# Hierarchical Image Segmentation by Structural Content

Nathir A. Rawashdeh
Lexmark International, Inc., Lexington, KY, USA
Email: nrawashd@lexmark.com

Shaun T. Love
Lexmark International, Inc., Lexington, KY, USA
Email: slove@lexmark.com

Kevin D. Donohue
University of Kentucky, Lexington, KY, USA
Department of Electrical and Computer Engineering
Email: donohue@engr.uky.edu

***Abstract*— Image quality loss resulting from artifacts depends on the nature and strength of the artifacts as well as the context or background in which they occur. In order to include the impact of image context in assessing artifact contribution to quality loss, regions must first be classified into general categories that have distinct effects on the subjective impact of the particular artifact. These effects can then be quantified to scale the artifact in a perceptually meaningful way. This paper formulates general context categories, develops automatic image region classifiers, and evaluates the classifier performance using images containing multiple categories. Linear classifiers are designed to identify three main classes which include random, textured, and transient regions. Features for identifying these areas over regions at multiple resolutions are based on the optical density histogram (ODH), the cortex transform, and the co-occurrence matrix. It was found that selecting features from the ODH and cortex transform provides classification results in agreement with human assessment, and performances comparable to those of classifiers using co-occurrence matrix features. Experiments to assess performance show misclassification rates ranging from 3.3% for the lowest resolutions to 32.2% at highest. This paper also presents a hierarchical classification algorithm that combines classifiers operating at multiple resolutions and achieves an overall misclassification rate as low as 4.8%.**

***Index Terms*— hierarchical classifier, classification confidence, image structure, image quality, image segmentation, cortex transform**

## I. Introduction

The loss in perceived image quality is often determined by the nature and level of an artifact along with the context in which it appears. For example, in a highly structured image containing lines and edges, sharpness will likely be the most critical attribute in ranking image quality; where as, low-frequency uniformity may have little impact on the quality decision. The reverse may be true if the image content consisted of low frequency fields (such as sky or sand). Such fields with no structure or periodicity are robust to blur, but can be susceptible to periodic artifacts such as banding, or structural artifacts such as contouring or blocking. textures tend to mask structural defects and noise, but are prone to periodic artifacts.

Prior work ranks the relative impact of artifacts on the judgment of subjective image quality for various image types and suggests local defect characterization for overall image quality evaluation [1, 2]. In addition, document analysis and understanding research has been given much attention [3-9]. For example, a scanned document page is likely to contain objects that require customized enhancement such as sharpening for text regions, color processing for images and graphics, and clipping to remove document background pixels [3, 4]. Here, classification is performed on each pixel, labeling it as text, image or background. This is a bottom-up approach, whereas the work presented here adopts a top-down approach.

This paper examines ways to classify image regions hierarchically based on the level of structural content. The image is divided into regions (or tiles) and assigned to one of three classes – random, texture, or transient. A classification confidence measure is applied to detect regions that cannot be classified at a set confidence level [10]. In this case it is assumed that the region content may belong to more than one class. Hence, a low-confidence region is then divided into smaller regions, which are again classified, using the next higher resolution classifier. The confidence rule allows for hierarchical subdivision resulting in a classification of image content over the whole image at appropriate resolution levels. In addition, image content that could not be classified confidently at the highest resolution is labeled as such, and is available for further analysis or exclusion from the image quality loss estimation.

The first image region class considered is a random field for which contrast variations are randomly distributed and on the same order of magnitude. This region is void of structure and is characterized by its power spectral density (PSD) and optical density histogram (ODH). The second region type is a textured field. Like the random field, the textured field has no dominant or isolated structures over the spatial region; however, this region contains quasi-periodic structures of relatively low frequency. The presence of structure results in PSD amplitude modulations; therefore, characterizations of textured fields require additional analyses, such as the harmonic analysis of the PSD or spatial autocorrelation. The third region type is a transient field, primarily consisting of isolated lines and edges. This region is highly structured with strong spatial localization, and high frequency spectral content. The spatial orientation of the edges or lines can be further detected and utilized in the classifier feature selection process through the cortex transform, which indicates both spatial frequency and orientation [11, 12].

This work examines feature selection for the design of linear classifiers [13, 14, 15]. It extends single resolution classification in [16] to include 3 window sizes in a hierarchical classifier. Exhaustive feature-combination searches are used to identify the best features from the ODH, cortex transform, and co-occurrence matrices.

The paper is organized as follows. Section II describes the experiments used to design classifiers and assess performance. Section III describes the process of feature selection. Section IV presents performance results for classifiers at the lowest resolution levels. Section V presents the best classifiers at different window sizes, and Section VI develops the hierarchical multi-resolution classifier. Section VII illustrates example image segmentations using the multi-resolution classifier. Finally, results are summarized and presented with conclusions in Section VIII.

## II. METHODS

A total of 150 image regions of size 256 x 256 pixels were extracted from natural images, ranked according to their structural content and assign to one of the three categories (random, texture, or transient) by human observers. This image set was divided into 60 training image regions (20 images per class), and 90 test image regions (30 images per class). Images showing complex scenes were created using a five megapixel color digital camera and converted to the CIE L*a*b* color space, where only the L-channel image was used for analysis.

The image columns in Fig. 1 show examples of the 3 image region classes used in his study. They represent the complex content found in real-life images. The left column contains random class images (table cloth; balloon rubber; brushed steel; blue sky; red paint). The middle and right columns contain examples of the texture and transient classes, respectively. From each training and test image regions, a total of 108 features are extracted at every resolution level. The features are listed in Table I and are based on the ODH, cortex filter, and the co-occurrence matrix. Initially, potential features were found using stepwise regression models [17]. This interactive examination helped identify the impact of individual features and resulted in a pool of 108 features.

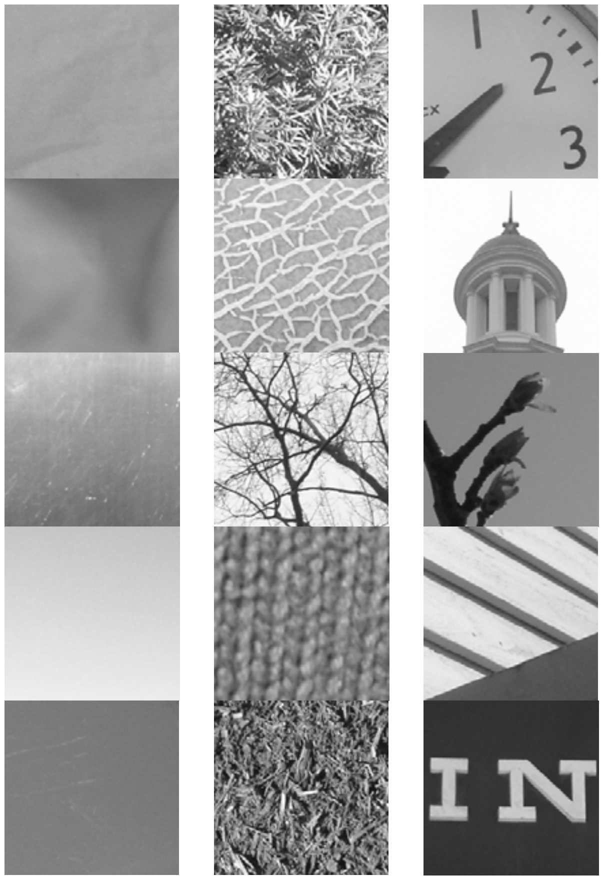


Figure 1. Columns show example image regions of the three classes - random, texture, and transient respectively.

The best performing classifier feature sets were identified through a Monte Carlo search where feature combinations were randomly selected. This work considered feature set sizes of eight, six, five and four features, and picked feature sets at random from the feature pool 2 x$10^5$ times per feature set size. The search program recorded the classification performances of the randomly selected feature sets for post analysis.

The best set size was determined to be 4 for the lowest resolution. Sets of 5 and 6 features did not outperform those of size 4. A similar pattern was observed at higher resolutions. The reduction in performance for larger feature sets is likely due to the fact that increasing the number of features caused the classifiers to learn the training set variations (i.e., overfitting) and as a result performed poorly on the test images.

Table II shows the classifier performances at the largest window size using the best sets of four features as determined by the Monte Carlo search. The classifiers implemented in this work are basic linear discriminant function classifiers that fit a multivariate normal density function to each class and use a pooled estimate of covariance [14, 15].

## III. CLASS FEATURES

The human visual system's (HVS) sensitivity exhibits orientation and radial spatial frequency selectivity. These properties can be modeled as visual channels using the cortex transform, which uses radial frequency (Dom) and orientation (Fan) filters [11]. The superposition of these filters is termed *cortex* filter, and their effects are cascaded to describe the combined radial frequency and orientation selectivity of cortical neurons. Fig. 2 shows a decomposition of the spatial frequency plane according to Daly's modified cortex filter [12].

The result of applying the cortex filter to an image is similar to computing the two dimensional power spectral density. Both the PSD and cortex filter output show the directional distribution of spatial frequency energy; however, the cortex filter performs a subband filtering analogous to the independent visual channels of the HVS. This can be used to predict masking of image content based on its distribution over the independent visual channels. The discrete energy distribution also helps classify image content, as this study shows.

As can be seen in Fig. 3, the cortex filter output for the three classes of interest differs qualitatively across the three classes. For the random-class image (top row), the energy is relatively low and uniform outside of the baseband. The texture-class image (middle row) has a similar uniform energy distribution; however, with more low frequency content. This means that Fan filter energies exhibit a sharper roll-off for the texture, compared to the random, case. For the transient-class image region (bottom row), the energy is directional and normal to the edges with notable high-frequency content. The directionality of energy distribution can be measured by the variation in Dom energies, in this case.

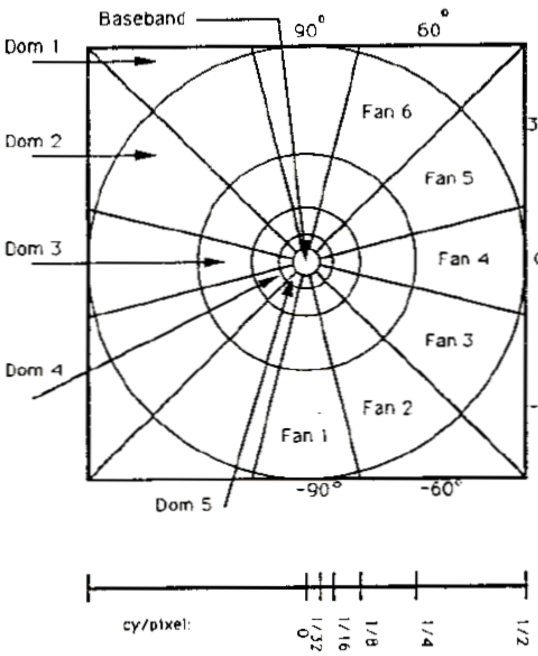


Figure 2. The cortex filter's radial and orientation channels. This figure is adapted from Daly's work [12].

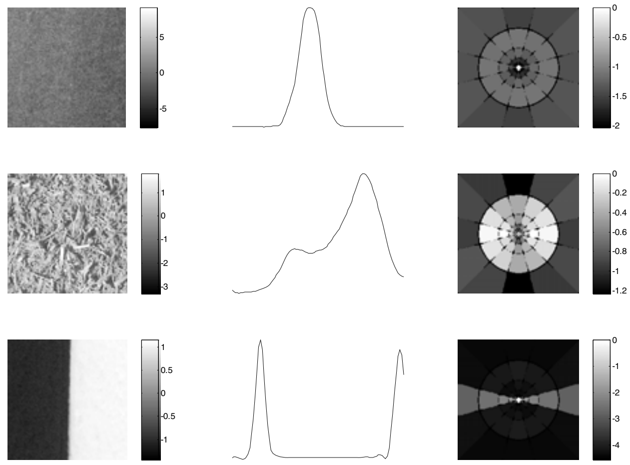

Figure 3. Optical density histograms and visual channel energies (using Daly's cortex filter) for the three classes.

The center column in Fig. 3 shows the typical ODH's of the three classes. A random region tends to have similar pixel values throughout; this is indicated by a single narrow peak of the ODH. For the texture class, the pixel values tend to spread out and have multiple peaks in the ODH. Lastly, the transient class ODH shows a bimodal shape because it mostly contains only two pixel values – dark and light for example. It is interesting to note that line textures, such as the cantaloupe skin in Fig. 1, also have bimodal ODH's; however, their visual channel energies tend to radiate in multiple directions.

Features related to changes in the energy distribution over the cortex channels, as well as shape changes in the ODH shape, were most critical for classification. The impact of overall gray level on the classification was removed by subtracting out the mean gray level from the local region, and normalizing by the standard deviation. Both the original mean and standard deviation were recorded as individual features. Features were grouped into two categories – features extracted from the ODH and cortex transform, and those extracted from the co-occurrence matrix. These are shown in Table I.

The features of the co-occurrence matrix used in this work are described below. These features are based on the normalized gray level co-occurrence (or spatial dependence) matrix as defined by Haralick *et al.* [18]. The sum of elements of the normalized co-occurrence matrix **P** equals unity after normalization. This allows **P** to be seen as a probability distribution of gray-level occurrences across the image region. For example, a textured region will have certain gray-level values repeat at pixel distances close to the periods present in the texture; those distances will appear as higher probabilities in the co-occurrence matrix. The following equations define features of the co-occurrence matrix which were considered for the implemented linear classifiers:

$$C_1 = \sum_{i,j} |i-j|^2 \cdot p(i,j)\,, \qquad (1)$$

$$C_2 = \frac{\sum_{i,j}(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \cdot \sigma_j}, \quad (2)$$

$$C_3 = \sum_{i,j} p(i,j)^2 , \quad (3)$$

$$C_4 = -\sum_{i,j} p(i,j) \cdot \ln[p(i,j)] , \quad (4)$$

$$C_5 = \sum_{i,j} \frac{p(i,j)}{1+|i-j|^2} , \quad (5)$$

where $p(i,j)$ are elements of $\mathbf{P}$, $C_1$ is the *contrast*, $C_2$ is *correlation*, $C_3$ is *energy*, $C_4$ is *entropy*, and $C_5$ is *homogeneity*. In addition, $\mu_i$ and $\mu_j$ are the means of the marginal probabilities $P_i$ and $P_j$, and $\sigma_i$ and $\sigma_j$ are their standard deviations - as follows:

$$p_i(i) = \sum_j p(i,j) , \quad (6)$$

$$\mu_i = \sum_i i \cdot p_i(i) , \quad (7)$$

$$\sigma_i = \sqrt{\sum_i (i-\mu_i)^2 p_i(i)} \quad (8)$$

The features defined in (1) through (5) are computed for each of the co-occurrence matrices, as seen Table I (features 29–108). For each image direction (0, 45, 90, and 135 degrees), we find the five features of the co-occurrence matrices, which have been computed for distances of $d = \{1, 4, 8, 16\}$ pixels.

The computed co-occurrence features provide measures of texture; for instance, contrast is zero for a constant image, i.e. close to zero for images of the random class. The absolute value of correlation is closer to one, the more pixel values correlate with their neighbors; and is relatively large for the random class and decreases for texture. Similarly, energy is unity for a constant image and decreases for textures. Entropy is largest for textures, and Homogeneity is large for the random and transient classes.

The features in Table I are computed for all the training and test images in this study. For the performance analysis, sets of features containing only ODH and cortex filter energies are compared to sets of only co-occurrence matrix features. In addition, features from both groups are also used for classification.

TABLE I.
DESCRIPTION OF CONSIDERED FEATURES

| Feature | Feature Description |
|---|---|
| 1 | Variance of RMS normalized L-channel image |
| 2 - 5 | Features related to peaks of the L-channel image ODH |
| 6 - 8 | Features related to peaks of the baseband ODH |
| 9 - 13 | Variance of Dom energies over 6 Fan directions |
| 14 - 18 | Range of Dom energies over 6 Fan directions |
| 19 - 21 | Variance, range and mean of f Fan energy slopes |
| 22, 23 | Magnitude of flattest and steepest slopes of linear fits to fan filter energies |
| 24 - 26 | Variance, range and mean of Fan filter slopes |
| 27, 28 | Maxima of ODH for L-channel image, and baseband filtered L-channel image |
| 29 - 48 | 0-deg. co-occurrence features at d=1,4,8,16 pixels |
| 49 - 68 | 45-deg. co-occurrence features at d=1,4,8,16 pixels |
| 69 - 88 | 90-deg. co-occurrence features at d=1,4,8,16 pixels |
| 89 - 108 | 135-deg. co-occurrence features at d=1,4,8,16 pixels |

## IV. CLASSIFIER PERFORMANCE

Five sets of four features FS1 trough FS5 are shown in Table II Those feature sets have been determined (through a Monte Carlo search) to have the best classification performances at the window size 256 x 256. The error measures presented in the table are computed as follows:

$$E_{test} = 100 \times \frac{M_{test}}{N_{test}} , \quad (9)$$

where $M_{test}$ is the number of misclassifications in the set of test images, which contains a total of $N_{test}$ = 90 images (30 per class). Similarly,

$$E_{train} = 100 \times \frac{M_{train}}{N_{train}} , \quad (10)$$

where $M_{train}$ is the number of misclassifications in the set of training images, which contains a total of $N_{train}$ = 60 images (20 per class). The feature sets in Table II are divided into three categories. The first category includes features from the ODH and cortex filter only (FS1). The second category includes only co-occurrence matrix features (FS2 and FS3), and is examined to contrast its classification performance with the features from the first category. The third group of feature sets includes features from both the first and second categories (FS4 and FS5).

Note that FS4 includes three features from the ODH and cortex filter, and only one feature from the co-occurrence matrix. Conversely, all of FS5's features are from the co- occurrence matrices, except for one; nevertheless, the performances of the two sets are identical.

The feature set with the best classifier performance is FS1. It has a classification error rate of 3.3% on the test images as well as on the training image set. The feature numbers are referenced in Table I, but are explained in more detail here. The features in FS1 are: the valley depth between the two highest peaks of the L-channel

TABLE II.
FEATURE SET PERFORMANCES AT THE LARGEST WINDOW SIZE

| Feature Set | $E_{train}$ (%) | $E_{test}$ (%) |
|---|---|---|
| FS1 = { 3, 5, 21, 26 } | 3.3 | 3.3 |
| FS2 = { 40, 42, 76, 85 } | 6.7 | 10 |
| FS3 = { 57, 74, 81, 104 } | 6.7 | 10 |
| FS4 = { 3, 4, 22, 88 } | 1.7 | 4.4 |
| FS5 = { 21, 35, 57, 101 } | 1.7 | 4.4 |

ODH as a percentage of the highest peak (feature 3); the full width at half maximum (FWHM) of the highest peak in the L-channel ODH (feature 5); the mean of the slopes of the linear fits to the log of fan filter energies (feature 21); the mean intercept of these lines (feature 26). It is evident that those features perform very well and are computationally inexpensive, for they only require computations on the ODH and cortex fan filters.

Conversely, the co-occurrence matrix feature sets e and FS3, with a test set error of 10%, perform about three times worse than FS1 features. It has been observed, that most of the misclassifications were for images of the random class that were assigned to the texture class. This indicates that the co-occurrence features are sensitive to subtle changes, on the pixel level, in random fields and weigh those more heavily than desired.

This idea is supported by the performance improvement of FS5 over FS2 and FS3; here, FS5 includes a single feature not belonging to the co-occurrence matrix features (feature 21 - the mean of the fan filter energy slopes), which helps separate random fields from textures. This is due to textures tending to have higher low-frequency content than random fields, i.e. the fan filter energies have a steeper roll-off (or slope) in the case of textures.

The FS2 features are detailed as follows: the energy at 0 degrees and d=16 (feature 40); the entropy at 0 degrees and d=4 (feature 42); the correlation at 90 degrees and d=16 (feature 76); the homogeneity at 90 degrees and d=1 (feature85). In addition, FS3 features are: energy at 45 degrees and d=1 (feature 57); correlation at 90 degrees and d=4 (feature 74); entropy at 90 degrees and d=1 (feature 81); entropy at 135 degrees and d=16 (feature 104).

The category of mixed features (ODH, cortex filter, and co-occurrence matrix) includes the sets FS4 and FS5. With a test set error of 4.4%, these features outperform the co-occurrence features alone and come close to the performance of FS1. Note that FS4 and FS5 perform equally well, and that all but one feature in FS4 is from the ODH and cortex filter - the opposite is true for FS5.

Contrasting the performance of FS5 with those of FS2 and FS3, it can be concluded that the addition of feature 21 (a cortex filter feature), to the three co-occurrence features in FS5, dramatically improved the classification performance – namely from 6.7% training and 10% testing errors to 1.7% and 4.4% respectively. This indicates that co-occurrence features alone cannot separate textures and random fields well (as applied in this work). The addition of cortex filter slope information (feature 21), helped discriminate between the steeper texture energy roll-off and the flat random field energy signature.

## V. WINDOW SIZE AND CLASSIFIER PERFORMANCE

This section describes the classifier performance as the region size is reduced. Five window sizes were considered: 256 x 256; 128 x 128; 64 x 64; 32 x 32; 16 x 16 pixels. It is expected that zooming into image content will affect the amount of structural content. For example, a line texture such that of a brick wall, looks more like transients as the window size decreases (or resolution increases). Likewise, processing a quasiperiodic texture with a smaller window size will capture less periods of the texture, which favors an assignment to the random class. Also, regions with random content can be seen as textures as the region shrinks due to pixelization introduced by the camera.

To measure the classification performance at various window sizes, the same training and test images as in Section II were used, with the difference that the images were cropped to the desired window sizes. Initial investigation showed that the classifier features that performed well on the 256 x 256 level, do not work well on the higher resolutions. To work around this, it was necessary to find new features for every region size level. This is achieved by repeating the Monte Carlo search for the best features for every resolution level.

The results show that features from the co-occurrence matrix do not significantly improve on features from the ODH and cortex filter. For this reason, only features from the latter feature pool are used. The classifier performances at different window sizes and the best features are shown in Table III.

As the window/region size decreases, the classification problem becomes more difficult. This is evident by the increased number of features required to discriminate between the 3 classes as the region sizes shrink. At the two largest sizes, four features are sufficient to achieve good classification; however, at the smaller region sizes the required number of features increase. As many as eight features do not produce good results. These classifiers are also computationally more expensive. At the same time, the misclassification rate increases rapidly with decreased region size, as illustrated in Fig. 4.

TABLE III.
WINDOW SIZE AND CLASSIFICATION PERFORMANCE

| Window Size (pixels) | $E_{test}$ (%) | Classifier Feature Sets |
|---|---|---|
| 256 x 256 | 3.3 | {3, 5, 21, 26} |
| 128 x 128 | 7.8 | {5, 13, 14, 16} |
| 64 x 64 | 13.3 | {1, 11, 16, 17, 20, 28} |
| 32 x 32 | 18.9 | {6, 13, 17, 18, 21, 24, 26, 27} |
| 16 x 16 | 32.2 | {3, 15, 18, 19, 20, 23, 25, 27} |

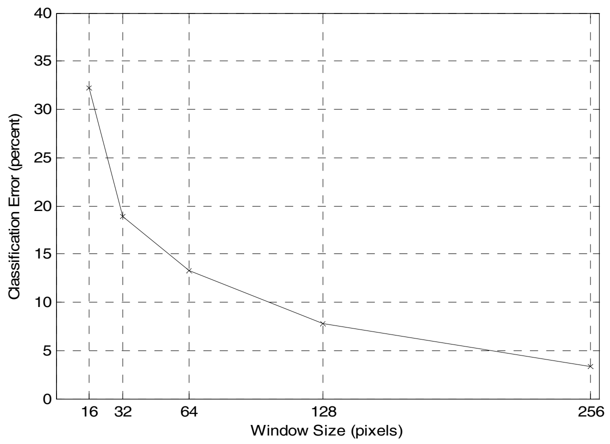


Figure 4. Classifier errors for different region sizes

## VI. Hierarchical Classification Approach

In order to combine these classifiers to segment image content, a hierarchical implementation is presented. It operates on the levels of 256 x 256, 128 x 128, and 64 x 64 pixels, which results in low misclassification errors using only 6 features. Higher resolution classifiers at 32 x 32 and 16 x 16 are not used because they performed poorly.

The implementation combines the classification results from the 3 resolutions by utilizing a classification confidence rule [10]. This rule is derived from the training data sets for each classifier. For each of the classifiers, it is necessary to compute the averages distances, and variances of these distances, between the training image regions and the means of the 3 classes. After normalizing the class scatters by the respective covariance matrices, it is possible to formulate a classification confidence rule.

Analytically, the classification of a feature vector *V* assigned to class $C_i$ is *confident* if the following inequality holds:

$$D_V \leq D_i + \alpha \cdot \sigma_i , \quad (11)$$

where $D_V$ is the distance between the feature vector of the classified image region and the class mean. $D_i$ is the average normalized distance of training feature vectors form the mean of class $C_i$, and $\sigma_i$ is the standard deviation of the distribution of those distances. The factor $\alpha$ is used to vary the upper bound of the confidence interval. For a specific classifier, this rule essentially defines a hypersphere in the feature space at each class center/mean. When a test image region is assigned to a specific class, a corresponding confidence is also computed. The classification is confident if the feature vector (or point) lies within the sphere, and not confident if it lies outside.

Fig. 5 shows a hierarchical classification algorithm, which utilizes the classification confidence rule at each region window size to classify the image content. When a

```
for every image segment of size 256x256 pixels
  => classify the region: random, texture, or transient
  => apply the classification confidence rule
  if the classification was confident at the 256x256 level,
     => report this classification as final for this region
  else

     for every subregion (quadrant) of size 128x128 pixels
       => classify the region: random, texture, or transient
       => apply the classification confidence rule
       if the classification was confident at 128x128,
          => report the classification as final for the subregion
       else

          for every sub region (quadrant) of size 64x64 pixels
            => classify the region: random, texture, or trans.
               => apply the classification confidence rule
            if the classification was confident at 64x64,
               => report this classification for the subregion
            else
                  => do not assign the subregion to any class
            end
          end

       end
     end

  end
end
```

Figure 5. Pseudo code for a hierarchical multi-resolution classification algorithm

region contains characteristics of more than one class, the classification will likely be labeled as not confident. This prompts the algorithm to divide the region into four non-overlapping quadrants, or subregions. These subregions are then classified individually using the next smallest window size classifier. The algorithm presented here operates on 3 window sizes, but can be extended to more.

The procedure starts with the 256 x 256 level, and further divides the window into four 128 x 128 windows if the classification was not confident at the 256 x 256 level. Similarly, if a classification on the 128 x 128 level is not confident, the window is divided into four 64 x 64 subregions, on which the appropriate classifier for this resolution is applied. If at this last level, a classification is not confident, the region is not classified and is assigned to a default fourth class of unclassifiable image content.

## VII. Multi-Resolution Performance

This section illustrates some classification examples using the hierarchical classifier algorithm. Fig.'s 6, 7 and 8 have a similar format. The original image is shown on top and the segmented categories are shown below for random, texture, and transient. A fifth image shows all content classified confidently, where the regions that were not confidently classified are blacked-out. The blacked-out areas in this last image are the areas that the classification algorithm was unable to assign confidently to any of the three classes. All these areas are regions of the smallest window size (64 x 64 in this work) because they cannot be subdivided further.

These examples show relatively good performance. In Fig. 6 the random image shows content that lacks structure, such as the sky, distant dense tree leaves, and incoherent elements of the building that might be blurred. The texture result captured the quasiperiodic elements such as the tree branches, text, and brick wall. Here, some parts of the sky were incorrectly detected as textures. The transient image shows all the structure dominated content, such as light pillars on a dark background, and areas where the building or vegetation meets the sky.

The last image shows the confidently classified portion of the content. It is a relatively large portion and the black-out areas (not classifiable) is small. The unclassified content includes the areas where the dark branches meet the lighter sky. These areas resemble textures and transients at the same time. Similarly, the screened arched windows above the clock also have features of transients and textures.

The image in Fig. 7 shows comparable results. The transient content appears to have been captured well in the transient result image; however, some of the regions classified as random might be better described as transient. It is interesting to note that the darker portions of the basket are correctly classified as texture, where the overexposed portions are assigned to the random class. This might be due to the reduced contrast, and thus weaker textural signatures in the overexposed portions.

Most of the content in the Fig. 8 image is confidently assigned to the 3 classes. The content consists of mostly transients and textures, and few random areas. The output correctly shows that most of the brick work facing the camera is considered transient, while it appears more texture-like when viewed at a wider angle. Also, the brick lined ground is texture-like; however, it is more blurred and random at longer viewing distances.

The performance of the proposed hierarchical classification algorithm is assessed by two metrics; first, the rate of *misclassified* content. It is the area incorrectly classified as a percentage of the total image area. Secondly, the rate of *unclassified* content, which is the percentage of the total image area that could not be confidently classified. The misclassifications were based on disagreements with visual inspection. Overall, the results show that the hierarchical algorithm performs relatively well with a misclassification rate of 4.8%, and an unclassified content rate of 4.1%. In other words, about 95.9% of the image area can be classified with 95.2% accuracy. This amounts to 91.3% of the original test image area being correctly classified.

The classification confidence rule mentioned in Section V allows for some control (using the $\alpha$ parameter in (11)) over the likelihood of dividing a region into smaller subregions for classification on a higher resolution level. Variation of this parameter can slightly affect the algorithm's performance [10].

For example, if $\alpha$ is reduced, classification results are trusted less, which increases the likelihood of subdivision of regions into smaller ones for classification. This will cause more image area to be classified at higher resolutions, where the classifier performance is worse. As a result, more image content is classified at the highest (64 x 64) resolution level, and thus, more content is likely to fall into the "not confident" category. This is detrimental to the classification performance over the image content as a whole.

## VIII. CONCLUSIONS

This work designed linear discriminant image-region classifiers and evaluated their performance using images with complex content. Three image region classes were considered - a random field, texture, and transient. Classifier features were selected using Monte Carlo searches form groups of features from the optical density histogram, the cortex transform, and co-occurrence matrix.

Various window sizes were applied to develop a multi-resolution hierarchical classifier algorithm. The algorithm combined classifier results at several window sizes for overall image content classification by the localized structure of image content. It allowed ambiguous regions to be divided into smaller subregions, which were then classified individually at the new resolution level. A powerful feature of this algorithm is the identification of image content that remained ambiguous and could not be classified confidently at the smallest window size. This is important for image quality metrics in that these regions can be excluded from the metric computations. They may also not impact the correlation with subjective evaluations because the human observer does not look at every region in making a quality judgment, especially complex regions where artifacts are harder to detect.

The content classification algorithm performs well with an overall misclassification rate of 4.8% and unclassifiable content corresponding to 4.1% of the total image area. This amounts to 91.3% of the original test image area being correctly classified. The classification confidence rule the hierarchical algorithm uses is tunable. By varying a single parameter, the likelihood of accepting a classification result at lower resolutions can be controlled, resulting in variable performances and different amounts of unclassified content.

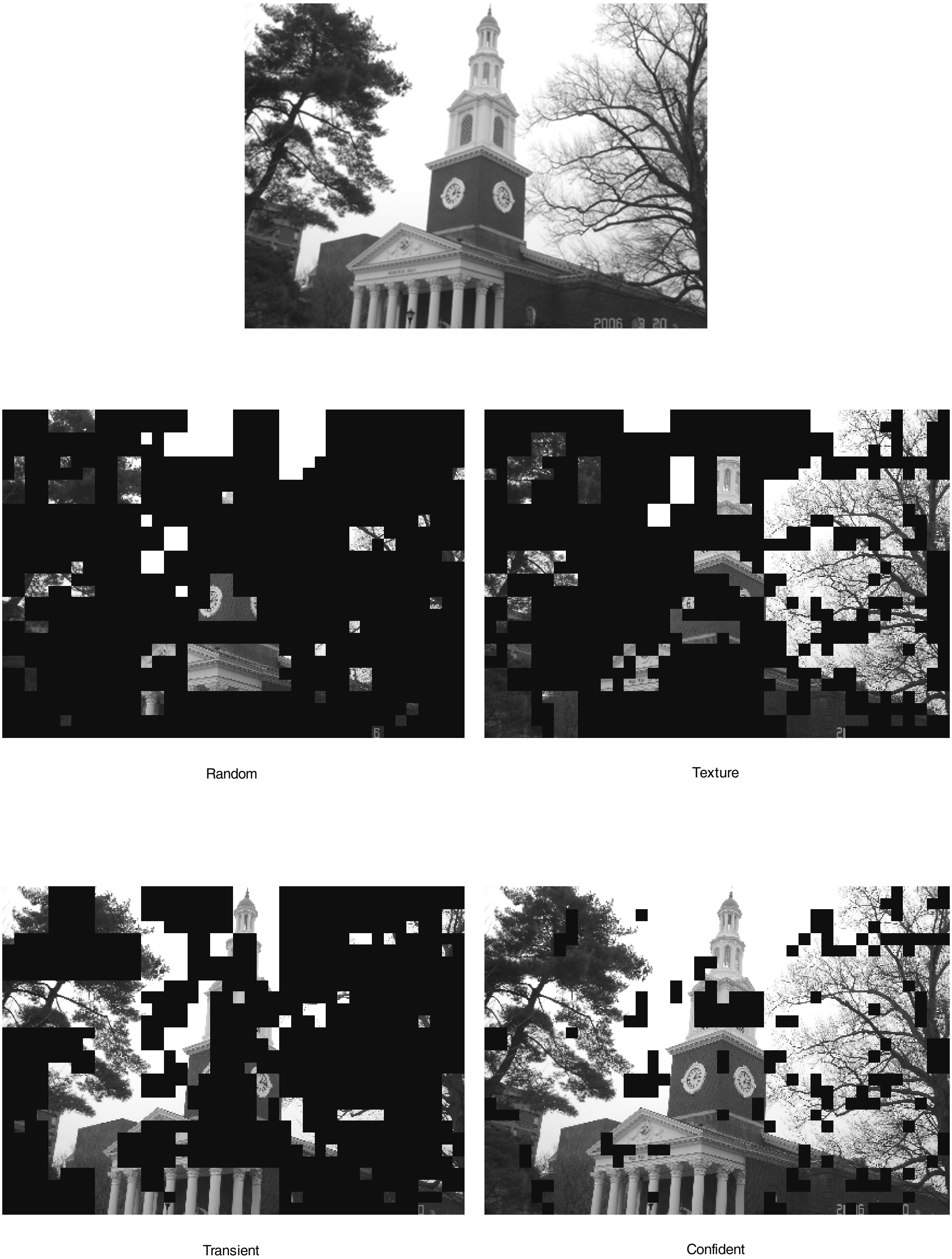


Figure 6. Example hierarchical image segmentation: building

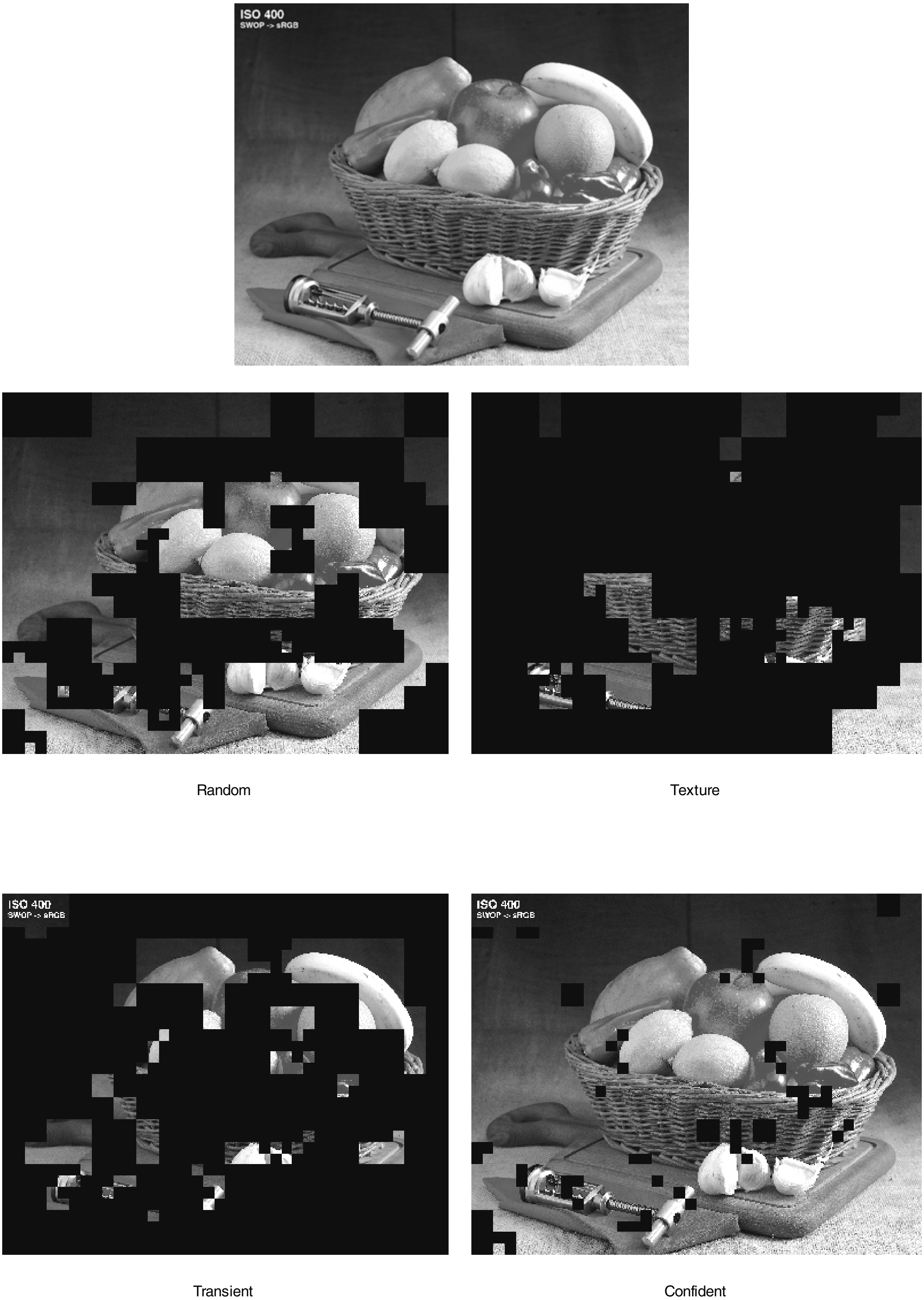


Figure 7. Example hierarchical image segmentation: fruit

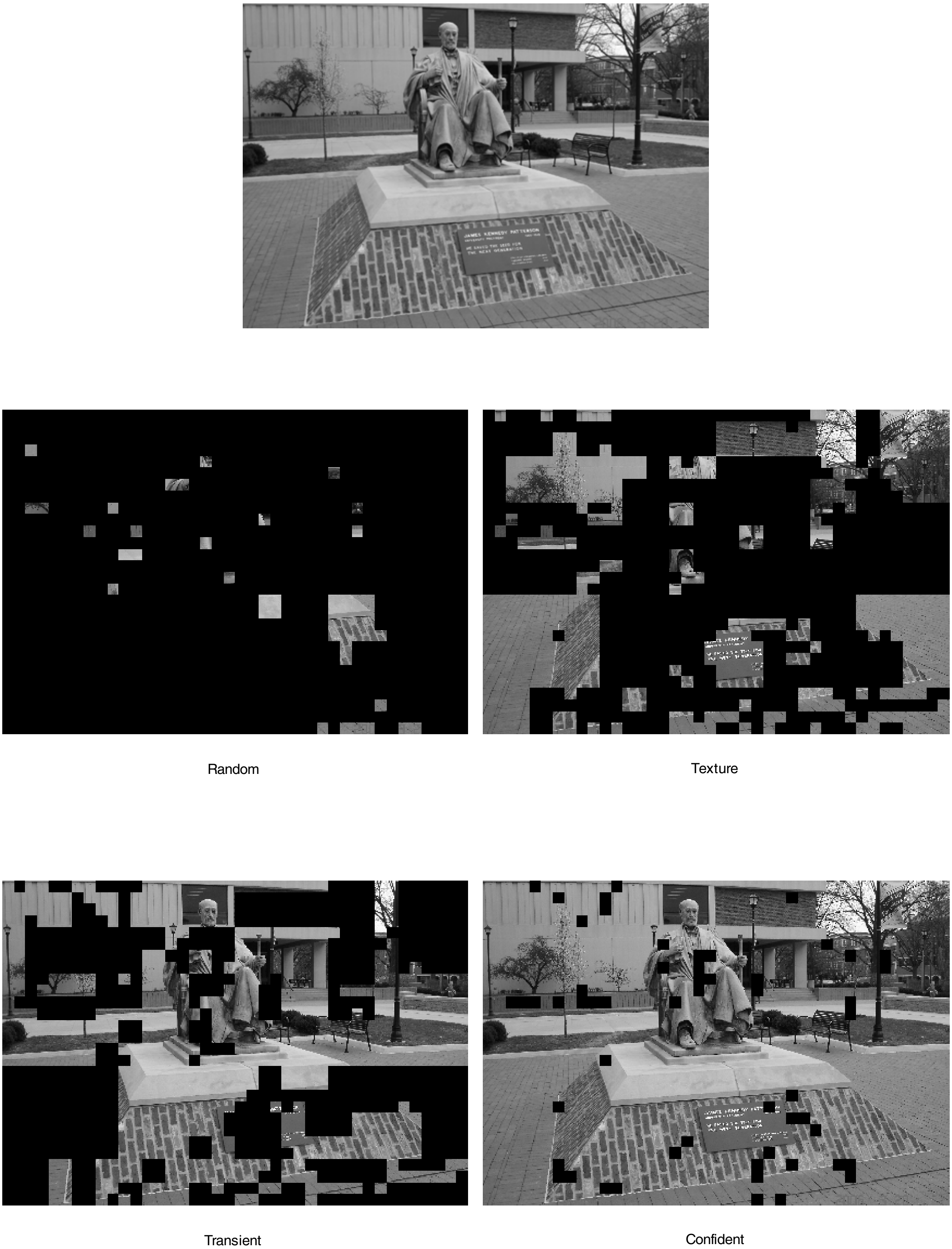


Figure 8. Example hierarchical image segmentation: statue

## REFERENCES


[1] B. Keelan, Handbook of Image Quality Characterization and Prediction, Marcel Dekker, Inc., New York, 2002.

[2] K. MacLennan-Brown and R. E. Jacobson, "Visual significance of digital artifacts," *PICS Conference, Imaging Science and Tech.*, pp. 119-123, Quebec, 2001.

[3] M. Ahmed, B. Cooper, and S. Love, "Content-based document enhancement and resizing," *Proceedings of the IS&T's NIP16 (16th Non-Impact Printing Conference)*, Vancouver, BC, October 2000.

[4] M. Ahmed, B. Cooper, and S. Love, "Document image segmentation using a two-stage neural network," *Proceedings of SPIE* , Vol. 3962, pp. 25-33, January 2000.

[5] Y. Tang, C. Suen, and M. Cheriet, "Document analysis and understanding: a brief survey," *Proceedings of 1'st International Conference on Document Analysis and Recognition*, pp. 17-31, 1991.

[6] L. O'Gorman, "The document spectrum for page layout analysis," IEEE Trans. on Pattern Matching and Machine *Intelligence (PAMI)*, Vol. 15, No. 11, November 1993.

[7] K. Etemad, D. Doermann, and R. Chellapa, "Page segmentation using decision integration and wavelet packets," *Proceedings of the International Conference on Pattern Recognition*, Vol. 2, pp. 345-349, 1994.

[8] A. Jain and Y. Zhong, "Page segmentation using texture analysis," *Pattern Rec.*, vol. 29, no. 5, pp. 343-770, 1996.

[9] H. Cheng, C. Bouman, and J. Allebach, "Multiscale document segmentation," *Proceedings of IS&T 50th Annual Conference*, pp. 417-425, May 1997.

[10] N. A. Rawashdeh, *Characterization of Defects in Images with Complex Content*, Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY, May 2007.

[11] A. Watson, "The cortex transform: rapid computation of simulated neural images." *Computer Vision Graphics and Image Processing 39*, Academic Press, pp. 311-327, 1987.

[12] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," *Digital Images and Human Vision*, MIT Press, Cambridge, pp. 179-206, 1993.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, 1990.

[14] W. Krzanowski, *Principles of Multivariate Analysis*, Oxford University Press, 1988.

[15] G. F. Seber, *Multivariate Observations*, John Wiley & Sons, Inc., 1984.

[16] N. A. Rawashdeh, Kevin D. Donohue, and S. T. Love, "An image quality motivated complex content classifier," *IEEE SoutheastCon 2007*, pp. 514-518, March 2007.

[17] N. Draper and H. Smith, *Applied Regression Analysis*, John Wiley and Sons, Inc., New York, 1998.

[18] R. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC 3, no. 6, pp. 610-621, 1973.


**Nathir A. Rawashdeh** received the Ph.D. and B.S. degrees in electrical engineering from the University of Kentucky in 2007 and 2000 respectively. He earned an M.S. degree in electrical and computer engineering from the University of Massachusetts, Amherst in 2003. He currently works as a Senior Software Engineer in the Color Science and Imaging Department at Lexmark International, Inc. in Lexington, Kentucky. Between 2000 and 2003, he worked for The MathWorks, Inc., in Natick, Massachusetts, as Quality Engineer on embedded DSP programming tools. His research interests include image quality analysis, color science, perceptual modeling, and nondestructive testing. Dr. Rawashdeh is a member of the IEEE, SPIE and ARRL.

**Shaun T. Love** received the Ph.D. degree in computer science from North Carolina State University in 1990. He joined the printing division of IBM, which became Lexmark International, in 1991. He currently manages the Security and Software Research Department at Lexmark in Lexington, Kentucky. Dr. Love is a member of IS&T.

**Kevin D. Donohue** received the B.A. degree in mathematics from Northeastern Illinois University in 1981. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Illinois Institute of Technology in 1983, 1984, and 1987, respectively. From 1987 to 1991 he worked as a Visiting Assistant Professor in the Electrical and Computer Engineering Department at Drexel University, Philadelphia, PA,. From 1991 to present he has been with the Electrical Engineering Department at the University of Kentucky, where he currently holds the Databeam Professorship. He directs the Audio Sensing and Rendering Laboratory in the Center for Visualization and Virtual Environment and co-directs the Signal and Image Processing Laboratory. His current research interests include novel signal characterizations, perceptual modeling, image quality assessment, and audio signal processing. Dr. Donohue is a senior member of the IEEE, a member of IS&T, and Sigma Xi.