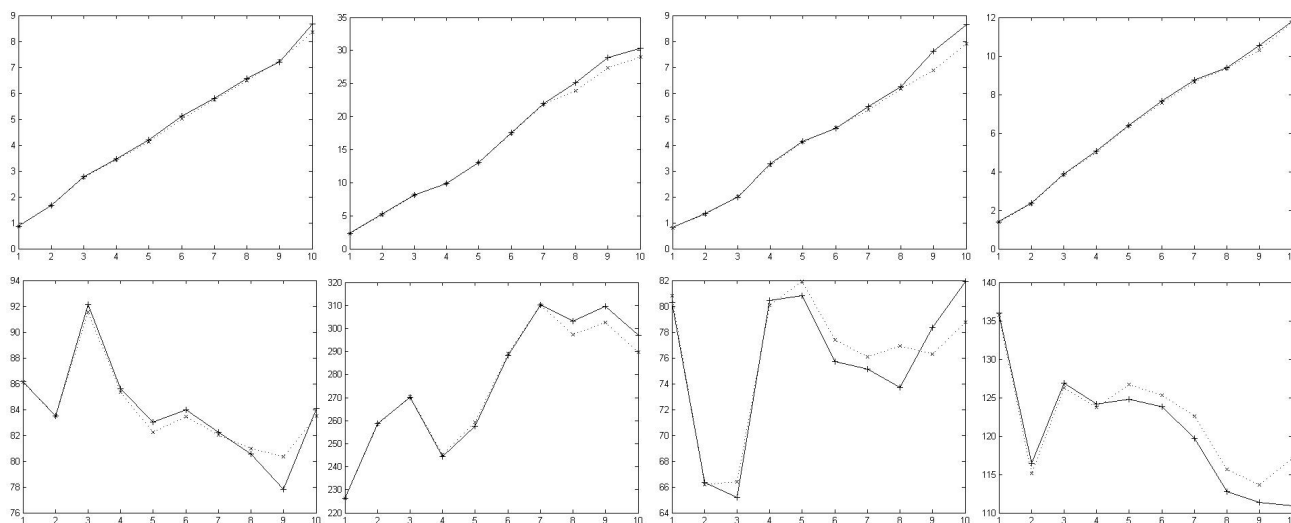


Fig. 3. From top to bottom: DN vs. percentage of pixels pseudorandomly selected for inpainting; ASVS vs. percentage of pixels selected for inpainting. From left to right: “A Tale of 1001 Nights,” “Mona Lisa,” *Wolf*, *Woman*. Metric scores for exemplar-based inpainting appear as solid lines (with upright hashes), while metric scores for raster-order linear interpolation appear as dashed lines (with angled hashes).

The results from inpainting of increasingly sized squares are more revealing, as a number of published examples (*e.g.*, “Removing large objects from images.” “Comparison with ‘Fragment-Based Image Completion’”)¹ corresponds to inpainting problems of similar form: medium-to-large thick regions with varied protruding structure and heterogeneous data available for completion. Both metric scores for square inpainting appeared as drastic peaks in an otherwise zeroed

* ASVS scores averaged at least 60.00 for all experiments for each image.

† DN scores averaged at least 5.000 for all experiments for each image, and were monotonically increasing at an average rate of at least 0.5000 per increased percentile for each image.

plane (Fig. 4), where DN peak frequency and amplitude increased with problem size for all images tested while ASVS displayed similar spiking but less clarity of spike trending. These results matched observer comments: inpainting was either subtle and reasonable, or contained large regions of copied texture that did not match local expectation (and therefore were immediately visible as well as almost always being considered visually unacceptable). Qualitative scores reiterate this: results of constant-valued thresholding for ASVS and DN are shown in the precision-recall graph of Fig. 5, indicating that simple threshold-based testing can be used as a reasonable means of rejecting inpainting results that are likely to contain noticeable artifacting.

The separation between exemplar and interpolation DN and ASVS scores is as expected: exemplar-based methods perform at least as well as the fast-but-imprecise interpolation process, as copied texture often matches the surrounds better than splotches of blurred color. It is when region completion results in incorrect copying (due to similar linear structures that appear in different textural neighborhoods) that threshold-detectable artifacting occurs; the relative rarity of these occasions (occurring in less than 20% of all result images) across the thousands of performed inpaintings encourages the continued use of exemplar-based methods for large region inpainting, confirming original author sentiments of generalized success. However, we strongly encourage the adoption of pre-output metric thresholding to determine if a result is likely to contain (potentially correctable) visual mistakes, as the introduction of a simple rejection threshold can greatly reduce the frequency of high-scoring images.

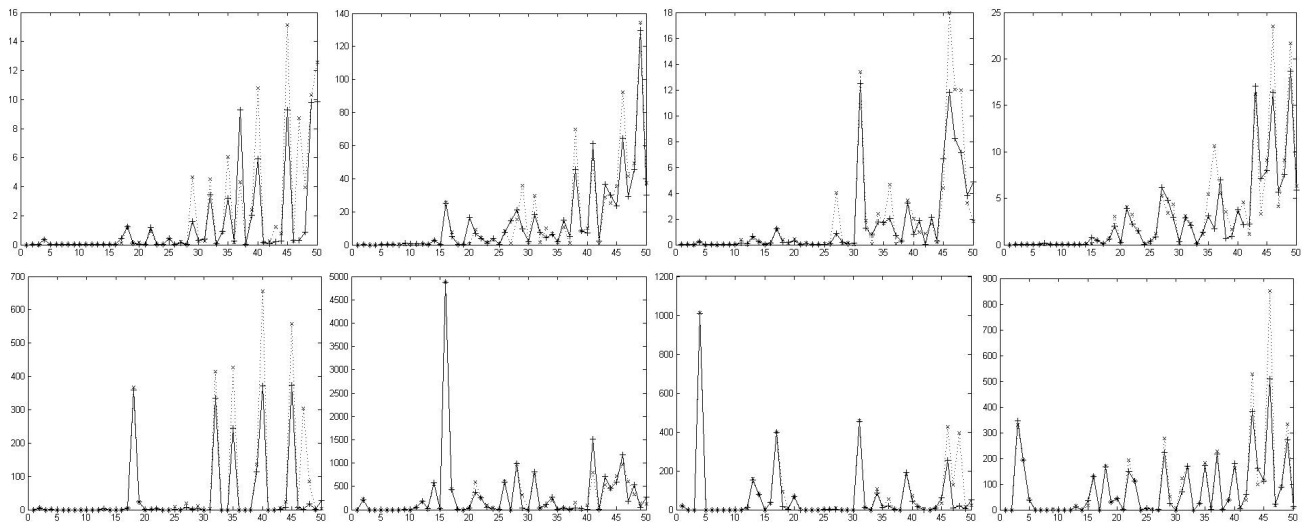


Fig. 4. From top to bottom: DN vs. width of block selected for inpainting (in pixels); ASVS vs. width of block selected for inpainting. From left to right: “A Tale of 1001 Nights,” “Mona Lisa,” *Wolf*, *Woman*. Metric scores for exemplar-based inpainting appear as solid lines, while metric scores for raster-order linear interpolation appear as dashed lines.

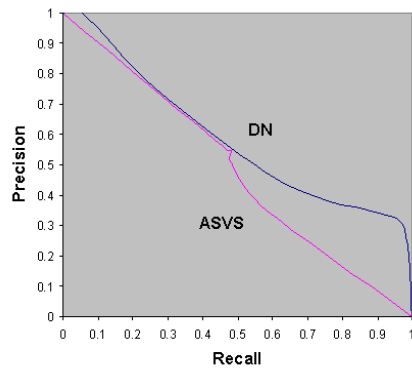


Fig. 5. Precision vs. Recall is shown, when attempting to classify inpainted results as “bad” (receiving at least one score of “4” during qualitative testing) vs. “good” (receiving only “1” to “3” scores). Ground truth is 31% “bad,” 69% “good.”

To further illustrate the use of DN and ASVS on large-region inpainting results, we tested the bungee jumper example of Fig. 6 (appearing as “Removing large objects from photographs” in the exemplar method’s journal publication¹ and originally appearing in an isophote continuity paper by Bertalmio *et al.*¹⁴). The exemplar-based approach produces a DN score of 2.247×10^2 and an ASVS score of 2.965×10^2 , while raster-order linear interpolation produces a DN score of 2.623×10^2 and an ASVS score of 4.085×10^2 , all easily appearing as metric peaks and breaking the aforementioned thresholds. Given the strongly salient central building (and its imprecise reconstruction), we consider it unsurprising that viewers gave both the exemplar-based and interpolation results a qualitative score of “4.”



Fig. 6. From left to right: original image (a), inpainting mask (b), exemplar-based result (c), linear interpolation result (d). Pixels selected for inpainting include the bungee jumper and attached rope (shown selected in (b)).

Inpainting of “scratches” produced metric scores that appeared to trend in a manner related to both those seen for pseudorandom pixels and squares: a general increasing trend in DN and often-elevated ASVS, with well-defined local peaks for both ASVS and DN (Fig. 7). This behavior appears to be the result of a decreasing number of exemplars (as with pseudorandom pixel selection) and medium-sized texture mismatches (as with square selection), with nearly all sampled results receiving a qualitative score of “3” or “4.” These findings appear consistent with the published examples (*e.g.*, “Comparison with ‘Texture and Structure Inpainting,’” “Comparison with Jia *et al.*,’ ‘Image Repairing’”),¹ including the growing likelihood and number of small visible artifacts with the increased number of thin lines removed (*e.g.*, “Image restoration example”).¹

Of interest is the minimal separation between interpolation and exemplar-based results for this scenario. While an exemplar-oriented approach may be expected to perform better due to its promises of linear structure continuity, the rapid breakdown associated with small-scale structural similarity (as well as a decreasing number of exemplars) leading to incorrect texture copying prevents its results from appearing visually superior to the known blurring artifacts of linear interpolation. We suggest that larger examples (*e.g.*, Fig. 8) indicate this situation more clearly than those previously published, and more strongly point out this important limitation of exemplar-based inpainting when considering the inpainting of a large number of thin objects; this includes compensating for occluding grates, removing large amounts of overlaid text, and “looking around” foreground objects when synthesizing stereo images from monoscopic data.

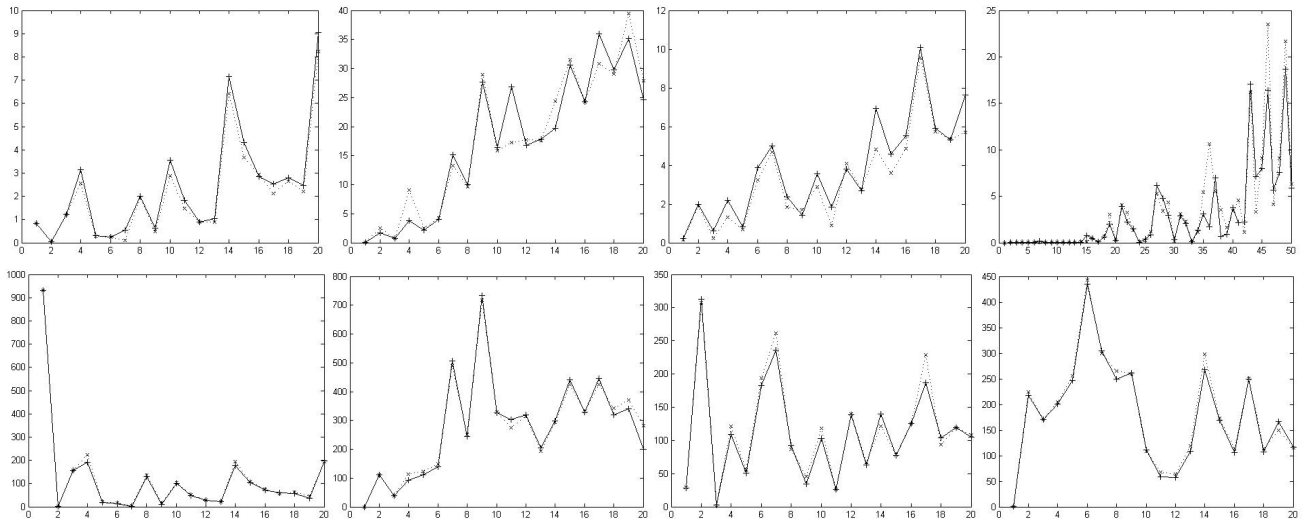


Fig. 7. From top to bottom: DN vs. number of “scratches” selected for inpainting; ASVS vs. number of “scratches” selected for inpainting. From left to right: “A Tale of 1001 Nights,” “Mona Lisa,” *Wolf, Woman*. Metric scores for exemplar-based inpainting appear as solid lines, while metric scores for raster-order linear interpolation appear as dashed lines.

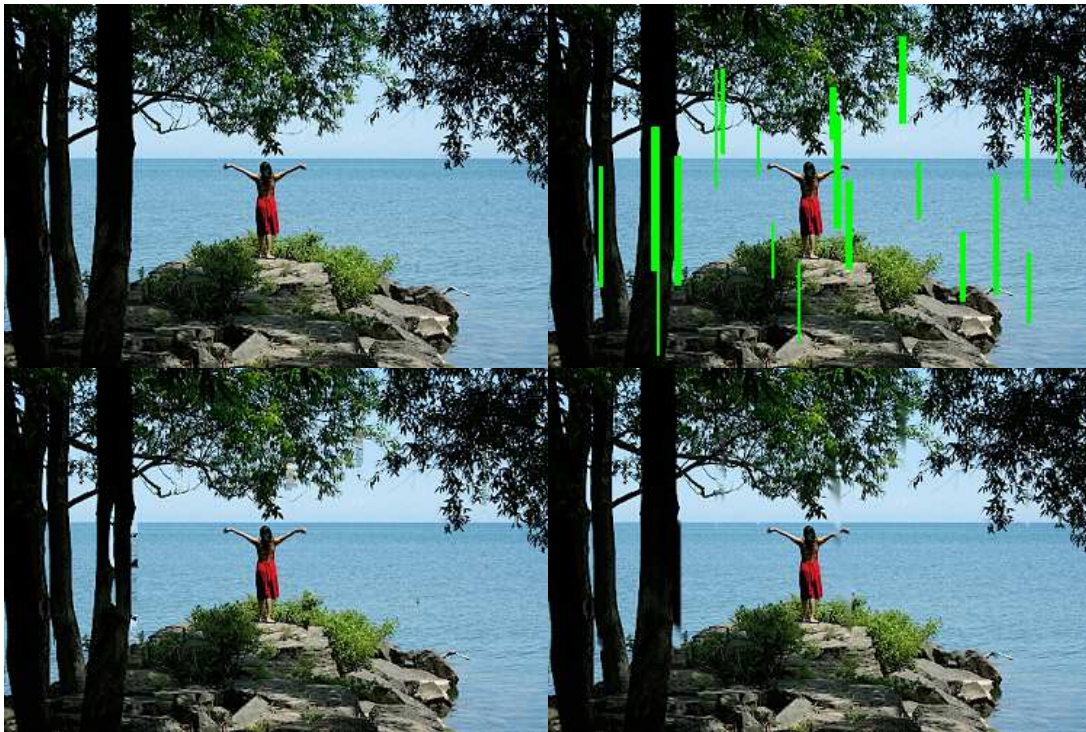


Fig. 8. From left to right, up to down: original image (a), inpainting mask (b), exemplar-based result (c), raster-order linear interpolation result (d). Although artifacting varies between (c) and (d), both images received a qualitative score of “4.” Pixels selected for inpainting are shown in bright green in (b).

CONCLUSIONS

The two proposed metrics correctly identify visually unacceptable inpainting results for a number of common uses, providing a first step toward quantifying inpainting success. For applications where original visual intent is not crucial, ASVS indicates the presence of highly visible inpainting artifacts. For applications where it is important to maintain original intended attention in addition to reducing visible artifacting, DN indicates when results have diverged from this goal.

Future work will include an expanded observer study (to reconfirm initial findings for a larger number of inpainted images and a wider observer base), metric evaluations of additional images (*e.g.*, motion picture frames) and scenarios (*e.g.*, ovoid removal), and derivative metrics for specific applications.

ACKNOWLEDGEMENTS

Special thanks to Sooraj Bhat for his assistance with the Matlab[®] implementation of the Criminisi *et al.* exemplar-based inpainting algorithm.

REFERENCES

- [1] Criminisi, A., Pérez, P. and Toyama, K., "Region filling and object removal by exemplar-based inpainting," *IEEE Trans. on Image Proc.* 13(9), 1200-1212 (2004).
- [2] Daly, S. J., "Visible differences predictor: An algorithm for the assessment of image fidelity," *Proc. SPIE* 1666(2) 179-206 (1992).
- [3] Osberger, W., Bergmann, N. and Maeder, A., "An automatic image quality assessment technique incorporating higher level perceptual factors," *Proc. ICIP* 3, 414-418 (1998).
- [4] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L. and Bovik, A. C., "Image quality assessment based on a degradation model," *IEEE Trans. on Image Proc.* 9(4), 636-650 (2000).
- [5] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Proc.* 13(4), 600-612 (2004).
- [6] Peters, R. J. and Itti, L., "Applying computational tools to predict gaze direction in interactive visual environments," *ACM Trans. on Applied Perception* 5(2), 1-21 (2008).
- [7] Peters, R. J. and Itti, L., "Congruence between model and human attention reveals unique signatures of critical visual events," *NIPS* 21, 1-8 (2008).
- [8] Itti, L., Koch, C. and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. PAMI* 20(11), 1254-1259 (1998).
- [9] Itti, L., Dhavale, N. and Pighin, F., "Realistic avatar eye and head animation using a neurobiological model of visual attention," *Proc. SPIE* 5200, 64-78 (2003).
- [10] Navalpakkam, V. and Itti, L., "Modeling the influence of task on attention," *Vision Research* 45(2), 205-231 (2005).
- [11] Zhaoping, L., "A saliency map in primary visual cortex," *TRENDS in Cog. Sci.* 6(1), 9-16 (2002).
- [12] Zhaoping, L., "Primary visual cortex creates a bottom-up saliency map," [*Neurobiology of Attention*], Elsevier, 570-575 (2005).
- [13] Criminisi, A., Pérez, P. and Toyama, K., "Object removal by exemplar-based inpainting," *Proc. CVPR* 2, 721-728 (2003).
- [14] Bertalmio, M., Sapiro, G., Caselles, V. and Ballester, C., "Image inpainting," *Proc. CGIT* 1, 417-424 (2000).
- [15] Bertalmio, M., Bertozzi, A. L. and Sapiro, G., "Navier-Stokes, fluid dynamics, and image and video inpainting," *Proc. CVPR* 1, 355-363 (2001).
- [16] Bertalmio, M., Vese, L., Sapiro, G. and Osher, S., "Simultaneous structure and texture image inpainting," *IEEE Trans. on Image Proc.* 12(8) 882-889 (2003).
- [17] Matsumoto, M. and Nishimura, T., "Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator," *ACM Trans. on Modeling and Comp. Simul.* 8(1) 3-30 (1998).