Complex Wavelet Structural Similarity: A New Image Similarity Index

Mehul P. Sampat, *Member, IEEE*, Zhou Wang, *Member, IEEE*, Shalini Gupta, Alan Conrad Bovik, *Fellow, IEEE*, and Mia K. Markey, *Senior Member, IEEE*

***Abstract*—We introduce a new measure of image similarity called the complex wavelet structural similarity (CW-SSIM) index and show its applicability as a general purpose image similarity index. The key idea behind CW-SSIM is that certain image distortions lead to consistent phase changes in the local wavelet coefficients, and that a consistent phase shift of the coefficients does not change the structural content of the image. By conducting four case studies, we have demonstrated the superiority of the CW-SSIM index against other indices (e.g., Dice, Hausdorff distance) commonly used for assessing the similarity of a given pair of images. In addition, we show that the CW-SSIM index has a number of advantages. It is robust to small rotations and translations. It provides useful comparisons even without a preprocessing image registration step, which is essential for other indices. Moreover, it is computationally less expensive.**

***Index Terms*—Complex wavelet structural similarity index (CW-SSIM), image similarity, structural similarity (SSIM) index.**

I. Introduction

IMAGE similarity indices play a crucial role in the development and evaluation of a large number of image processing and pattern recognition algorithms. Examples include image coding, restoration, denoising, halftoning, segmentation, communication, target detection, image registration, and object recognition. An image can be viewed as a 2-D function $x(i,j)$ of intensity. A similarity index for images should account for both intensity variations and geometric distortions. Image similarity indices can be roughly classified according to how they handle these two properties. One class assumes that the images being compared are at the same scale and are perfectly registered, and, thus, their similarity is determined from a comparison of the corresponding pixel intensities. We refer to such indices as "intensity-based." Alternatively, pixel correspondences may first be established between the images based on intensity, then similarity determined by comparing the geometric transformations between corresponding pixels. Such indices will be referred to as "geometry-based."

The simplest intensity-based indices are designed to compare two binary images. For example, a typical application of such indices is to compare the segmentation of a dense object (binary output) to a surrogate of ground truth. These indices quantify "spatial overlap" by applying Boolean operations to the corresponding pixel intensities. The Boolean operations make them straightforward and easy to implement. Examples include the Dice [1], Jaccard [2], Simpson [3], Ochiai [4], Braun-Blanquet [5], and Sokal-Sneeth [6] similarity indices. As a consequence of their design, these indices penalize images for being different by even just one pixel. While this property is probably desirable for some applications, such as where the goal is to compare segmentations of dense objects encompassing many pixels, it is of questionable value for other applications, such as assessing the similarity of tracings of linear structures comprised of relatively few pixels. Another commonly used intensity-based index is the mean square error (MSE), that can be applied to both bi-valued and grayscale images [7]. The grayscale images may be intensity images, range images, or maps of some other scene property. While other intensity-based indices report agreement between two images, the MSE is a measure of disagreement. The MSE is lower-bounded at zero (when the two images are identical) and has no upper bound. By comparison, many intensity-based indices are designed to have an upper-bound of unity and a lower-bound of zero.

All the intensity-based indices described above are point operations. In other words, the similarity evaluation at one pixel is independent of all other pixels in the image. However, neighboring image pixels are highly correlated with each other. To take advantage of such correlations, and also to take into account properties of the human visual system (HVS), a variety of transform-domain methods have been proposed [7]–[11]. Recently, a substantially different index called the structural similarity (SSIM) index [12] was proposed that also accounts for spatial correlations. In SSIM, the structural information of an image is defined as those attributes that represent the structures of the objects in the visual scene, apart from the mean intensity and contrast. Thus, the SSIM index compares local patterns of pixel intensities that have been normalized for mean intensity and contrast. Compared with HVS-based methods, SSIM has

Manuscript received May 09, 2008; revised December 20, 2008. First published June 23, 2009; current version published October 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun Abraham Ross.

M. P. Sampat is with the Advanced Imaging in Multiple Sclerosis Laboratory, Department of Neurology, University of California San Francisco, San Francisco, CA 94107 USA (e-mail: mehul.sampat@ieee.org).

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, ON N2L 3G1 Canada (e-mail: zhouwang@ieee.org).

S. Gupta is with the Laboratory for Image and Video Engineering and the Biomedical Informatics Lab, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: shalinig@ece.utexas.edu).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

M. K. Markey is with the Biomedical Informatics Lab, Department of Biomedical Engineering, University of Texas Austin, TX 78712 USA (e-mail: mia.markey@mail.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2025923

lower computational complexity, and superior performance in image quality assessment tasks [12].

A common drawback of existing intensity-based indices is their high sensitivity to geometric and scale distortions. This becomes a big problem when there are small translations, rotations, or scale differences between the images being compared. For example, when comparing 2-D binary edge images it is likely that outlines representing the same objects will have the same general shape, but when generated by different users and/or algorithms will have displaced coordinates. Intensity-based indices would substantially penalize such distortions, even if they are visually negligible. Similarly, in higher dimensions, we may be interested in comparing the geometric structure of free-form curved surfaces in space. Thin surfaces in 3-D space computed as boundaries/edges of 3-D objects may be displaced relative to each other.

Indeed, comparing edge images is the primary motivating application for geometry-based similarity indices. These include the Figure of Merit (FOM) [13], Closest Distance Metric (CDM) [14], [15], Pixel Correspondence Metric (PCM) [15], the Partial Hausdorff Distance Metric (PHDM) [16], and a modified version of MSE called the point-to-closest-point mean squared error ($\mathrm{MSE_{CP}}$) [17]. All of these indices allow for small localization errors (user-specified) between the structures being compared. Most of these indices operate in the spatial domain. In essence, each pixel in one image is "matched" with a corresponding pixel in the other image. Thus, for geometry-based methods, correspondences between pairs of pixels in the two images is not assumed, but is established before the index is computed. This process can be computationally expensive. Some indices (e.g., FOM, PHDM, and $\mathrm{MSE_{CP}}$) allow more than one pixel in an image to correspond to the same pixel in the other image (many-to-one mapping). This is sometimes undesirable, and, thus, CDM and PCM were designed so that there is a one-to-one mapping between pixels of the two images. It was reported that PCM produces a more accurate matching than CDM [15]. While most geometry-based indices account only for edge distances in 2-D images, a few indices (e.g., PCM) also account for the intensity of edges. This has advantages for gray-scale applications.

We propose a novel complex wavelet SSIM index (CW-SSIM, pronounced "C-W-S-Seem") as a general index for image similarity measurement.[1] The key idea behind CW-SSIM is that small geometric image distortions lead to consistent phase changes in the local wavelet coefficients, and that a consistent phase shift of the coefficients does not change the structural content of the image. The advantages of the CW-SSIM approach are manifold. In some sense, it addresses the drawbacks of both intensity-based and geometry-based methods. First, CW-SSIM does not require explicit correspondences between pixels being compared. Second, CW-SSIM is insensitive to small geometric distortions (small rotations, translations and small differences in scale). Third, CW-SSIM compares the textural and structural properties of localized regions of image pairs. This feature is lost in geometric-based methods. We empirically demonstrate that CW-SSIM generalizes well to a wide variety of image and pattern similarity tasks, using four disparate case studies. These include (a) comparison of 2-D binary edge images with simulated geometric distortions, (b) comparison of human segmentations of arbitrary shaped natural objects, (c) comparison of human segmentations of linear structures in medical images, and (d) accurate recognition of 3-D human facial surfaces using face range images. We further demonstrate the performance of CW-SSIM relative to the established indices that are currently employed for each of these tasks.

[1]Preliminary results were published in Wang & Simoncelli, "Translation insensitive image similarity in complex wavelet domain," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2005; Sampat *et al.*, "Measuring intra- and interobserver agreement in identifying and localizing structures in medical images", IEEE Int. Conf. Image Processing, 2006; and Gupta *et al.*, "Facial range image matching using the complex-wavelet structural similarity metric," IEEE Workshop on Applications of Computer Vision, 2007.

TABLE I
TWO-BY-TWO CONTINGENCY TABLE: a IS THE NUMBER OF CORRESPONDING PIXELS OF VALUE 1 IN BOTH IMAGES; b IS THE NUMBER OF PIXELS TAKING VALUE 1 IN IMAGE # 1 ONLY; c IS THE NUMBER OF PIXELS TAKING VALUE 1 IN IMAGE # 2 ONLY; d IS THE NUMBER OF CORRESPONDING PIXELS THAT ARE 0 IN BOTH IMAGES

		Image 2		
		One	Zero	Total
Image 1	One	a	b	a+b
	Zero	c	d	c+d
	Total	a+c	b+d	a+b+c+d

II. EXISTING SIMILARITY INDICES

A. Intensity-Based Similarity Indices

The simplest similarity indices for binary images are based on direct measurement of the degree of spatial overlap between these images. Let a represent the number of corresponding pixels of value "1" in both binary images. The greater the spatial overlap between objects in the two images, the greater the value of a. Let c be the number of pixels taking value "1" in image # 1 only and b be the number of pixels taking value "1" in image 2 only. Finally, d is the number of corresponding pixels that are "0" in both images. The terms a, b, c, and d form a two-by-two contingency table as shown in Table I. Indices designed to measure the spatial overlap between objects in two binary images are computed by dividing the term a by a normalizing factor, so that the resulting values fall in the range 0 to 1. The normalizing factor is formed by a combination of the terms a, b, c, and d.

A comprehensive review of intensity-based overlap indices was published by Shi [18]. Competitive spatial-overlap-based indices are summarized in Tables II and III. These can be divided into two categories. The indices in Table II do not use the term d whereas the indices in Table III do. For evaluating similarity of binary images, the measures in Table II are more intuitively appealing than those in Table III, since most users probably prefer that a similarity measure for images be independent of the size of the background. Of course, the indices in Table III may be suitable for other applications where both agreement variables (a and d) are equally informative. The Dice similarity coefficient [1] is commonly used (e.g., in medical imaging studies) to quantify the degree of overlap between objects in two binary images (e.g., two segmentations, [19], [20]). Dice is defined as

$$\mathrm{Dice} = (2 \times a)/(2 \times a + b + c). \tag{1}$$

TABLE II
SIMILARITY INDICES FOR BINARY DATA: THESE USE THE TERMS a, b, AND c ONLY (FROM TABLE I)

Name of Index, Year, Source	Formula	Range
Dice, 1945 [1]	$\frac{2 \cdot a}{2 \cdot a + b + c}$	0 to 1
Jaccard[2], 1912 [2]	$\frac{a}{a + b + c}$	0 to 1
Kulczynski1, 1928 [21]	$\frac{a}{b + c}$	0 to inf
Kulczynski2, 1928 [21]	$\frac{(a/2) \cdot (2 \cdot a + b + c)}{(a + b) \cdot (a + c)}$	0 to 1
Simpson, 1960 [3]	$\frac{a}{min(a + b, a + c)}$	0 to 1
Ochiai[3], 1957 [4]	$\frac{a}{((a + b) \cdot (a + c))^{1/2}}$	0 to 1
McConnaughey, 1982 [22]	$\frac{a^2 - b \cdot c}{(a + b) \cdot (a + c)}$	-1 to 1
Braun-Blanquet, 1932 [5]	$\frac{a}{max(a + b, a + c)}$	0 to 1
Sokal & Sneath(2), 1963 [6]	$\frac{a}{a + 2 \cdot b + 2 \cdot c}$	0 to 1

TABLE III
INDICES FOR BINARY DATA: THESE USE THE TERMS a, b, c, AND d (FROM TABLE I)

Name of Index, Year, Source	Formula	Range
Russell & Rao, 1940 [23]	$\frac{a}{a + b + c + d}$	0 to 1
Simple Matching, 1958 [24]	$\frac{a + d}{a + b + c + d}$	0 to 1
Yule,1900 [25]	$\frac{(a \cdot d) - (b \cdot c)}{(a \cdot d) + (b \cdot c)}$	0 to 1
Rogers & Tanimoto, 1960 [26]	$\frac{a + d}{a + d + 2 \cdot (b + c)}$	0 to 1
Sokal and Sneath(1), 1963 [6]	$\frac{2 \cdot (a + d)}{2 \cdot (a + d) + b + c}$	0 to 1

If objects in the two binary images overlap perfectly then $\mathrm{Dice} = 1$, and if there is no overlap then $\mathrm{Dice} = 0$. Although it has been argued that a Dice value >0.7 should be taken to indicate "excellent" agreement [20], it is difficult to analyze the absolute value of Dice since "similarity" is application dependent and can be interpreted differently in different contexts.

Another index commonly used is the Jaccard index [2]. The Jaccard and Dice indices are closely related: $\mathrm{Dice} = (2 \times \mathrm{Jaccard})/(\mathrm{Jaccard} + 1)$. Thus, the Jaccard index does not provide independent information relative to Dice, so we do not compute both indices in this study. Since all the indices listed in Table II are related, some are equivalent under certain circumstances. For example, if $(b = c)$ then the Dice, Kulczynski2, Braun-Blanquet, Simpson, and Ochiai indices are equivalent and simplify to the formula $a/(a + b)$. This corresponds to the case where the total number of pixels labeled as "1" in each of the two binary images is equal.

1) Mean Squared Error: The Mean Squared Error (MSE) is another commonly used intensity-based index [7] that can be applied to both binary and grayscale images. The MSE is a popular index as it is intuitive, and computationally and analytically tractable. For two images x and y, of size $N \times M$, $\mathrm{MSE} = (1/NM) \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) - y(i,j))^2$. For binary images, the MSE is a linear function of the Simple Matching index [24] (Table III).

[2]Dice = (2∗Jaccard)/(Jaccard + 1).

[3]Also called Otuska or Fager index.

B. Geometry-Based Similarity Indices

Geometry-based similarity indices compare the overall shapes of objects regardless of the presence of small geometric distortions between them. They can be employed, e.g., to compare the shapes of the outlines/edges of objects in 2-D images, or to compare the shapes of surfaces in three dimensions. Geometry-based indices include the Figure of Merit (FOM, [13]), Closest Distance Metric (CDM, [14]), Pixel Correspondence Metric (PCM, [15]), the Partial Hausdorff Distance Metric (PHDM, [16]), and a modified version of the MSE called the point-to-closest-point mean squared error ($\mathrm{MSE_{CP}}$, [17]).

1) Pixel Correspondence Metric (PCM): The PCM was proposed by Prieto *et al.* [15]. Let x and y represent two images that are to be compared. Let $x(i,j)$ represent a particular pixel in x and $y(k,l)$ represent the corresponding pixel in y. For every pixel $x(i,j)$, PCM tries to find its match within a neighborhood of radius r of the corresponding pixel $y(k,l)$. The PCM index is computed as

$$\mathrm{PCM}_r(x,y) = 100 \cdot \left(1 - \frac{C(M_{\mathrm{opt}}(x,y))}{|x \cup y|}\right) \tag{2}$$

where $C(M_{\mathrm{opt}}(x,y))$ is the cost of optimal matching between the images, r denotes the localization error allowed between the pixels and $|x \cup y|$ is the number of nonzero pixels in x or y. (Note that in in terms of the quantities described in Table I, $|x \cup y| = a + b + c$). The search for the optimal match is solved by modeling the task as a weighted matching task in bipartite graphs [15]. In this work, the search radius was $r = 2$.

2) Closest Distance Metric (CDM): The CDM was proposed by Bowyer *et al.* [14]. The goal is to find a possible match in y (within a certain neighborhood), for every pixel $x(i,j)$ in image x. If multiple matches exist, then the edge pixel $y(k,l)$ that is closest to $x(i,j)$ is selected. Finally, the number of matched and unmatched pixels are counted. The index is computed as

$$\mathrm{CDM}_r(x,y) = 100 \cdot \left(1 - \frac{C(M_{cd}(x,y))}{|x \cup y|}\right) \tag{3}$$

where $\mathrm{CDM}_r(x,y)$ is the cost of the matching obtained using the closest-distance condition, r is the neighborhood radius used in the matching.

3) Point-to-Closest-Point Mean Squared Error ($\mathrm{MSE_{CP}}$)*:* The $\mathrm{MSE_{CP}}$ is a modified version of the MSE which has been employed for applications such as free-form object matching and registration [17]. In order to calculate this index, the edge points in each image are represented as a cloud of points. For example, edge pixels in 2-D images are represented a collection of 2-D (i,j) locations and a 3-D surface is represented as a cloud of 3-D (i,j,k) points. For two point clouds $\chi = \{a_1, a_2, a_3 \ldots a_M\}$ and $\psi = \{b_1, b_2, b_3 \ldots b_N\}$, containing M and N-dimensional points, respectively, first compute the directed $\mathrm{MSE_{CP}}$ from set χ to set ψ as: $\mathrm{MSE_{CP}}(\chi,\psi) = 1/M \sum_{i=1}^{M} D(a_i, b_{i'})$, where D is the square of the Euclidean distance between point $a_i \in \chi$ and the point closest to it $b_{i'} \in \psi$. The directed $\mathrm{MSE_{CP}}(\psi,\chi)$ is similarly computed and the undirected $\mathrm{MSE_{CP}}$ between the two surfaces is defined as

$$\mathrm{MSE_{CP}} = \max(\mathrm{MSE_{CP}}(\chi,\psi), \mathrm{MSE_{CP}}(\psi,\chi)). \tag{4}$$

4) Partial Hausdorff Distance Metric (PHDM): The partial Hausdorff distance metric is a measure of similarity between two point sets [16]. The procedure to calculate PHDM is similar to that of the $\mathrm{MSE_{CP}}$ index. For calculating the PHDM from point set χ to ψ, the distances of all points $a_i \in \chi$ to their closest points in ψ are sorted in ascending order, and the P^{th} distance in the ordered set quantifies the directed partial Hausdorff distance metric, i.e., $h_P(\chi,\psi) = P^{\mathrm{th}}_{a\in\chi} D(a_i, b_{i'})$. We can similarly calculate $h_Q(\psi,\chi)$ the directed partial Hausdorff distance metric from point set ψ to point set χ. The undirected partial Hausdorff distance between the two point sets is then defined as

$$\mathrm{PHDM} = \max(h_P(\chi,\psi), h_Q(\psi,\chi)). \tag{5}$$

The partial Hausdorff distance has the advantage of being robust to outliers produced by noise and occlusions of objects. However, its performance depends on the selection of optimal values for the heuristic parameters P and Q that quantify the extent of overlap between the two point sets.

III. Structural Similarity Index (SSIM)

The SSIM index was originally proposed to predict human preferences in evaluating image quality [12], [27]. The fundamental principle of the structural approach is that the HVS is highly adapted to extract structural information (the structures of the objects) from the visual scene, and, therefore, a measurement of structural similarity (or distortion) should provide a good approximation of perceptual image quality. In particular, SSIM attempts to discount those distortions that do not affect the structures (local intensity patterns) of the image. The SSIM algorithm provides surprisingly good image quality prediction performance for a wide variety of image distortions [12], [27]. In the spatial domain, the SSIM index between two image patches $\mathbf{x} = \{x_i \,|\, i = 1, \ldots M\}$ and $\mathbf{y} = \{y_i \,|\, i = 1, \ldots, M\}$ is defined as

$$S(\mathbf{x},\mathbf{y}) = \frac{(2\,\mu_x\,\mu_y + C_1)\,(2\,\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

where C_1 and C_2 are two small positive constants, and $\mu_x = 1/M\sum_{i=1}^{M} x_i$, $\sigma_x^2 = 1/M\sum_{i=1}^{M}(x_i - \mu_x)^2$, $\sigma_{xy} = 1/M\sum_{i=1}^{M}(x_i - \mu_x)(y_i - \mu_y)$. The maximum SSIM index value 1 is achieved if and only if $\mathbf{x}$ and $\mathbf{y}$ are identical [12], [27].

IV. Complex Wavelet Structural Similarity Index

The CW-SSIM index is an extension of the SSIM method to the complex wavelet domain. The major drawback of the spatial domain SSIM algorithm is that it is highly sensitive to translation, scaling, and rotation of images. This was our original motivation to extend the SSIM idea. The goal is to design a measurement that is insensitive to "nonstructured" geometric image distortions that are typically caused by nuisance factors, such as, relative movement of the image acquisition device, rather than actual change in the structure of the objects. The CW-SSIM index is also inspired by the impressive pattern recognition capabilities of the HVS. In the last three decades, scientists have found that neurons in the primary visual cortex are well-modeled localized multiscale bandpass oriented filters (loosely referred to as "wavelets") that decompose images into multiple visual *channels* [28]–[30]. Interestingly, some psychophysical evidence suggests that the same set of visual channels may also be used in image pattern recognition tasks [31]. Furthermore, phase contains more structural information than magnitude in typical natural images [32], and rigid translation of image structures leads to consistent phase shift. In previous work, consistency of global (Fourier) and local (wavelet) phase across scale and space has been used to characterize image features [33]–[35]. Wavelet phase has also been employed in a number of machine vision and image processing applications [36]–[38]. Based on these observations, we attempt to design a complex wavelet domain index that 1) separates the measurement of magnitude and phase distortions; 2) is more sensitive to phase than magnitude distortions; and 3) is insensitive to consistent *relative* phase distortions.

We consider symmetric complex wavelets whose "mother wavelets" can be written as a modulation of a low-pass filter $w(u) = g(u)e^{j\omega_c u}$, where ω_c is the center frequency of the modulated band-pass filter, and $g(u)$ is a slowly varying and symmetric function. The family of wavelets are dilated/contracted and translated versions of the mother wavelet

$$w_{s,p}(u) = \frac{1}{\sqrt{s}} w\left(\frac{u-p}{s}\right) = \frac{1}{\sqrt{s}} g\left(\frac{u-p}{s}\right) e^{j\omega_c(u-p)/s} \tag{7}$$

where $s \in R^+$ is the scale factor, and $p \in R$ is the translation factor. The continuous wavelet transform of a real signal $x(u)$ is

$$X(s,p) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\sqrt{s}\,G(s\,\omega - \omega_c)e^{j\omega p}d\omega \tag{8}$$

where $X(\omega)$ and $G(\omega)$ are the Fourier transforms of $x(u)$ and $g(u)$, respectively. The discrete wavelet coefficients are sampled versions of the continuous wavelet transform. In the complex wavelet transform domain, suppose $\mathbf{c}_x = \{c_{x,i} \,|\, i = 1, \ldots, N\}$ and $\mathbf{c}_y = \{c_{y,i} \,|\, i = 1, \ldots, N\}$ are two sets of coefficients extracted at the same spatial location in the same wavelet subbands of the two images being compared, respectively. The CW-SSIM index is defined as

$$\tilde{S}(\mathbf{c}_x,\mathbf{c}_y) = \frac{2\left|\sum_{i=1}^{N} c_{x,i}c_{y,i}^*\right| + K}{\sum_{i=1}^{N}|c_{x,i}|^2 + \sum_{i=1}^{N}|c_{y,i}|^2 + K}. \tag{9}$$

Here c^* denotes the complex conjugate of c and K is a small positive constant. The purpose of the small constant K is mainly to improve the robustness of the CW-SSIM measure when the local signal to noise ratios are low.

To better understand the CW-SSIM index, rewrite it as a product of two components

$$\tilde{S}(\mathbf{c}_x,\mathbf{c}_y) = \frac{2\sum_{i=1}^{N}|c_{x,i}||c_{y,i}| + K}{\sum_{i=1}^{N}|c_{x,i}|^2 + \sum_{i=1}^{N}|c_{y,i}|^2 + K} \cdot \frac{2\left|\sum_{i=1}^{N} c_{x,i}c_{y,i}^*\right| + K}{2\sum_{i=1}^{N}|c_{x,i}c_{y,i}^*| + K}. \tag{10}$$

The first component is completely determined by the magnitudes of the coefficients and the maximum value 1 is achieved if and only if $|c_{x,i}| = |c_{y,i}|$ for all i. It is equivalent to the SSIM

index (6) applied to the magnitudes of the coefficients (notice that the coefficients are zero mean, due to the bandpass nature of the wavelet filters). The second component, is fully determined by the consistency of phase changes between $\mathbf{c}_x$ and $\mathbf{c}_y$. It achieves the maximum value 1 when the phase difference between $c_{x,i}$ and $c_{y,i}$ is a constant for all i. We consider this component as a useful measure of image structural similarity based on the beliefs that 1) the structural information of local image features is mainly contained in the relative phase patterns of the wavelet coefficients and 2) constant phase shift of all coefficients does not change the structure of the local image feature. In previous work, a similar phase correlation idea had been employed for feature localization [33], [34], texture description [37], and blur detection [35], but, to the best of our knowledge, has not been used for image similarity measurement. We now analyze the sensitivity of CW-SSIM with respect to different types of image distortions. In all the analysis below, we assume that $\mathbf{x}$ corresponds to a reference image and $\mathbf{y}$ is an altered version of the image whose similarity to the reference image is being evaluated.

Luminance and contrast changes can be roughly described as a point-wise linear transform of local pixel intensities: $y_i = ax_i + b$ for all i. Due to the linear and bandpass nature of the wavelet transform, the effect in the wavelet domain is a constant scaling of all the coefficients, i.e., $c_{y,i} = ac_{x,i}$ for all i. Substituting into (10), value 1 is obtained for the second component and the first component is

$$\tilde{S}(\mathbf{c}_x, \mathbf{c}_y) = \frac{2a + K/\sum_{i=1}^{N} |c_{x,i}|^2}{1 + a^2 + K/\sum_{i=1}^{N} |c_{x,i}|^2} > \frac{2a}{1+a^2}. \quad (11)$$

The inequality follows since $2a \leq 1 + a^2$ and $K/\sum_{i=1}^{N} |c_{x,i}|^2 > 0$. This is an insensitive measure (compared with many spatial domain methods such as the MSE)-scaling the magnitude by a factor of 10% ($a = 1.1$) only causes reduction of the CW-SSIM value from 1 to a value >0.996. The effect is more manifest at weaker image features (which create small coefficient magnitudes) than stronger image features (large coefficient magnitudes).

Translation in the 2-D spatial domain can be written as: $y\binom{u_1}{u_2} = x\binom{u_1+\Delta u_1}{u_2+\Delta u_2}$ where Δu_1 and Δu_2 represent the horizontal and vertical shifts, respectively. For easy analysis, first consider the 1-D case $y(u) = x(u + \Delta u)$. This corresponds to a linear phase shift in the Fourier domain $Y(\omega) = X(\omega)e^{j\omega\Delta u}$. Substituting this into (8) yields

$$\begin{aligned} Y(s,p) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\sqrt{s}\,G(s\omega - \omega_c) \\ &\quad \times e^{j\omega(p+\Delta u)} d\omega \\ &= e^{j\omega_c \Delta u/s} \cdot \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\sqrt{s}\,G(s\,\omega - \omega_c) \\ &\quad \times e^{j\omega p} e^{j(\omega-\omega_c/s)\Delta u} d\omega \\ &\approx X(s,p)\, e^{j\omega_c \Delta u/s}. \end{aligned} \quad (12)$$

Here the approximation is valid when the translation Δu is small (compared with the size of the wavelet filter) and the envelope $g(u)$ of the wavelet filter is slowly varying. (The approximation becomes exact when $g(u) \equiv 1$, i.e., $G(\omega) = \delta(\omega)$.) Similar result are obtained for the 2-D case. Consequently, the discrete wavelet coefficients $\{c_{y,i}\}$ and $\{c_{x,i}\}$ (discrete samples of $X(s,p)$ and $Y(s,p)$) at the same location in the same wavelet subband are approximately phase shifted versions of each other. Therefore, we can write $c_{y,i} \approx c_{x,i}e^{j\phi}$ for all i, and thus

$$\tilde{S}(\mathbf{c}_x, \mathbf{c}_y) \approx \frac{2\left|\sum_{i=1}^{N} c_{x,i}\, c_{x,i} e^{-j\phi}\right| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{x,i}e^{j\phi}|^2 + K} = 1. \quad (13)$$

Here, the accuracy of the approximation depends on the magnitudes of the translation as well as the shape of the envelope of the wavelet filter.

Scaling and rotation in the 2-D spatial domain can be written as

$$y\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = x\left(\begin{pmatrix} 1+\Delta s_1 & 0 \\ 0 & 1+\Delta s_2 \end{pmatrix} \times \begin{pmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\right) \quad (14)$$

where $(1+\Delta s_1, 1+\Delta s_2)$ and $\Delta\theta$ are the scaling and rotation factors, respectively. When $\Delta\theta$ is small, we have $\cos\Delta\theta \approx 1$ and $\sin\Delta\theta \approx \Delta\theta$, and therefore

$$\begin{aligned} y\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &\approx x\begin{pmatrix} u_1 + (u_1\Delta s_1 - u_2\Delta\theta - u_2\Delta s_1\Delta\theta) \\ u_2 + (u_2\Delta s_2 + u_1\Delta\theta + u_1\Delta s_2\Delta\theta) \end{pmatrix} \\ &= x\begin{pmatrix} u_1 + \Delta u_1 \\ u_2 + \Delta u_2 \end{pmatrix} \end{aligned} \quad (15)$$

where $\Delta u_1 = u_1\Delta s_1 - u_2\Delta\theta - u_2\Delta s_1\Delta\theta$ and $\Delta u_2 = u_2\Delta s_2 + u_1\Delta\theta + u_1\Delta s_2\Delta\theta$, respectively. From (15), we see that when (u_1, u_2) is not far away from the origin, a small amount of scaling and rotation can be locally approximated by a small translation $(\Delta u_1, \Delta u_2)$. Based on the same analysis as in the small translation case described above, $\tilde{S}(\mathbf{c}_x, \mathbf{c}_y) \approx 1$.

To summarize the sensitivity analysis given above, we conclude that the CW-SSIM index is simultaneously insensitive to luminance change, contrast change, and geometric translation, scaling and rotation. In other words, the CW-SSIM index is resilient to small distortions only and provides smaller similarity index value for large distortions. From the viewpoint of the structural similarity philosophy explained earlier, all of these distortions belong to the category of nonstructural distortions, e.g., caused by localization errors or by changes of lighting conditions or by the movement of the image acquisition systems. These are not due to changes in the structures of the objects in the scene. On the other hand, CW-SSIM is sensitive to structural distortions (e.g., JPEG compression) because they lead to significant variations of the local relative phase patterns. It is important to note that the *insensitive*, rather than *invariant*, property to nonstructural distortions is exactly what we want, because it is often desirable to penalize these distortions when they are extreme.

The CW-SSIM index has some interesting connections with computational models that have been used to account for visual behavior. These models include: 1) the involvement of bandpass

visual channels in image pattern recognition tasks [31]; 2) representation of phase information in primary visual cortex using quadrature pairs of localized bandpass filters [39]; 3) the computation of complex-valued products in visual cortex [40]; 4) the computation of local energy (using sums of squared responses of quadrature-pair filters) by complex cells in visual cortex [41]; and 5) divisive normalization of filter responses (using summed energy of neighboring filter responses) in both visual and auditory neurons [42], [43]. All of these suggest that the computation involved in the CW-SSIM measure is biologically plausible.

To implement the CW-SSIM index for the comparison of images, we first decompose the images using a complex version [37] of a 6-scale, 16-orientation steerable pyramid decomposition [44], which is a type of redundant wavelet transform that avoids aliasing in subbands. The CW-SSIM indices are then computed locally using a sliding 7×7 window that moves across each wavelet subband. For parameter K (9), in our experiments we found that small perturbations do not result in noticeable performance degradations and use $K = 0$ for all the results reported in this paper. More careful tuning of this parameter is needed in future studies especially in the presence of noise. Finally, the resulting CW-SSIM index map is combined into a scalar similarity measure using a weighted summation. The weighting function is obtained using a Gaussian profile with a standard deviation equaling a quarter of the image size at finest level of pyramid decomposition.

V. Case Study I: Simulated Images

A. Introduction

The goal of this case study was to study the effect of simple transformations on similarity indices. For this task, a simulated image was created [Fig. 1(a), solid lines] and variants of this image were created by rotation, translation, and adjusting the edge thickness. The simulated image was compared to each of the rotated and translated images and similarity indices were computed. In this case study, the CW-SSIM index was compared against an intensity-based index (Dice) and a geometry-based index (PCM). We only selected one intensity-based index, since the intensity-based indices (Table II) are all closely related, and several are identical in the simulations for this case study since $b = c$.

B. Methods

Three tests were conducted in this study. In the first test, the test image was rotated from 0 to 4 degrees in increments of 0.4 degrees and was compared to each of the rotated images. Translation was not applied to these images and the edge width was fixed at one pixel. In the second test, the simulated image was translated by shifting in the horizontal or vertical directions by a few pixels. Let Δu_1 and Δu_2 denote the amount of spatial translation applied in the horizontal and vertical directions. The Euclidean distance is $\sqrt{(\Delta {u_1}^2 + \Delta {u_2}^2)}$. Ten spatial translations of 1 to 7 units (Euclidean distance) were applied to the simulated image and for each translation, it was compared to the translated image using all of the three indices described in Section V-A. Rotation was not applied to these images and the edge width was fixed at one pixel. In the third test, we analyzed the effect of the width of edges in combination with rotations and translations. For this, the line segments in the simulated image were dilated once and twice with a 3-by-3 square window. The two dilated images are shown (with solid lines) in Fig. 1(b) and (c), respectively. The dilated images were also rotated and translated as described previously and each pair of images were compared using the three indices described in Section V-A.

C. Results

1) Effect of Rotation: Fig. 1(d) shows the effect of rotation on the Dice, PCM and CW-SSIM indices. The values of the Dice index decrease drastically as the amount of rotation is increased (this is true for other intensity-based indices as well). In this simulation, the value of the Dice index is zero for any rotation greater than 1.6 degrees. For example, the reported similarity between the original image and a copy rotated 2 degrees [Fig. 1(d)] is the same as that between the original image and a copy that is rotated 4 degrees [Fig. 1(d)]. In comparison, PCM has a greater robustness than Dice does to rotational distortions. PCM is 0.67 for 2 degrees of rotation and 0.13 for 4 degrees of rotation [Fig. 1(d)]. Similarly, note that the CW-SSIM is more robust to rotational distortions than PCM. The CW-SSIM value is the same for 2 or 4 degrees of rotations [Fig. 1(d)]. That is, the response of the CW-SSIM index is invariant for all rotations while this is not case for the Dice and PCM indices.

2) Effect of Translation: Fig. 1(j) shows the effect of translation on the Dice, PCM and CW-SSIM indices. The Dice index decreases significantly as the amount of translation is increased [Fig. 1(j)]. In this simulation, the value of both indices is zero for any translation greater than 3 units (Euclidean distance). For example, the reported similarity between the original image and a copy translated by approximately 5 units [Fig. 1(j)] is the same as that between the original image and a copy that is translated by approximately 7 units [Fig. 1(j)]. Intensity-based indices cannot quantify such a difference in similarity. As was the case for rotational distortions, PCM has a greater robustness than the Dice index to translational distortions. For example, PCM is 0.43 for the first case and 0.22 for the second case [Fig. 1(j)]. The CW-SSIM is more robust to translational distortions than PCM [Fig. 1(j)]. Finally note that the response of the CW-SSIM index is invariant for all translations while this is not case for the Dice and PCM indices.

3) Effect of Width of Edges: The performance of the indices for three different edge widths and various rotations are depicted in Fig. 1(d), (e), and (f). Similarly, the performance of the indices for three different edge widths and various translations are depicted in Fig. 1(j), (k), and (l). The behavior of the Dice and PCM indices as a function of rotation or translation depends on the width used to represent the segmentations. For smaller values of the edge-width, the responses of both Dice and PCM are nonlinear. By comparison, for larger values of the edge-width, the responses of both Dice and PCM appear linear. In contrast, the response of the CW-SSIM index is the same for all widths and it is invariant to the rotations and translations for all widths.

4) Combined Effect Rotations, Translations and Width of Edges on CW-SSIM Index: The combined effect of various

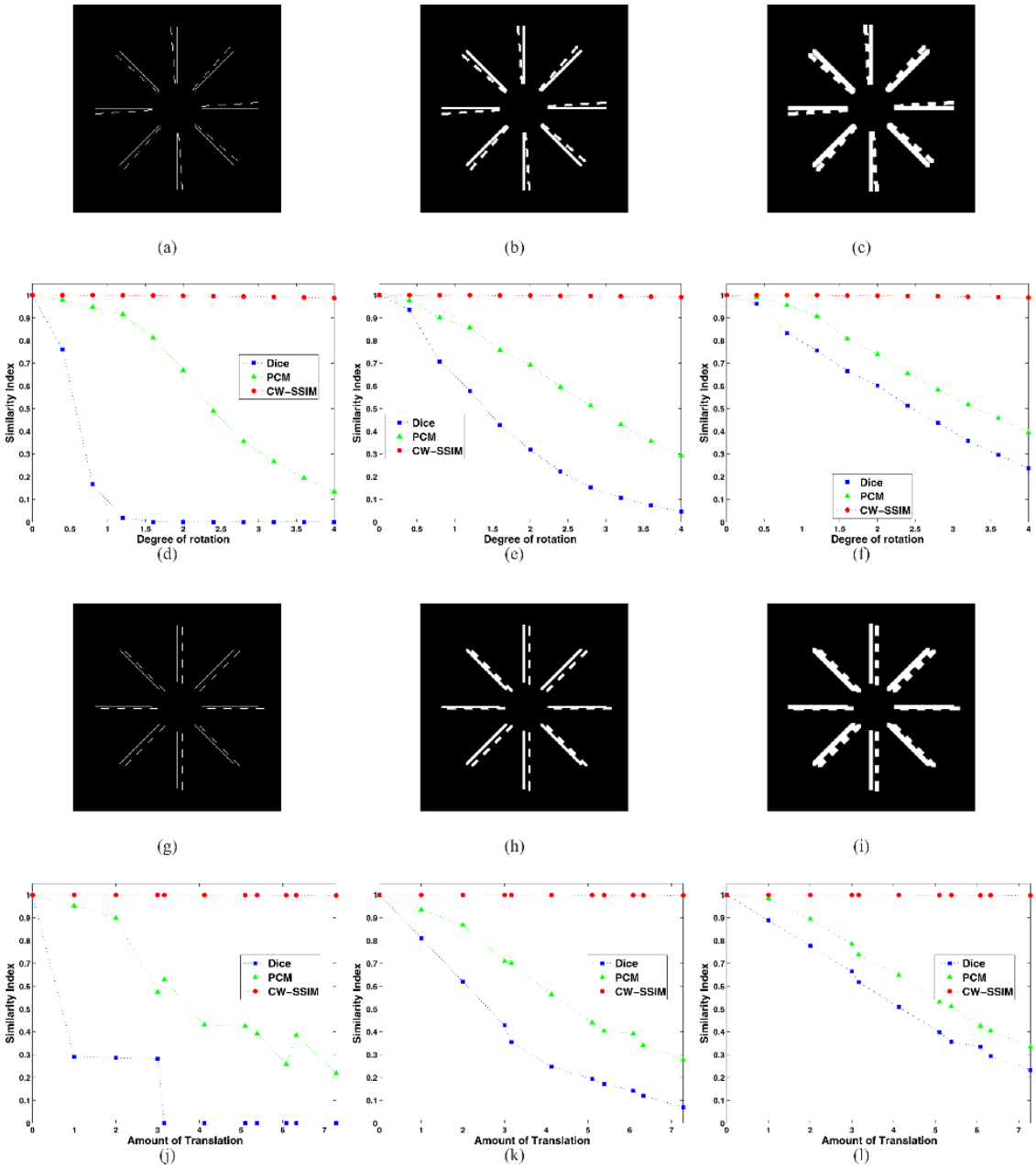


Fig. 1. *Effects of rotation and edge width on the similarity indices:* The original image was rotated from 0 to 4 degrees in increments of 0.4 degrees. The locations of the lines of the image that was obtained after applying a rotation are shown with dotted lines [for example, in (a), the original image is rotated by 4 degrees]. The original image was compared to each rotated image, and for each pair of images the similarity indices were computed. To study the effect of edge width, the image was dilated and the indices were recalculated; (b) and (c) show the simulated images after dilation by factors of 1 and 2, respectively, and (e) and (f) show the similarity indices as a function of rotation. *Effects of translation and edge width on the similarity indices:* The original image was translated in the x and y directions by different amounts. The locations of the lines of the image that was obtained after applying spatial translations are shown with dotted lines [for example, in (g), a translation of 7 pixels to the right and 2 pixels to the bottom is applied]. The original image was compared to each translated image and for each pair of images the similarity indices were computed. To study the effect of edge width on the indices, the image was dilated and the indices recalculated; (h) and (i) show the simulated images after dilation by factors of 1 and 2, respectively, and (k) and (l) show the similarity indices as a function of translation.

rotations and translations on the CW-SSIM index were also analyzed. For this purpose, rotations in the range 0 to 20 degrees and translations in the range 0 to 21 units (Euclidean distance) were simultaneously applied to create a transformed image. The original image was compared to each transformed image, and for each pair of images the CW-SSIM index was computed. To study the effect of edge width, the image was dilated by factors of 1 and 2 and the CW-SSIM indices were recalculated. The results are shown in Fig. 2(d), (e), and (f) in which each bar represents the CW-SSIM value for a particular pair of rotation and translation factors. Note that for all edge widths the CW-SSIM index is robust to the combined effects of rotation and translation.

5) Effect of Parameters of the Steerable Pyramid on the CW-SSIM Index: The two main adjustable parameters of the steerable pyramid used in the computation of the CW-SSIM index are: (a) the number of levels of decomposition (level) and (b) the number of orientations (ori). The effect of dif-

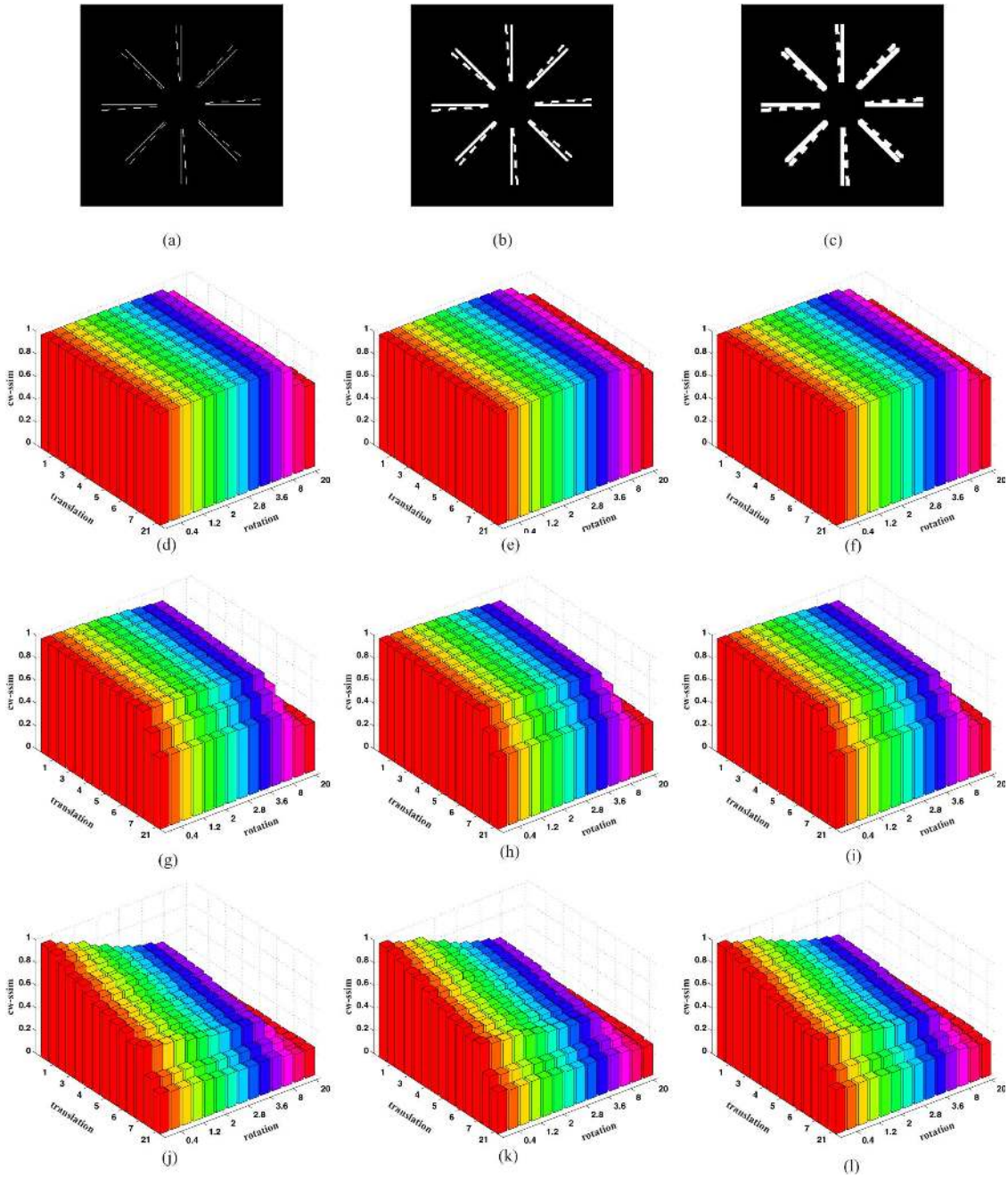


Fig. 2. *Combined effects of rotation, translation on the CW-SSIM index:* Rotations in the range 0 to 20 degrees and translations in the range 0 to 21 units (Euclidean distance) were simultaneously applied to create a transformed image. The original image was compared to each transformed image, and for each pair of images the CW-SSIM index was computed. To study the effect of edge width, the image was dilated by factors of 1 and 2 and the CW-SSIM indices were recalculated. The results are shown in (d), (e), and (f) in which each bar represents the CW-SSIM value for a particular pair of rotation and translation factors. Note that for all edge widths the CW-SSIM index is robust to the combined effects of rotation and translation. *The effect of parameters of the steerable pyramid on the CW-SSIM index:* The steerable pyramid used in the computation of the CW-SSIM index has two main parameters: (a) the number of levels of decomposition (level) and (b) the number of orientations (ori). The effect of different values for the level and ori parameters was analyzed. In (d), (e), and (f), the parameters are: $\mathrm{level} = 6, \mathrm{ori} = 16$. In (g), (h), and (i), the parameters are: $\mathrm{level} = 4, \mathrm{ori} = 16$. In (j), (k), and (l), the parameters are: $\mathrm{level} = 2, \mathrm{ori} = 16$. The CW-SSIM index obtained with $\mathrm{level} = 2, \mathrm{ori} = 16$ is the most sensitive to translations and rotations. The CW-SSIM index computed with multiple levels of decomposition shows the most robustness to translations and rotations. (d) $\mathrm{level} = 6, \mathrm{ori} = 16, \mathrm{width} = 1$ (e) $\mathrm{level} = 6, \mathrm{ori} = 16, \mathrm{width} = 2$, (f) $\mathrm{level} = 6, \mathrm{ori} = 16, \mathrm{width} = 3$, (g) $\mathrm{level} = 4, \mathrm{ori} = 16, \mathrm{width} = 1$, (h) $\mathrm{level} = 4, \mathrm{ori} = 16, \mathrm{width} = 2$, (i) $\mathrm{level} = 4, \mathrm{ori} = 16, \mathrm{width} = 3$, (j) $\mathrm{level} = 2, \mathrm{ori} = 16, \mathrm{width} = 1$, (k)$\mathrm{level} = 2, \mathrm{ori} = 16, \mathrm{width} = 2$, (l)$\mathrm{level} = 2, \mathrm{ori} = 16, \mathrm{width} = 3$.

ferent values for the level and ori parameters was analyzed. In Fig. 2(d)–(f), the parameters are: $\mathrm{level} = 6, \mathrm{ori} = 16$. In Fig. 2(g)–(i), the parameters are: $\mathrm{level} = 4, \mathrm{ori} = 16$. In Fig. 2(j)–(l), the parameters are: $\mathrm{level} = 2, \mathrm{ori} = 16$. The CW-SSIM index obtained with $\mathrm{level} = 2, \mathrm{ori} = 16$ is the most sensitive to translations and rotations. Note that, as expected the CW-SSIM index computed with multiple levels of decomposition shows the most robustness to translations and rotations. Note that the parameters $\mathrm{level} = 6$ and $\mathrm{ori} = 16$ are used to compute the CW-SSIM index for all experiments.

VI. Case Study II: Interobserver Similarity of Segmented Objects, Context Based

A. Introduction

In this case study, we quantify the similarity of multiple segmentations (generated by different human observers), of a given image. The segmentations are obtained from the publicly available Berkeley Segmentation Data-set and Benchmark (BSDB) [45]. The segmentations are "context-based" and the guidelines provided to the observers were as follows: "Divide each image into pieces, where each piece represents a distinguished thing in the image. It is important that all of the pieces have approximately equal importance. The number of things in each image is up to you. Something between 2 and 20 should be reasonable for any of our images" [45]. This database was created as a ground truth for fair comparisons of boundary detection and image segmentation algorithms and is an ideal test set for comparing the performance of image similarity indices.

B. Methods

The BSDB contains segmentations made on 300 grayscale and color images. These images are divided into a training set of 200 images and a test set of 100 images. In this study, we used all segmentations of the 100 grayscale images from the test set. Fig. 3 shows a few of the images and the segmentations made by the observers. Various experiments were carried out with the MSE, $\mathrm{MSE_{CP}}$, PHDM, and CW-SSIM indices. First, all segmentations of a given image were compared to each other and an average similarity measure was calculated. This procedure was repeated for all 100 images and histograms for each similarity index were computed. Second, each segmentation of an image was compared to every segmentation of another image and the average similarity was analyzed for each pair of different images. A total of 3179 pairs of images were compared and the histograms of the similarity indices were computed. To quantify the discriminability of the similarity indices, the following procedure was adopted. For each pair of distributions, a Receiver Operating Characteristic (ROC) curve was produced and the area under the curve (AUC) was used to measure the overlap between the two distributions of similarity values.

Finally, the BSDB also provides a measure of how much the various human observers agreed with each other. Each ground truth human segmentation was compared to the segmentations of the other observers in terms of precision, recall, and F-measure for each image [46]. Larger values of the F-measure denote greater agreement among the human observers [46]. The relationship between the F-measure and the various indices was studied with linear regression analysis.

C. Results

Fig. 3(b), (c), and (g) shows the segmentations made by three observers, on the image shown in Fig. 3(a). Notice the remarkable similarity between the segmentations in Fig. 3(b) and (c). This is reflected by a CW-SSIM index of 0.99 for this pair of images [Fig. 3(d)]. By comparison, one can see a number of differences between the segmentations Fig. 3(b) and (g) and the CW-SSIM index for this pair is 0.69. The $\mathrm{MSE_{CP}}$ and PHDM indices also capture this observation; however, the MSE index does not reflect this finding [Fig. 3(d) and (h)].

Fig. 3(j), (k), and (o) are the segmentations made by three observers, on the image shown in Fig. 3(i). The correspondence between the segmentations in Fig. 3(j) and (k) is more than that between the segmentations in Fig. 3(j) and (o) and this is also seen with the CW-SSIM, $\mathrm{MSE_{CP}}$ and PHDM indices. In addition, the segmentations in Fig. 3(b) and (c) show more similarity to each other than the pair of segmentations in Fig. 3(j) and (k). The CW-SSIM index for the former case is 0.99 [Fig. 3(d)], whereas it is 0.90 [Fig. 3(l)] for the latter.

Finally, note that the agreement between the segmentations in Fig. 3(j) and (k) $(\mathrm{CW{-}SSIM} = 0.90)$ is similar to that for the segmentations in Fig. 3(r) and (s) $(\mathrm{CW{-}SSIM} = 0.89)$. However, the $\mathrm{MSE_{CP}}$ and the PHDM values are quite different for the two cases [see Fig. 3(l) and (t)]. This would make it difficult to assign a "label" (e.g., high or medium similarity) to particular values of the $\mathrm{MSE_{CP}}$ and PHDM indices since the range of these indices is dependent on the contents of the images being compared. By comparison, by conducting human observer studies, it could be more feasible to create an "interpretation guideline" for the CW-SSIM index.

Similarity between segmentations of same and different images were reported and the maximum, median, and minimum values of the various indices are reported in Table IV. Fig. 4 shows the two histograms (for each index), when the similarity between segmentations of the same and different images are evaluated. One can see that the two histograms for the CW-SSIM index have the smallest overlap [Fig. 4(a)] whereas the two histograms for the MSE index [Fig. 4(g)] seem to have the largest overlap. This observation is confirmed by using the area under the ROC curve to quantify the overlap. The AUC value for the CW-SSIM, $\mathrm{MSE_{CP}}$, PHDM and MSE indices are 0.999, 0.978, 0.975, and 0.808, respectively. The results of the regression analysis are presented in Fig. 4(b), (d), (f), and (h). We observer a linear relationship between the F-measure and the CW-SSIM and PHDM indices.

VII. Case Study III: Similarity of Segmentations of Linear Structures; Spicules on Mammography

A. Introduction

Accurate image segmentation is critical for the detection and classification of objects-of-interest in medical images. Image segmentations may be generated by human observers or by computer algorithms. A segmentation is represented as a binary image where all pixels on and inside the indicated boundary are assigned a value of one, and zero otherwise. For dense objects, e.g., tumors, the segmentation may contain many pixels. By comparison, for linear structures, e.g., blood-vessels, the segmentation can be sparse. Similarly, small dense objects are represented by a few pixels (sparse segmentations). In medical imaging the segmentation of dense objects is encountered more frequently and so many indices have been developed for the comparison of segmentations of dense objects. However, few indices have been proposed for or tested on segmentations of linear structures in medical images.

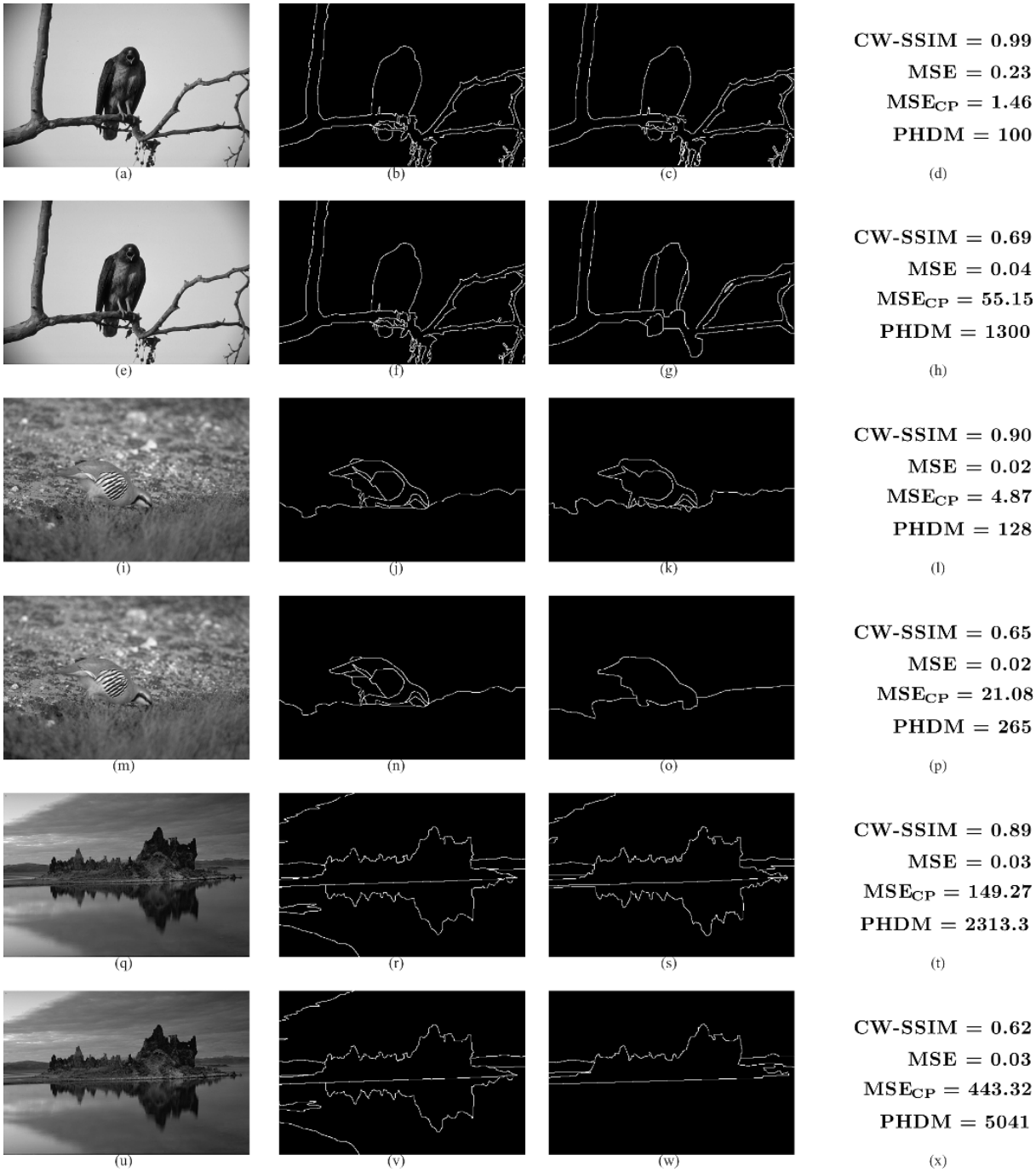


Fig. 3. Segmentations created by observers for a given gray-scale image [45]. The first column shows the original grayscale images and a pair of segmentations corresponding to that images are displayed in columns two and three. Column four shows the CW-SSIM, MSE, $\mathrm{MSE}_{\mathrm{CP}}$ and PHDM indices for each pair of segmentations.

B. Methods

Similarity indices were compared on tracings of linear objects on mammograms made by two experienced radiologists. The images for this study were obtained from the Digital Database for Screening Mammography (DDSM) [47], the largest publicly available data-set of digitized mammograms. Two radiologists (reader1 and reader2) traced all linear structures of interest (spiculations) on a set of 12 spiculated mass images. To compute an estimate of the intraobserver agreement, reader1 repeated the process. The analysis was conducted on regions-of-interest containing the spiculated masses using the ROI Manager plugin of NIH ImageJ.

C. Results

Quantitative measures of both the intra- and interobserver agreement were computed using the Dice, PCM, and CW-SSIM indices. Fig. 5 shows the outlines of spicules created by two radiologists (reader1 and reader2) on four spiculated masses. The first column shows the original images and the first and second

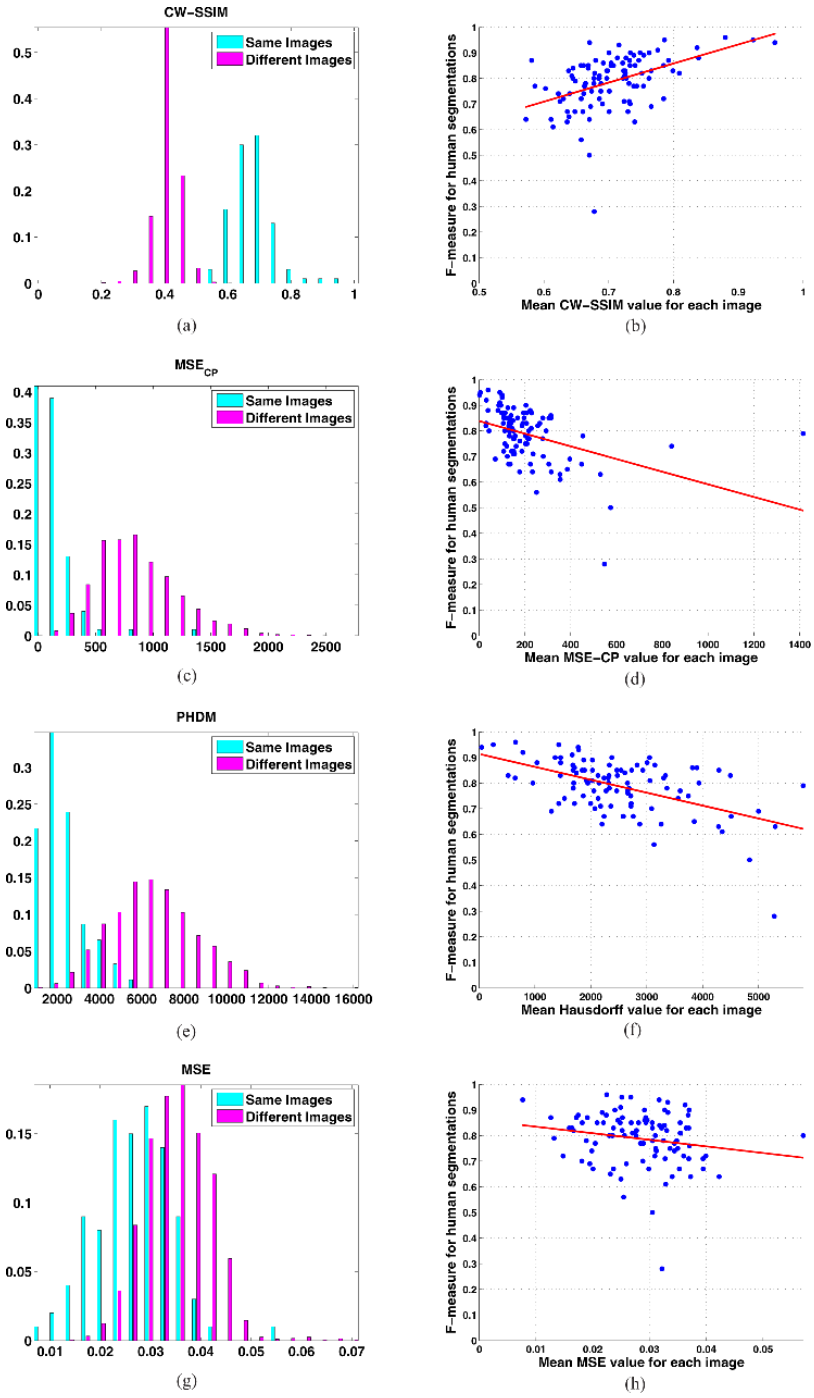


Fig. 4. (a), (c), (e), (g) Two histograms when segmentations of same and different images are compared. The overlap between each pair of histograms is measured by the area under the ROC curve (AUC). The CW-SSIM index shows the largest AUC value, and, thus, the greatest discrimination between these groups. (b), (d), (f), (h) Results of the regression analysis. We observer a linear relationship between the F-measure (of human observations) and the CW-SSIM and PHDM indices. (a) CW-SSIM, $\mathrm{AUC} = 0.999$, (b) CW-SSIM versus F-measure, (c) $\mathrm{MSE_{CP}}$, $\mathrm{AUC} = 0.978$ (d) $\mathrm{MSE_{CP}}$ versus F-measure, (e) PHDM, $\mathrm{AUC} = 0.975$, (f) PHDM versus F-measure, (g) MSE, $\mathrm{AUC} = 0.808$, (h) MSE versus F-measure.

sets of outlines made by reader1 are displayed in columns two and three. The outlines made by reader2 are shown in the fourth column. The values of the Dice, PCM, and CW-SSIM indices, for the intra- and interobserver measurements are shown in the fifth column. By visual inspection, it is apparent that there is substantial intraobserver agreement in the segmentations of some images [e.g., Fig. 5(a)].However, low values of the Dice index are obtained [e.g., $\mathrm{Dice} = 0.29$ for Fig. 5(a)]; note that a Dice value of >0.7 is considered to denote "excellent" agreement [20]. In contrast, the PCM and CW-SSIM values are much higher for these images [e.g., $\mathrm{PCM} = 0.36$ and $\mathrm{CW{-}SSIM} = 0.81$, for Fig. 5(a)]. On the other hand, the relative values of all three indices are generally consistent with subjective assessments. For example, by visual inspection, we note that the intraobserver agreement in the segmentations of the spiculated mass shown in Fig. 5(a) is more than that for the

TABLE IV
CW-SSIM, MSE, $\mathrm{MSE_{CP}}$, AND PHDM INDICES WERE USED TO COMPARE THE SEGMENTATIONS OF THE SAME IMAGES AND SEGMENTATIONS OF DIFFERENT IMAGES

	Segmentations of Same Images (100 images)		
Index	**Min. Value**	**Max. Value**	**Median**
CW-SSIM	0.57	0.96	0.70
MSE	0.01	0.06	0.03
$\mathbf{MSE_{CP}}$	2.36	$1.42 \cdot 10^3$	152.53
PHDM	44.30	$5.80 \cdot 10^3$	$2.33 \cdot 10^3$
	Segmentations of Different Images (3179 pairs of images)		
Index	**Min. Value**	**Max. Value**	**Median**
CW-SSIM	0.22	0.63	0.43
MSE	0.02	0.07	0.04
$\mathbf{MSE_{CP}}$	124.63	$2.74 \cdot 10^3$	868.68
PHDM	$1.11 \cdot 10^3$	$1.60 \cdot 10^4$	$6.77 \cdot 10^3$

spiculated mass in Fig. 5(p). This difference is reflected by all indices: $\mathrm{Dice} = 0.29$ versus 0.09, $\mathrm{PCM} = 0.36$ versus 0.20, and $\mathrm{CW{-}SSIM} = 0.81$ vs 0.62.

Similar trends are observed for the measurements of the interobserver agreement. For example, the interobserver agreement for the spiculated mass shown in Fig. 5(k) is only 0.10 by Dice. In comparison, PCM (0.26) and CW-SSIM (0.66) indicate more substantial agreement. As was the case for intraobserver agreement, all indices generally captured the trends in the interobserver agreement expected from visual inspection. For example, we can see that the interobserver agreement in the segmentations of the spiculated mass shown in Fig. 5(a) is more than that for the spiculated mass in Fig. 5(f). All three indices reflect this observation: $\mathrm{Dice} = 0.13$ versus 0.07, $\mathrm{PCM} = 0.32$ versus 0.16, and $\mathrm{CW{-}SSIM} = 0.69$ versus 0.58. Finally, one would expect that the intraobserver agreement should be higher than the interobserver agreement. All of the indices exhibit this trend for most of the segmentation pairs. The PCM values are greater for intraobserver agreement than for the corresponding interobserver agreement for 8 out of the 12 mammography cases, and the Dice and CW-SSIM values are greater for intraobserver agreement for 9 out of the 12 of spiculated masses.

VIII. CASE STUDY IV: 3-D FACE RECOGNITION

A. Introduction

A number of applications including access control, surveillance, criminal justice, and human computer interaction require automatic human identification. One of the most widely analyzed biometric methods for human identification is face recognition. However, it is difficult to develop robust, fully automated face recognition systems. For example, despite considerable progress in 2-D face recognition systems they are inadequate for robust face recognition. Their performance is reported to decrease significantly with varying facial pose, illumination and expression [48]. 3-D face recognition systems are less sensitive to changes in ambient illumination conditions than 2-D systems [49] and the pose of 3-D face models can also be easily corrected. Hence, considerable effort is now being directed toward 3-D face recognition.

In the computer vision literature, a 3-D object that cannot be recognized as either planar or naturally quadric is referred to as a "free form" object, e.g., the surface of the human face. One class of existing techniques for 3-D face recognition is based on the direct comparison of facial surfaces. The surface of a human face is represented as a point cloud and two facial surfaces are compared using the geometry-based indices of $\mathrm{MSE_{CP}}$ [49]–[51], [52] or PDHM [53]–[56]. These indices require that correspondences between pairs points on the two surfaces be established before the indices are computed. An alternative existing approach to facial surface matching is to compare range images generated from 3-D facial models using the intensity-based MSE [57]–[59]. A range image, also referred to as a 2.5D surface or depth map, consists of (x, y) points on a regular rectangular grid. Each (x, y) point is associated with a z value or depth of the point on the surface of the face, which is closest to an 3-D acquisition device. 3-D face recognition techniques that employ the MSE, however, are very sensitive to misalignment errors between facial range images. For such 3-D face recognition systems, the similarity/dissimilarity score between an incoming "probe" face and each face in a "gallery," is employed to index the gallery face closest in appearance to the probe face. Hence, the performance of 3-D face recognition systems that employ 3-D facial surface matching depends on the accuracy and robustness of the index employed. In this case study, we summarize our recent work on the use of the CW-SSIM index for facial surface matching [60].

B. Methods

Three-dimensional face models for the study were acquired by an MU-2 stereo imaging system manufactured by 3Q Technologies, Ltd. (Atlanta, GA). The data contained 360 models of 12 subjects. It was partitioned into a gallery set containing one image each of the 12 subjects in a neutral expression. The probe set contained 348 images of the gallery subjects in a neutral or an arbitrary expression. The probe set contained 29 range images of each subject. All 3-D face models were transformed to a frontal pose by iteratively aligning them to a reference model in a canonical frontal upright pose. Range images were constructed by orthographic projection of the 3-D models onto a regularly spaced rectangular grid. The range images were of size 751×501 pixels with a resolution of 0.32 mm in the x, y, and z directions. Range images were median filtered to remove spike noise, and interpolated to remove large holes. Fig. 6 presents example range images after applying these preprocessing steps.

We implemented four 3-D face recognition algorithms based on facial surface matching. The first one employed the CW-SSIM index to obtain similarity scores between pairs of facial range images. These similarity scores were then converted into distance values. For the second algorithm, distances between pairs of range images were obtained using the MSE index. For the third and the fourth algorithm, 3-D faces were regarded as point clouds and distances between them were quantified by means of the $\mathrm{MSE_{CP}}$ and PHDM with $P = Q = 0.9$, respectively. Verification performance of the four 3-D face recognition algorithms was evaluated using the receiver operating characteristic (ROC) methodology. The equal error rates (EER) and the areas under the ROC curves (AUC) of all algorithms were reported. Identification performance was evaluated by means of a cumulative match characteristic (CMC)

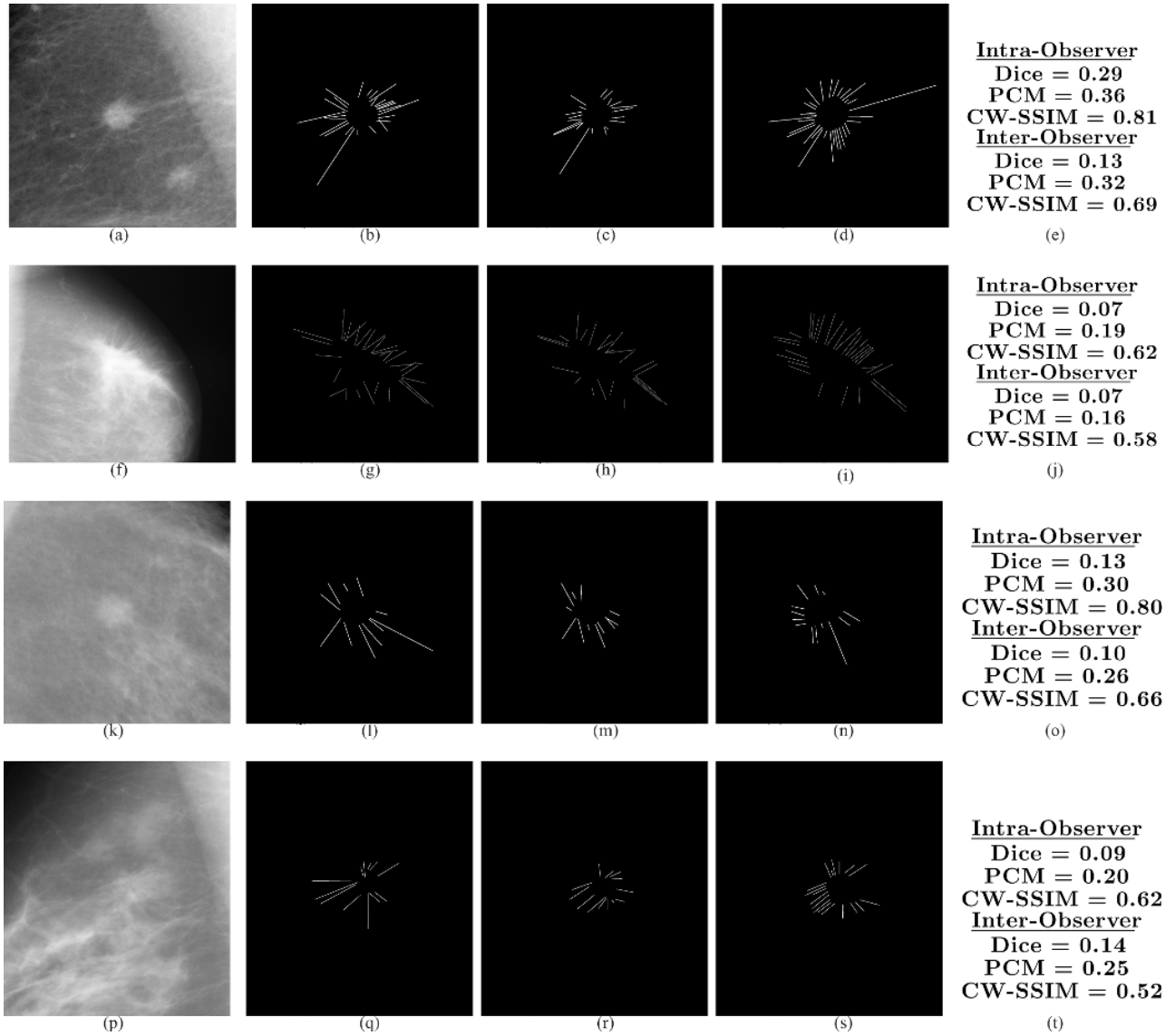


Fig. 5. Intra- and interobserver agreement: This figure shows the outlines of spicules created by two radiologists (reader1 and reader2) for four spiculated masses. The first column shows the original images and the first and second sets of outlines made by reader1 are displayed in columns two and three. The outlines made by reader2 are shown in the fourth column. The values of the Dice, PCM, and CW-SSIM indices, for the intra- and interobserver measurements are shown in the fifth column. (a), (b) reader1—(1st), (c) reader1—(2nd), (d) reader2—(1st), (e), (f), (g) reader1—(1st), (h) reader1—(2nd), (i) reader2—(1st), (j), (k), (l) reader1—(1st), (m) reader1—(2nd), (n) reader2—(1st), (o), (p), (q) reader1—(1st), (r) reader1—(2nd), (s) reader2—(1st), (t).

curve and the rank 1 recognition rates (RR) were observed. The 95% confidence intervals for each observed quantity was obtained by applying bootstrap sampling. The performance of each algorithm was evaluated separately for the entire probe set, for neutral probes only and for expressive probes only.

C. Results

The equal error rates, AUC values and rank 1 recognition rates of the four 3-D face recognition algorithms are presented in Table V. The CMC curves are presented in Fig. 7(a) and the ROC curves of the four 3-D face recognition algorithms are presented in Fig. 7(b). The proposed algorithm that employed CW-SSIM for matching 3-D facial range images, performed significantly better at identifying human subjects than the algorithms based on the other indices. It had a rank 1 $\mathrm{RR} = 98.6\%$, $\mathrm{CI} = [97.1 \quad 99.7]$ for all probe images. Among the other 3-D facial surface matching techniques that were implemented, the overall identification performance of the algorithm that used MSE was slightly better than the one that used the PHDM [Fig. 7(a)]. The algorithm that employed the $\mathrm{MSE_{CP}}$ index for surface matching performed the worst (rank 1 $\mathrm{RR} = 63.2\%$ for all probes).

Analogous to the identification performance, the verification performance of the algorithm based on the CW-SSIM index was superior to all the other algorithms with $\mathrm{EER} = 9.13\%$, $\mathrm{CI} = [7.71 \quad 10.5]$ for all probes [Fig. 7(b)]. AUC values of the CW-SSIM algorithm for both neutral and expressive probes was significantly lower than the AUC values for the other algorithms. The performances of algorithms that employed MSE and PHDM, respectively were not statistically significantly different, with $\mathrm{EER} = 17.7\%$ and $\mathrm{EER} = 16\%$, respectively, for all probes. Their AUC values for all probes were also not

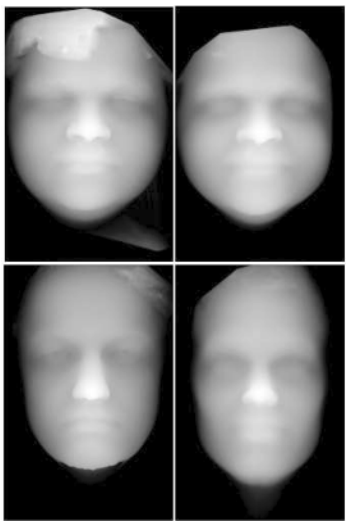


Fig. 6. Examples of the facial range images that were employed in case study IV. The images were preprocessed to remove noise and holes. The two images in the top row are of the same subject.

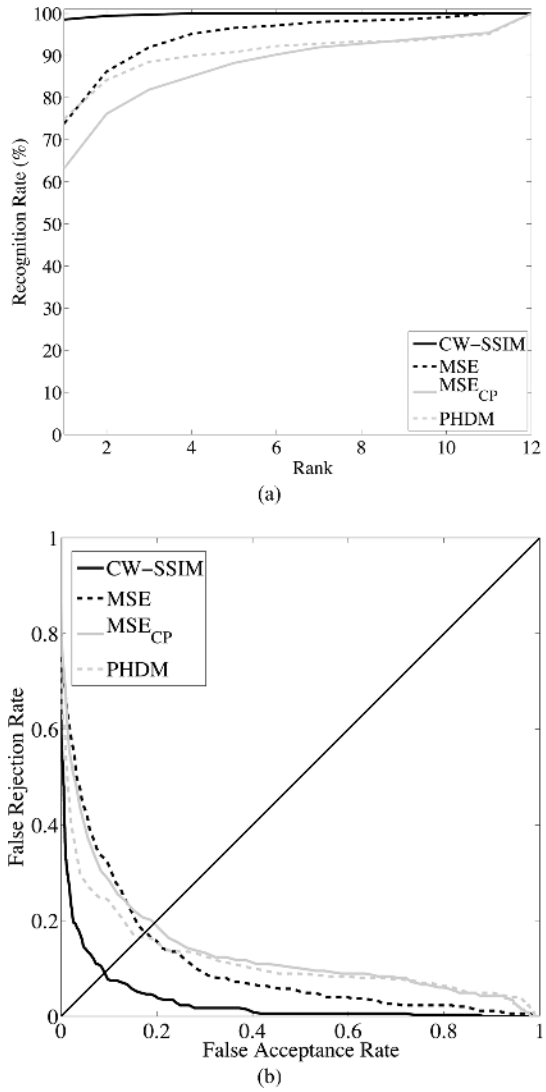


Fig. 7. (a) CMC curves for the identification performance of 3-D face recognition algorithms that matched pairs of facial surfaces using CW-SSIM, MSE, $\mathrm{MSE_{CP}}$, and PHDM. (b) ROC curves depicting the verification performance of these algorithms.

statistically significantly different (Table V). The highest EER of 19.5% for all probe images was observed for the algorithm that employed $\mathrm{MSE_{CP}}$. The CW-SSIM-based-algorithm for matching 3-D facial range images performed significantly better than existing algorithms based on the MSE, $\mathrm{MSE_{CP}}$, and PHDM.

IX. Discussion

We have compared indices for assessing the similarity of a given pair of images. The indices include intensity-based measures such as Dice and geometry-based approaches such as the Hausdorff distance. We have introduced the use of the CW-SSIM index to quantify image similarity and we have shown its applicability as a general purpose image similarity index for various image processing tasks. Towards this goal, the CW-SSIM index and the other indices were tested in four case studies. In the first case study, we applied the indices to simulated data and simulations were conducted to study the effect of small geometric distortions such as rotation and translation. This case study showed that the CW-SSIM index is robust to rotations, translations and to various combinations of these transformations.

In the second case study, the indices were applied to quantify the similarity of manual segmentations obtained from a publicly available database. The performance of the similarity indices was quantified with ROC analysis and it was shown that CW-SSIM provided the greatest discriminability. It is clear that when human observers provide ground truth segmentations, the consensus between their segmentations can vary and the F-measure has been used to quantify this agreement. It was encouraging to see a strong linear relationship between the CW-SSIM index and this measure of concordance.

The similarity indices were also compared for the task of quantifying intra- and interobserver agreement of segmentations of linear structures on mammograms. The Dice, PCM, and CW-SSIM indices were applied to mammography images in which observers traced spicules on spiculated masses. Both the intra- and interobserver agreement was quantified. One would expect that the intraobserver agreement should be higher than the interobserver agreement. All of the indices exhibit this trend for most of the segmentation pairs. However, as with the simulated data, the values of the Dice index for the intra- and interobserver agreement are low relative to the reported cutoff of 0.7 for "excellent" agreement.

In the final case study, we summarized a novel 3-D face recognition algorithm based on range image matching using the CW-SSIM index. We demonstrated that the proposed algorithm is more accurate and robust than some existing face recognition algorithms. In this work CW-SSIM was applied for quantifying similarity between 3-D facial surface but the techniques are not limited to face recognition. The study points towards the potential applicability of the CW-SSIM index to other 3-D pattern matching tasks as well. One limitation of the technique is that in its current form, the index can only be applied to range images. For matching 3-D objects, range images of coarsely registered objects would have to be created.

The success of the CW-SSIM index in these case studies can be attributed in part to the fact that CW-SSIM is robust to small

TABLE V
OBSERVED RANK 1 RR, EER, AND AUC VALUES AND THEIR 0.025 AND 0.975 QUANTILES FOR THE VERIFICATION AND IDENTIFICATION PERFORMANCE OF THE VARIOUS 3-D FACE RECOGNITION ALGORITHMS THAT COMPUTE SIMILARITY BETWEEN PAIRS OF 3-D FACIAL SURFACES USING CW-SSIM, MSE, $\mathrm{MSE_{CP}}$, AND PDHM

	Neutral Probes		Expressive Probes		All Probes	
Algorithm	RR (%)	95% CI	RR (%)	95% CI	RR (%)	95% CI
CW-SSIM	98.6	[96.8 100]	98.5	[96.2 100]	98.6	[97.1 99.7]
MSE	71.3	[65.3 77.3]	78.0	[70.4 84.8]	73.8	[69.1 79.0]
$\mathrm{MSE_{CP}}$	63.0	[56.5 69.0]	63.6	[54.5 72.0]	63.2	[58.0 68.2]
PHDM	75.0	[69.7 80.1]	75.0	[67.4 81.8]	75.0	[70.1 79.3]
	EER (%)	95% CI	EER (%)	95% CI	EER (%)	95% CI
CW-SSIM	8.29	[6.19 10.3]	9.37	[7.13 12.5]	9.13	[7.71 10.5]
MSE	18.7	[16.5 22.2]	15.5	[12.5 18.9]	17.7	[14.7 19.4]
$\mathrm{MSE_{CP}}$	20.6	[16.6 23.0]	18.2	[14.9 22.1]	19.5	[16.7 22.0]
PHDM	17.0	[14.1 21.8]	15.7	[12.4 19.4]	16.0	[15.5 19.3]
	AUC $\times 10^{-2}$	95% CI	AUC $\times 10^{-2}$	95% CI	AUC $\times 10^{-2}$	95% CI
CW-SSIM	2.87	[1.88 3.74]	3.29	[2.06 5.60]	3.11	[2.21 3.99]
MSE	11.2	[9.28 13.4]	10.1	[6.75 14.0]	10.7	[8.56 12.6]
$\mathrm{MSE_{CP}}$	14.1	[10.8 18.2]	12.7	[8.77 17.2]	13.7	[11.4 16.7]
PHDM	12.1	[9.30 17.1]	11.3	[8.18 16.2]	11.9	[10.1 14.7]

geometric distortions including small translations and rotations. Furthermore, the index is tailored to capture the local structure about a pixel irrespective of the local contrast or luminance values. Second, since CW-SSIM is robust to small image translations and rotations, it does not require registration as a preprocessing step. For example, some existing 3-D facial surface matching techniques [51], [56], require that every time a probe is presented to the gallery, it be finely registered to every model in the gallery before a measure of similarity can be reliably computed. For both 2-D and 3-D applications the CW-SSIM index has computational advantages. For example, computing CW-SSIM between a pair of range images/3-D models is less expensive than computing either the $\mathrm{MSE_{CP}}$ or the PHDM between a pair 3-D models. Similarly, for 2-D applications the computational requirements of the PCM index are significantly higher and a search neighborhood has to be specified upfront.

Another complex wavelet domain image similarity method is the index proposed by Lee *et al.* [61], which employs the Dual-Tree Wavelet transform. We believe that the steerable pyramid used in our approach has useful rotation-invariance property and is more flexible in adjusting the number of orientations. Gabor wavelets could also be used as an alternative. However, the Gabor wavelets are computationally more expensive and less convenient than the steerable pyramid.

In this paper, we propose the use of CW-SSIM as a general purpose similarity index. The generalizability is the key feature and, to us, the most interesting merit of our approach. We have demonstrated that the proposed CW-SSIM index, which does not use any knowledge about these specific types of distortions (by using low level image features, i.e., complex wavelet coefficient patterns), can work in a variety of applications. There are many other applications where CW-SSIM can be conveniently applied, such as: (a) evaluation of image compression and denoising algorithms; (b) comparing image segmentation results; (c) image registration; and (d) detecting and tracking objects in video.

Another application where image similarity plays an essential role is content-based image retrieval (CBIR). In CBIR, given a query image one is interested in finding an image from an existing database that most closely matches the query image. In most of the modern work in CBIR, feature vectors are extracted from the images and the "image similarity" is quantified by evaluating the similarity of the feature vectors. Typically features invariant to rotation, translation are designed [62]. While CW-SSIM could be used to directly compare the query image to database images, such comparisons would require much higher computational power than comparisons in low-dimensional feature vector spaces. Thus, we do not suggest that the current CW-SSIM method should be applied for CBIR without significant modifications. But for evaluating various image processing techniques (e.g., image segmentations, image denoising) CW-SSIM is an ideal image similarity index.

Assessing the similarity of a given pair of images is a challenging problem. While it is easy to define a multitude of potential indices, it is difficult to evaluate their performance since there is not a "ground truth" for the concept of similarity. Through careful simulations and demonstrations with real data, such as we have presented in this study, one can provide some insight into the relative strengths and weakness of different similarity indices. However, how to best interpret the values computed for a given similarity index remains elusive. In contrast, evaluation scales have been developed for indices for assessing similarity or agreement in other tasks. For example, there are established scales for interpreting the intraclass correlation coefficient (ICC), which is used to assess observer agreement in measurement studies (e.g., compare observers' measurements of tumor size), and the Kappa statistic, which is used to assess observer agreement in rating studies (e.g., compare observers' descriptions of mass margin). This work shows that it could be feasible to construct an evaluation scale for the CW-SSIM index. For example, such a scale could categorize a pair of images as different if the CW-SSIM index is below 0.6, and highly similar if it is greater than 0.9. Such criteria would tremendously increase the utility of the index and future work is needed to develop an "interpretability guideline" for the CW-SSIM index.

In conclusion, we have demonstrated the superiority of the CW-SSIM index against the other indices commonly used for assessing the similarity of a given pair of images. The CW-SSIM index has a number of advantages. It is robust to

small rotations and translations and it provides useful comparisons even without a preprocessing image registration step at the front end, which is essential for other indices.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Allen for sharing his implementation of the CDM and PCM indices [15], as well as Dr. G. Whitman and Dr. T. Stephens for tracing all spicules of the spiculated masses and we thank the anonymous reviewers and associate editor for their valuable comments.

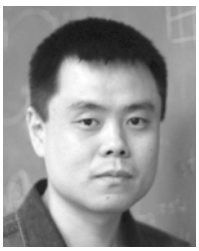
REFERENCES

[1] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
[2] P. Jaccard, "The distribution of flora in the Alpine zone," *New Phytol.*, vol. 11, pp. 37–50, 1912.
[3] G. Simpson, "Notes on the measurement of faunal resemblance," *Amer. J. Sci.*, vol. 258, pp. 300–311, 1960.
[4] A. Ochiai, "Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions," *Bull. Jpn. Soc. Sci. Fish*, vol. 22, pp. 526–530, 1957.
[5] J. Braun-Blanquet, *Plant Sociology: The Study of Plant Communities*. New York: McGraw Hill, 1932.
[6] R. Sokal and P. Sneath, *Principles of Numerical Taxonomy*. San Francisco, CA: W. H. Freeman, 1963.
[7] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York: Morgan & Claypool, Mar. 2006.
[8] A. B. Watson, "Dctune: A technique for visual optimization of dct quantization matrices for individual images," *Soc. Inf. Display Dig. Tech. Papers*, vol. XXIV, pp. 946–949, 1993.
[9] T. Pappas and R. Safranek, *Handbook of Image and Video Processing*. New York: Academic, 2000, ch. Perceptual Criteria for Image Quality Evaluation.
[10] S. Daly, *Digital Image and Human Vision*. Cambridge, MA: MIT, 1993, ch. The visible differences predictor: An algorithm for the assessment of image fidelity, pp. 179–206.
[11] J. Lubin, *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, ch. The use of psychophysical data and models in the analysis of display system performance, pp. 163–178.
[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[13] W. Pratt, *Digital Image Processing*. New York: Wiley, 2001.
[14] K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical roc curves," *Comput. Vis. Image Understand.*, vol. 84, no. 1, pp. 77–103, 2001.
[15] M. Prieto and A. Allen, "A similarity metric for edge images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1265–1273, Oct. 2003.
[16] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.
[17] P. Besl and H. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[18] G. R. Shi, "Multivariate data analysis in palaeoecology and palaeobiogeography-a review," *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, vol. 105, no. 3–4, pp. 199–234, Nov. 1993.
[19] K. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004.
[20] A. Zijdenbos, B. Dawant, R. Margolin, and A. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 716–724, Apr. 1994.
[21] S. Kulczynski, "Zespoly rslin w pieninach," *Bull. Int. Acad. Pol. Sci. Lettres*, vol. 2, pp. 57–203, 1928.
[22] Z. Hubalek, "Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation," *Biol. Rev.*, vol. 57, no. 4, pp. 669–689, 1982.
[23] P. Russel and T. Rao, "On habitat and association of species of anopheline larvae in south-eastern Madras," *J. Malaria Inst. India*, vol. 3, pp. 153–178, 1940.
[24] R. Sokal and C. Michener, "A statistical method for evaluating systematic relationships," *Univ. Kansas Sci. Bull.*, vol. 38, pp. 1409–1438, 1958.
[25] U. Yule, "On the association of attributes in statistics," *Philosoph. Trans. Roy. Soc. London*, vol. 194, pp. 257–319, 1900.
[26] D. Rogers and T. Tanimoto, "A computer program for classifying plants," *Science*, vol. 132, pp. 1115–1118, 1960.
[27] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[28] M. Clark, A. C. Bovik, and W. S. Geisler, "Texture segmentation using gabor modulation/demodulation," *Pattern Recognit. Lett.*, vol. 6, no. 4, pp. 261–267, Sep. 1987.
[29] M. Clark and A. C. Bovik, "Experiments in segmenting texton patterns using localized spatial filters," *Pattern Recognit.*, vol. 22, no. 6, pp. 707–717, 1989.
[30] A. Bovik, M. Clark, and W. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.
[31] J. A. Solomon and D. G. Pelli, "The visual filter mediating letter identification," *Nature*, vol. 369, pp. 395–397, 1994.
[32] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, pp. 529–541, 1981.
[33] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recognit. Lett.*, vol. 6, pp. 303–313, 1987.
[34] P. Kovesi, "Phase congruency: A low-level image invariant," *Psych. Res.*, vol. 64, pp. 136–148, 2000.
[35] Z. Wang and E. P. Simoncelli, "Local phase coherence and the perception of blur," presented at the Adv. Neural Information Processing Systems (NIPS03), vol. 16 (Cambridge, MA), MIT Press, May 2004.
[36] D. J. Fleet and A. D. Jepson, "Stability of phase information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1253–1268, Dec. 1993.
[37] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, pp. 49–71, 2000.
[38] J. Daugman, "Statistical richness of visual phase information: Update on recognizing persons by iris patterns," *Int. J. Comput. Vis.*, no. 45, pp. 25–38, 2001.
[39] D. A. Pollen and S. F. Ronner, "Phase relationships between adjacent simple cells in the cat," *Science*, no. 212, pp. 1409–1411, 1981.
[40] I. Ohzawa, G. DeAngelis, and R. Freeman, "Stereoscopic depth descrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, no. 249, pp. 1037–1041, 1990.
[41] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Amer.*, vol. 2, no. 2, pp. 284–299, Feb. 1985.
[42] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, no. 9, pp. 181–197, 1992.
[43] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neurosci.*, no. 4, pp. 819–825, 2001.
[44] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.
[45] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.
[46] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
[47] M. Heath, K. Bowyer, and D. Kopans, "Current status of the digital database for screening mammography," in *Digital Mammography*. Norwell, MA: Kluwer, 1998, pp. 457–460.
[48] P. J. Phillips, P. Grother, and R. J. T. Micheals, Frvt 2002: Overview and Summary [Online]. Available: http://www.frvt.org Mar. 2003
[49] E. P. Kukula, S. J. Elliott, R. Waupotitsch, and B. Pesenti, "Effects of illumination changes on the performance of Geometrix FaceVision® 3D FRS," in *Proc. 38th Int. Carnahan Conf. Security Technology*, 2004, pp. 331–337.
[50] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-D," in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp. 232–233.
[51] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5d face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.
[52] T. Papatheodorou and D. Rueckert, "Evaluation of automatic 4d face recognition using surface and texture registration," in *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 321–326.
[53] B. Achermann and H. Bunke, "Classifying range images of human faces with hausdorff distance," in *Proc. 15th Int. Conf. Pattern Recognition*, 2000, vol. 2, pp. 809–813, vol. 2.

[54] G. Pan, Y. Wu, and Z. Wu, "Investigating profile extracted from range data for 3D face recognition," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, 2003, vol. 2, pp. 1396–1399, vol. 2.
[55] T. D. Russ, M. W. Koch, and C. Q. Little, "3D facial recognition: A quantitative analysis," in *Proc. 38th Int. Carnahan Conf. Security Technology*, 2004, pp. 338–344.
[56] T. D. Russ, M. W. Koch, and C. Q. Little, "A 2D range Hausdorff approach for 3D face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition*, 2005, vol. 3, pp. 169–169.
[57] G. G. Gordon, "Face recognition based on depth maps and surface curvature," in *Proc. SPIE Geometric Methods in Computer Vision*, 1991, vol. 1570, pp. 234–247.
[58] S. Lao, Y. Sumi, M. Kawade, and F. Tomita, "3D template matching for pose invariant face recognition using 3D facial model built with isoluminance line based stereo vision," in *Proc. 15th Int. Conf. Pattern Recognition*, 2000, vol. 2, pp. 911–916.
[59] Y. Lee and T. Yi, "3D face recognition using multiple features for local depth information," in *Proc. 4th EURASIP Conf. Video/Image Processing and Multimedia Communications*, 2003, vol. 1, pp. 429–434.
[60] S. Gupta, M. P. Sampat, M. K. Markey, A. C. Bovik, and Z. Wang, "Facial range image matching using the complex wavelet structural similarity metric," in *Proc. IEEE Workshop on Applications of Computer Vision*, Austin, TX, Feb. 21-22, 2007, pp. 4–9.
[61] M.-S. Lee, L.-y. Liu, and F.-S. Lin, "Image similarity comparison using dual-tree wavelet transform," in *Proc. Advances in Image and Video Technology*, 2006, pp. 189–197, Lecture Notes in Computer Science.
[62] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, Apr. 2008.

Mehul P. Sampat (S'97–M'04) received the B.E. degree in biomedical engineering from the University of Mumbai, India, in 2000, and the M.S. and Ph.D. degrees in biomedical engineering from The University of Texas (UT) at Austin in 2002 and 2006, respectively.

At UT Austin, he was awarded a predoctoral traineeship award through the Department of Defense Breast Cancer Research Program. He was one of the 50 students selected to attend the 2004 IEEE International Summer School on Medical Imaging. He is currently an Associate Specialist in the Neurology Department at the University of San Francisco. Before that, he was a Research Fellow at Brigham and Women's Hospital and Harvard Medical School and a Research Engineer at Advanced Digital Imaging Research, LLC. His research interests include image and signal processing, pattern recognition, medical image analysis, biomedical informatics, neuro-imaging, and computer-aided detection and diagnosis



Zhou Wang (S'99–A'01–M'02) received the Ph.D. degree from The University of Texas at Austin in 2001.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. Before that, he was an Assistant Professor in the Department of Electrical Engineering, The University of Texas at Arlington; a Research Associate at Howard Hughes Medical Institute and New York University; and a Research Engineer at AutoQuant Imaging, Inc. His research interests include image processing, coding, communication, and quality assessment; computational vision and pattern analysis; multimedia coding and communications, and biomedical signal processing. He has more than 60 publications and one U.S. patent in these fields, and is an author of *Modern Image Quality Assessment* (Morgan & Claypool, 2006).

Dr. Wang is an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and *Pattern Recognition* and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING Special Issue on Visual Media Quality Assessment.

Shalini Gupta received the B.E. degree in electronics and electrical communication engineering from Punjab Engineering College, India, in 2002, and the M.S.E. and the Ph.D. degrees in electrical and computer engineering from the University of Texas at Austin in 2004 and 2008, respectively.

Currently, she is an Imaging and Architecture Scientist at the Wireless Division of Texas Instruments Incorporated. In the past, she worked with AT&T Laboratories and Advanced Digital Imaging Research, LLC. Her research interests include signal, image, video, and 3-D graphics processing, 3-D face recognition, computer vision, machine learning, detection theory, statistical pattern classification, medical image analysis, bioinformatics, computer aided diagnosis, and digital still image processing for wireless platforms.

Dr. Gupta is a recipient of the Summer Research Fellowship awarded by the Jawaharlal Nehru Center for Advanced Scientific Research, Bangalore, India.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1980, 1982, and 1984, respectively.

He is currently the Curry/Cullen Trust Endowed Professor at The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering (LIVE) in the Center for Perceptual Systems. His research interests include image and video processing, computational vision, digital microscopy, and modeling of biological visual perception. He has published over 450 technical articles in these areas and holds two U.S. patents. He is also the author of *The Handbook of Image and Video Processing* (Elsevier, 2005, 2nd ed.) and *Modern Image Quality Assessment* (Morgan & Claypool, 2006).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: the Education Award (2007); the Technical Achievement Award (2005); the Distinguished Lecturer Award (2000); and the Meritorious Service Award (1998). He is also a recipient of the Distinguished Alumni Award from the University of Illinois at Urbana-Champaign (2008), the IEEE Third Millennium Medal (2000), and two journal paper awards from the International Pattern Recognition Society (1988 and 1993). He is a Fellow of the Optical Society of America the Society of Photo-Optical and Instrumentation Engineers. He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 19961998; Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 19962002; Editorial Board, PROCEEDINGS OF THE IEEE, 19982004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003–present; and Founding General Chairman, First IEEE International Conference on Image Processing, Austin, TX, November 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.

Mia K. Markey (S'00–M'02–SM'09) received the B.S. degree in computational biology in 1998 from Carnegie Mellon University, Pittsburgh, PA, and the Ph.D. degree in biomedical engineering in 2002 from Duke University, Durham, NC.

She is currently an Associate Professor of biomedical engineering at the University of Texas at Austin, where she directs the Biomedical Informatics Laboratory (BMIL). Her lab designs cost-effective, computer-based decision aids. The BMIL develops decision support systems for clinical decision making and scientific discovery using artificial intelligence and signal processing technologies. The BMIL's research portfolio also includes projects in biometrics.

Dr. Markey has been recognized for excellence in research and teaching with awards from organizations such as the American Medical Informatics Association and the American Society for Engineering Education.