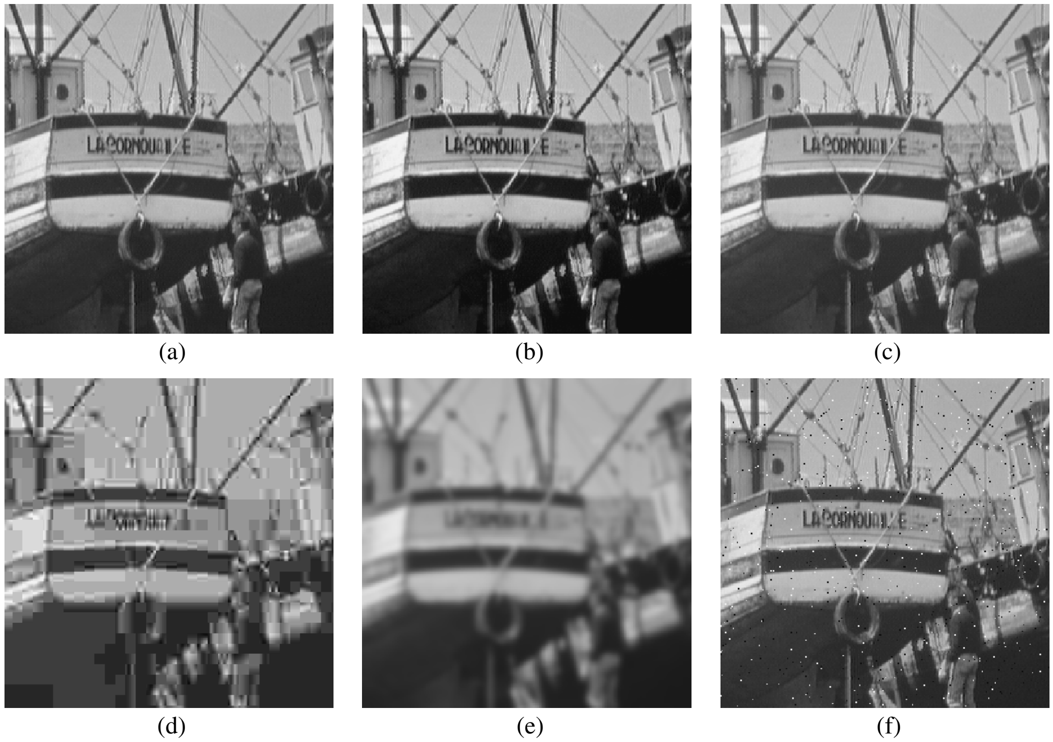

Fig. 2. Comparison of "Boat" images with different types of distortions, all with $\mathrm{MSE} = 210$. (a) Original image (8 bits/pixel; cropped from 512 × 512 to 256 × 256 for visibility). (b) Contrast-stretched image, $\mathrm{MSSIM} = 0.9168$. (c) Mean-shifted image, $\mathrm{MSSIM} = 0.9900$. (d) JPEG compressed image, $\mathrm{MSSIM} = 0.6949$. (e) Blurred image, $\mathrm{MSSIM} = 0.7052$. (f) Salt-pepper impulsive noise contaminated image, $\mathrm{MSSIM} = 0.7748$.

This new philosophy can be best understood through comparison with the error sensitivity philosophy. First, the error sensitivity approach estimates *perceived errors* to quantify image degradations, while the new philosophy considers image degradations as *perceived changes in structural information variation*. A motivating example is shown in Fig. 2, where the original "Boat" image is altered with different distortions, each adjusted to yield nearly identical MSE relative to the original image. Despite this, the images can be seen to have drastically different perceptual quality. With the error sensitivity philosophy, it is difficult to explain why the contrast-stretched image has very high quality in consideration of the fact that its visual difference from the reference image is easily discerned. But it is easily understood with the new philosophy since nearly all the structural information of the reference image is preserved, in the sense that the original information can be nearly fully recovered via a simple pointwise inverse linear luminance transform (except perhaps for the very bright and dark regions where saturation occurs). On the other hand, some structural information from the original image is permanently lost in the JPEG compressed and the blurred images, and therefore they should be given lower quality scores than the contrast-stretched and mean-shifted images.

Second, the error-sensitivity paradigm is a *bottom-up* approach, simulating the function of relevant early-stage components in the HVS. The new paradigm is a *top-down* approach, mimicking the hypothesized functionality of the overall HVS. This, on the one hand, avoids the suprathreshold problem mentioned in the previous section because it does not rely on threshold psychophysics to quantify the perceived distortions. On the other hand, the cognitive interaction problem is also reduced to a certain extent because probing the structures of the objects being observed is thought of as the purpose of the entire process of visual observation, including high level and interactive processes.

Third, the problems of natural image complexity and decorrelation are also avoided to some extent because the new philosophy does not attempt to predict image quality by accumulating the errors associated with psychophysically understood simple patterns. Instead, the new philosophy proposes to evaluate the structural changes between two complex-structured signals directly.

B. The SSIM Index

We construct a specific example of a SSIM quality measure from the perspective of image formation. A previous instantiation of this approach was made in [6]–[8] and promising results on simple tests were achieved. In this paper, we generalize this algorithm, and provide a more extensive set of validation results.

The luminance of the surface of an object being observed is the product of the illumination and the reflectance, but the structures of the objects in the scene are independent of the illumination. Consequently, to explore the structural information in an image, we wish to separate the influence of the illumination. We define the structural information in an image as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. Since luminance