Scene classification with respect to image quality measurements

Kyung Hoon Oh, Sophie Triantaphillidou and Ralph E. Jacobson
Imaging Technology Research Group, University of Westminster, HA1 3TP, Harrow, UK

ABSTRACT

Psychophysical image quality assessments have shown that subjective quality depended upon the pictorial content of the test images. This study is concerned with the nature of *scene dependency*, which causes problems in modeling and predicting image quality. This paper focuses on scene classification to resolve this issue and used K-means clustering to classify test scenes. The aim was to classify thirty two original test scenes that were previously used in a psychophysical investigation conducted by the authors, according to their *susceptibility* to sharpness and noisiness. The objective scene classification involved: 1) investigation of various *scene descriptors*, derived to describe properties that influence image quality, and 2) investigation of the degree of correlation between scene descriptors and scene *susceptibility parameters*. Scene descriptors that correlated with scene susceptibility in sharpness and in noisiness are assumed to be useful in the objective scene classification. The work successfully derived three groups of scenes. The findings indicate that there is a potential for tackling the problem of sharpness and noisiness scene susceptibility when modeling image quality. In addition, more extensive investigations of scene descriptors would be required at global and local image levels in order to achieve sufficient accuracy of objective scene classification.

Keywords: scene dependency (scene susceptibility) of image quality, scene classification, scene descriptors (image analysis tools)

1. INTRODUCTION

Image quality can be defined as the overall impression of image excellence and depends upon the pictorial content of the test images [1, 2]. This study is concerned with the nature of scene dependency, which causes problems in modeling and predicting image quality, especially in device dependent image quality measures. This is because objective quality measures tend to perform relatively well on individual average-looking scenes, but they provide lower correlation with subjective assessments when working with non-standard looking scenes.

There are several ways of overcoming the problems caused by scene dependency [3]. One commonly employed is to exclude results obtaining from ‘odd scenes’ in quality measurements. These, however, do not effectively represent the range and variety of different scenes that photographers, artists and consumers may wish to record and reproduce faithfully [3]. Furthermore, scenes that deviate in content from a representative set (e.g. ISO set of test scenes [4]) may not be reproduced appropriately, since they are not in accordance with the ‘average’ reproduction derived from image quality results.

Keelan [5] suggests test scene classification with respect to image quality. The classification he proposes which is based on test scene content and it’s impact of quality attributes, is as follows; a) most susceptible scenes 25%, b) least susceptible scenes 25% and c) intermediately susceptible scenes 50%. In addition, Triantaphillidou *et al* [3] propose test scene classification, using objective *scene descriptors* that correlate with subjective criteria on scene susceptibility to image quality attributes. Scene descriptors are derived to describe basic inherent scene properties that human observers refer to when they judge the quality of images.

The aim of the research describe here was to classify thirty two original test scenes that were previously used in a psychophysical investigation conducted by the authors [2], according to their *susceptibility* (see section 3.1) to sharpness and noisiness. The objective scene classification involved: 1) investigation of various scene descriptors, derived to describe properties that influence image quality, and 2) investigation of the degree of correlation between scene descriptors and scene susceptibility parameters.

2. SCENE DESCRIPTORS

The first step in the objective scene classification was to investigate scene descriptors, derived to describe a number of scene properties. The algorithms deriving these descriptors were implemented in global and local image regions. The reason for local region implementation was that some researchers [6, 7] believe that a local measure of image quality is probably more useful than a global one. A Kadir and Brade's saliency model [8] was applied in MATLAB [9] for this purpose. The implementation involved in following:

- the division of a 20x20 grid on the image
- the calculation of the local entropy in each grid, using a radius from 3 to 70 pixels
- the detection of 30 high in saliency points
- the erosion of the non-saliency areas to amplify the saliency areas

Figure 1 illustrates the saliency process for one test image and presents local regions derived from the saliency model implementation for another four test scenes.

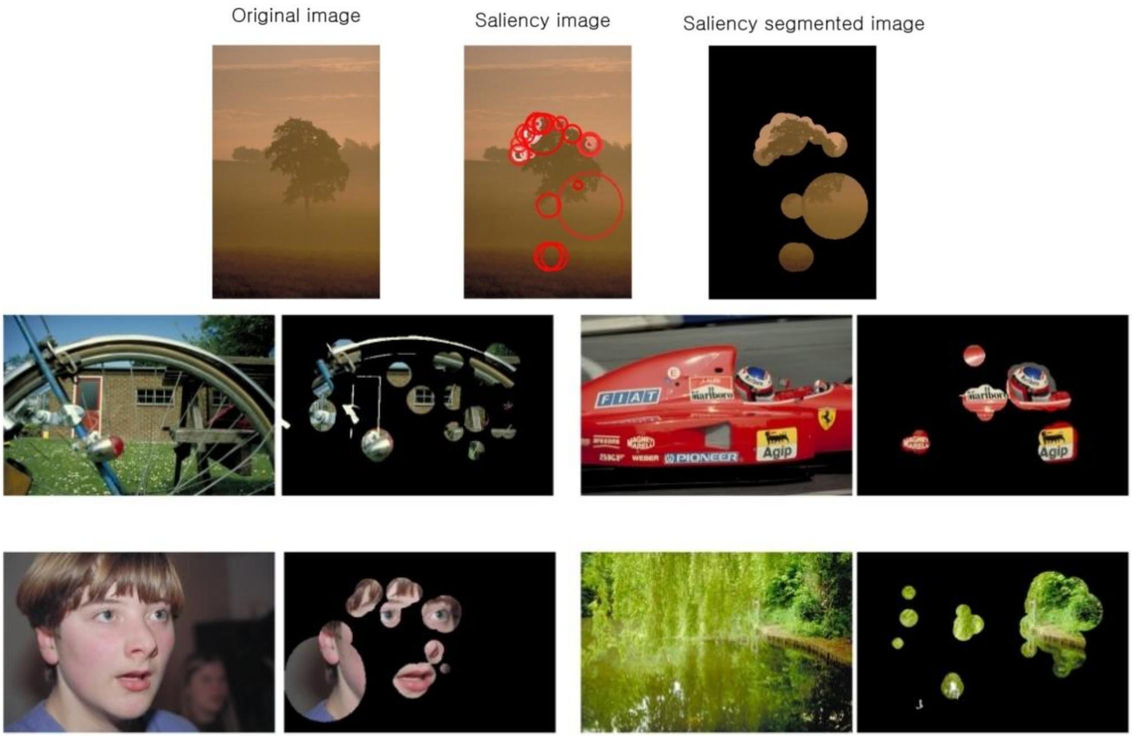

Figure 1. Saliency process for one test image (top row) and local regions for another four test scenes.

A number of scene descriptors were derived using first-order and second-order statistical measures as well as edge detection. Some of these measures (sections 2.1, 2.2 and 2.3) were applied to the *grayscale version* of the image, which was obtained from the 8-bit per channel sRGB image by [9]:

$$\text{Grayscale\_image} = 0.2989R + 0.5870G + 0.1140B.$$

where R, G, B correspond to the pixel value of the R,G and B channels, respectively.

Further, first-order statistical measures were employed to derived measures from the image represented in CIELAB coordinates (section 2.4).

2.1 First-order statistical measures

First-order statistical measures were derived from the Probability Density Functions (PDF) of the grayscale image. The ones investigated in this work are listed below:

- Mean: is the average value in PDF.
- Median: is that value of the middle term of PDF when all the observations are arranged is ascending or descending order.
- Mode: is the value that occurred most often in PDF.
- Variance: is a measure of contrast in PDF, the second power of standard deviation.
- Skewness: is a measure of imbalance of the PDF. We get a value close to zero when the distribution of grey level is balanced (symmetric PDF).
- Entropy: is a measure of information content of the PDF.

2.2 Second-order statistical measures

Second-order statistical measurements, which reveal textural information in images, were calculated from the gray-level co-occurrence matrix (GLCM) [9, 10]. Implementation was carried out in MATLAB [9] using default angle and distance values: 0 and 1 in pixels, respectively. The ones investigated for this work are listed below:

- Inertia (or contrast (Co)):

$$Co = \sum_{i=0} \sum_{j=0} |i-j|^2 P(i,j)$$

- Homogeneity (H):

$$H = \sum_{i=0} \sum_{j=0} \frac{P(i,j)}{1+|i-j|}$$

- Correlation (or linearity (Cor)):

$$Cor = \sum_{i=0} \sum_{j=0} \frac{(i-m_i)(j-m_j)P^2(i,j)}{\sigma_i \sigma_j}$$

- Energy (Ene):

$$Ene = \sum_{i=0} \sum_{j=0} P(i,j)^2$$

Where $P(i.j)$ is the the joint probability distribution of pairs of pixels (i, j). m_i and m_j are the mean values of the pair of gray levels i and j. σ_i and σ_j are the standard deviation value of the pair of gray levels i and j [10].

2.3 Measurement from edge detection

The Sobel, Prewitt and LOG (Laplacian of Gaussian) edge detection algorithms were used to quantify the presence and strength of edges in the grayscale image [11]. The Sobel and Prewitt edge detectors performed using a 3×3 kernel size and 0.04 for sigma [3]. The LOG edge detector was set to a 5×5 kernel size and 0.5 for sigma, which is the default value employed in MATLAB [9]. All edge detectors were operated with the 'replicate' boundary option in MATLAB, where the boundaries were assumed to be equal the nearest border value. During the edge detection, the magnitude of edge (G) was computed by the square-root operation [11]:

$$\mathrm{G} = \sqrt{\mathrm{G}_x^2 + \mathrm{G}_y^2}$$

where G_x and G_y are the horizontal and vertical edge gradients of the image respectively.

Then all individual edge gradients were averaged. Figure 2 illustrates two original images and the corresponding threshold images after Sobel edge detection with the average edge gradient, related to the edges' strength as well as the amount of edge information in the image.

Figure 2. Example of average edge gradient

2.4 Measurement from the CIELAB image

The variance in chroma and saturation were considered as measures of color information. They have been shown to correlate successfully with the perceived image colorfulness [3] and perceived color strength, respectively [12]. The variance in chroma *(VC*)* was calculated [3]:

$$VC^* \cong \sqrt{\sigma_{a*}^2 + \sigma_{b*}^2}$$

In addition, color strength metric *(VS*)*, based on the definition of saturation: $\mathrm{Saturation} = \mathrm{Chroma}/\mathrm{Lighntness}$, derived by [12]:

$$VS^* = VC^*/L^*$$

where the lightness (L^*) is $L^* = L_{mid}^* + \|L_{mid}^* - L_i^*\|$

where L^*_{mid} =50.

3. CORRELATION BETWEEN SCENE DESCRIPTORS AND SCENE SUSCEPTIBILITY PARAMETERS

The second step in the objective scene classification was to investigate the degree of correlation between scene descriptors and scene susceptibility parameters, described in reference [2]. Scene descriptors that successfully correlated with scene susceptibility in sharpness and in noisiness provided means toward the objective scene classification.

3.1. Scene susceptibility parameters

The scene susceptibility parameters were collected from previous experimental work on 'Perceptual image attribute scales derived from overall image quality assessments' [2] (Table 1). They were based on the visual quality loss that occurred to *individual test scenes* with sharpness and noisiness distortions.

	Susceptibility to sharpness	Susceptibility to noisiness		Susceptibility to sharpness	Susceptibility to noisiness
African tree	0.32	1.96	**Baby**	1.06	1.06
Bike	1.21	0.72	**China town**	0.95	0.97
Exercise	1.15	0.51	**Formula**	1.00	1.14
Glasses	0.86	1.17	**Group**	1.07	0.66
Human	0.91	1.10	**Human2**	0.32	1.07
Human3	1.24	0.55	**Human4**	1.08	1.11
Kids	1.18	1.15	**Landscape**	0.86	1.44
Landscape2	0.63	1.82	**Landscape3**	1.05	1.31
London Eye	0.86	1.09	**London Eye2**	0.93	1.28
Louvre	1.16	1.03	**National gallery**	1.07	0.96
Old building	1.29	0.92	**Plant1**	1.15	0.34
Plant2	0.79	0.85	**Plant3**	0.87	1.13
Plant4	0.80	1.12	**Plant5**	1.09	0.99
Plant6	0.97	1.04	**St. Pauls**	1.40	0.50
St. Pauls2	1.10	0.87	**Saules**	1.43	0.19
Sungsil	1.24	0.83	**Yellow flower**	0.91	1.14

Table 1. Subjective scene susceptibility parameters for sharpness and noisiness

A scene susceptibility parameter was identified for each test scene, by calculating the gradient of the straight line connecting average subjective quality ratings (calculated from the entire test-set) and individual quality ratings for the test scene. When the gradient of the line is one, the subjective scale values for the individual scene are the same with these of the combined scenes - for the specific attribute. When the gradient is larger than one the individual scene is more susceptible than the 'average scene' to changes in the specific attribute. The reverse is true when the gradient is smaller than one. An example is shown in Figure 3 for the test scene 'Saules', with gradients for scene susceptibility to noisiness and sharpness equal to 0.1858 and 1.4289 respectively.

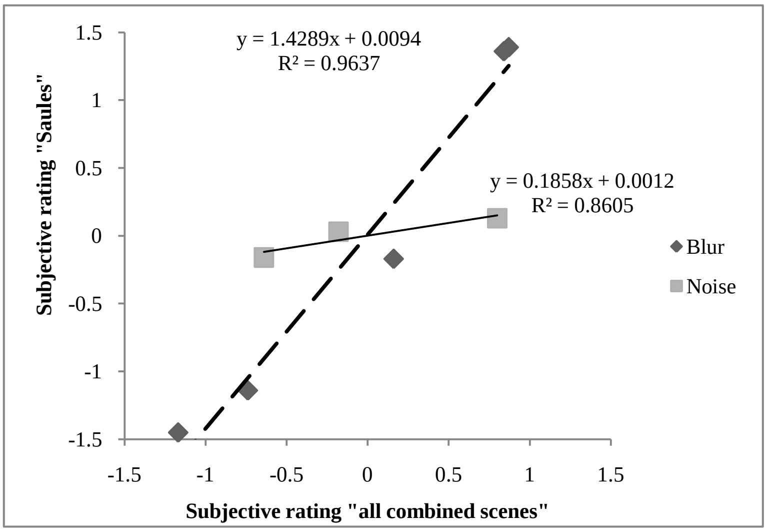

Figure 3. Scene susceptibility parameters for the test scene ″Saules″ (shown in Figure 7)

3.2. Scene descriptors versus susceptibility parameter for noisiness and sharpness

The Spearman's correlation coefficient, r_s, was derived to investigate correlation between scene descriptors and scene susceptibility to noisiness and to sharpness. The coefficient is useful when data have a ranking but no clear numerical interpretation, such as when assessing preferences for data on an ordinal scale [13]. The correlation coefficients range between -1.0 (indicating perfect anti-correlation) and 1.0 (indicating perfect correlation), with 0 denoting no correlation at all.

Successful correlations[1] were obtained between noisiness susceptibility parameters and most second order statistical measures, as well as measures derived from edge detection. Table 2 shows the successful correlation coefficients for noisiness. An example of correlating susceptibility with a scene descriptor is shown in Figure 4.

Scene descriptors	Correlation coefficient (r_s) for scene susceptibility to noisiness	Correlation coefficient (r_s) for scene susceptibility to sharpness
Inertia (Contrast):	-0.694	0.802
Homogeneity	0.738	-0.781
Correlation (Linearity)	0.644	-0.550
Energy	0.577	-0.647
Average Sobel gradient	-0.701	0.786
Average Prewitt gradient	-0.701	0.786
Average LOG gradient	-0.593	0.747

Table 2. Successful correlation coefficients

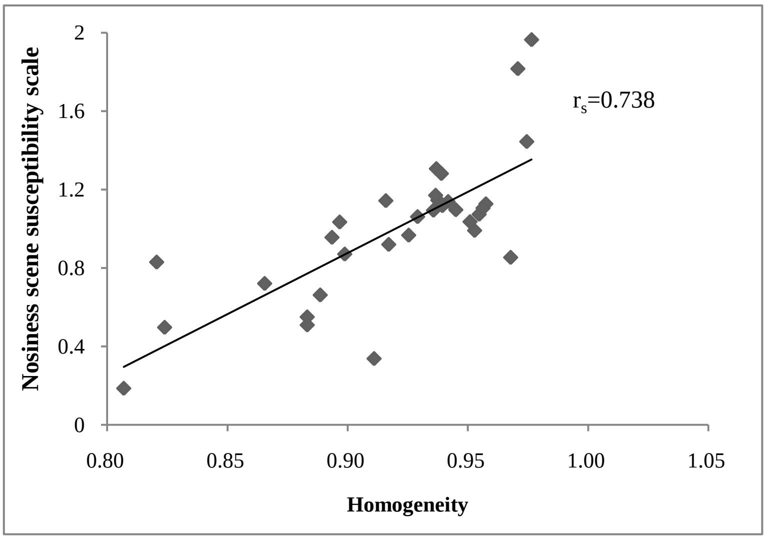

Figure 4. Relationship between the homogeneity descriptor and the susceptibility parameter for noisiness

Successful correlations were also obtained between sharpness susceptibility parameters and, again, most second order statistical measures and measures derived from the edge detection[2]. Table 2 shows also the successful correlation coefficients obtained for sharpness. An example is shown in Figure 5.

[1] When a correlation coefficient is larger than a level of significance at 1% probability level, it indicates statically significant [13].
[2] For both sharpness and noisiness susceptibility predictions, correlations were more successful when the measures were applied at global image level.

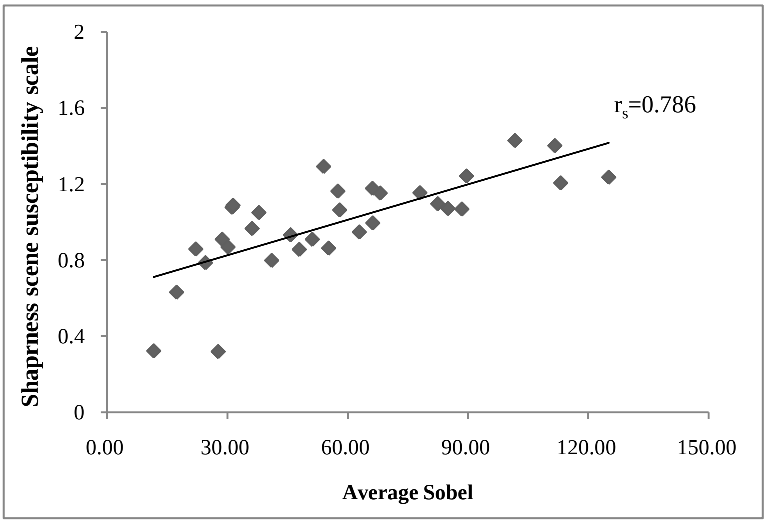

Figure 5. Relationship between the average Sobel descriptor and the susceptibility parameter for sharpness

There were several interesting relationships between scene content and scene susceptibility to noisiness and sharpness. The results confirmed that the higher the texture in the scene content, the lower the susceptibility to noisiness and the higher the susceptibility to sharpness. For example, the correlation coefficients between the homogeneity and scene susceptibility to noisiness and to sharpness were 0.738 and -0.781 respectively. In addition, high presence and strength of edges in the image significantly decreased the perception of noise and increased the susceptibility to sharpness. For example, the correlation coefficients between the average Sobel metric and scene susceptibility to noisiness and to sharpness were -0.701 and 0.786 respectively. It is also evident and confirmed that the relationship between sharpness and noisiness is complimentary i.e. high amount of blur in the image significantly decreased the perception of noise, and high noise decreased perceived blur [2, 7].

Correlations were more significant when the descriptors were derived from the entire image (algorithms were applied globally). Further investigation is required for the derivation of scene descriptors from specific image regions of interest (algorithm application locally). For example, using the central part of the image, as by Keelan and Jin have suggested [7] as a sharpness-critical region and the periphery of the image as a noisiness-critical region. Also, further investigation is required toward the combination of various scene descriptors to derive scene metrics that may describe more successfully the susceptibility of test scenes to noisiness and sharpness.

Overall the results indicated that there is association between selected scene descriptors and scene susceptibility parameters. Thus, the scene descriptors that correlated with sharpness and noisiness scene susceptibility can be used to objectively classify scenes.

4. CLUSTERING FOR NATURAL SCENES

Finally, k-means partitional clustering was implemented to objectively group the 32 test scenes according to their susceptibility to both sharpness and noisiness.

The k-means partitional clustering consists of several steps [14]. The first step of is to define a fixed number of clusters, k. The choice of k is exceedingly important in clustering: an inappropriate choice of k may yield poor results while the correct choice of k is often ambiguous. Possible methods for choosing k include empirical and numerical methods [15]. The empirical method is usually preferred [15]. In relevant image quality investigations k is usually chosen to be equal to 3.0 [5, 16]. Once k is chosen, then modifications of the distances between all points in n^{th} cluster (n varying from 1 to k) and the centre of the cluster are applied. The main idea for modification is that the average distances between all points in cluster and the central point is minimal. During these modifications, new cluster centers are allocated using Euclidean distances. The modification stops when the averages distance from all points in n^{th} cluster and the new central point is minimized.

Two scene descriptors that correlated successfully with both noisiness and sharpness susceptibility, i.e. the *homogeneity* and *average Sobel edge gradient descriptor*, were used for testing the clustering. Clustering was implemented in SPSS programming environment [17]. Figure 6 presents the three clusters with the initial and final centres of the cluster, and then the images corresponding to each of the three clusters (or groups) are shown in Figure 7.

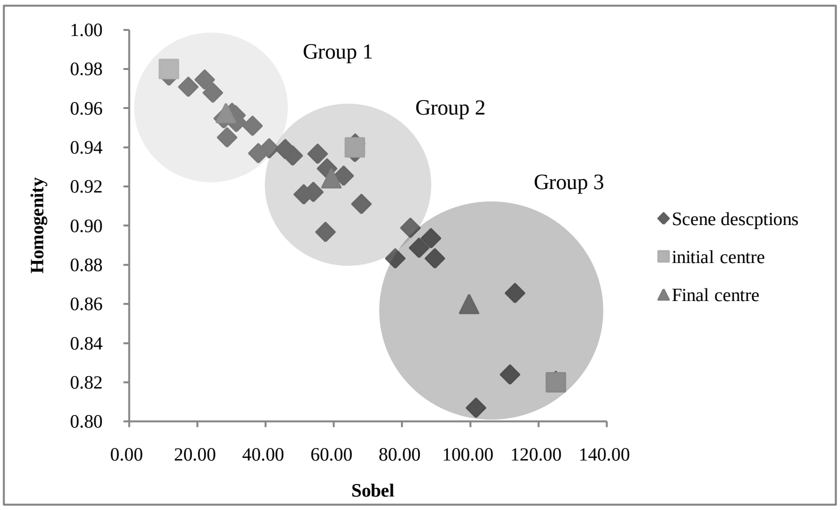

Figure 6. Initial and final centre in three groups

5. CONCULSION

A number of scene descriptors were successfully derived from first-order and second-order statistical measurements as well as edge detection. They were concerned with the extraction of image features, such as brightness, contrast, texture, edges, color contrast etc.

The degree of correlation between scene descriptors and scene susceptibility parameters was investigated using Spearman's correlation coefficient. Successful correlations were obtained between: scene susceptibility parameters for *noisiness* and the *homogeneity descriptor*; and scene susceptibility parameters for *sharpness* and the *average edge gradient descriptors.* These correlations indicated that the selected scene descriptors successfully represented sharpness and noisiness susceptibility and can be used to classify the test scenes used in image quality investigations.

Using the selected scene descriptors and applying K-mean clustering, three groups of scenes were successfully derived, i.e. scenes with: 1) low susceptibility to sharpness distortions and high susceptibility to noisiness 2) average susceptibility to sharpness distortions and noisiness, 3) high susceptibility to sharpness distortions and low susceptibility to noisiness.

The findings indicate that there is a potential for tackling the problem of sharpness and noisiness scene susceptibility when modeling image quality. More extensive investigations of scene descriptors with respect to both global and local image features will help further toward objective scene classification of test scene used in image quality investigations.

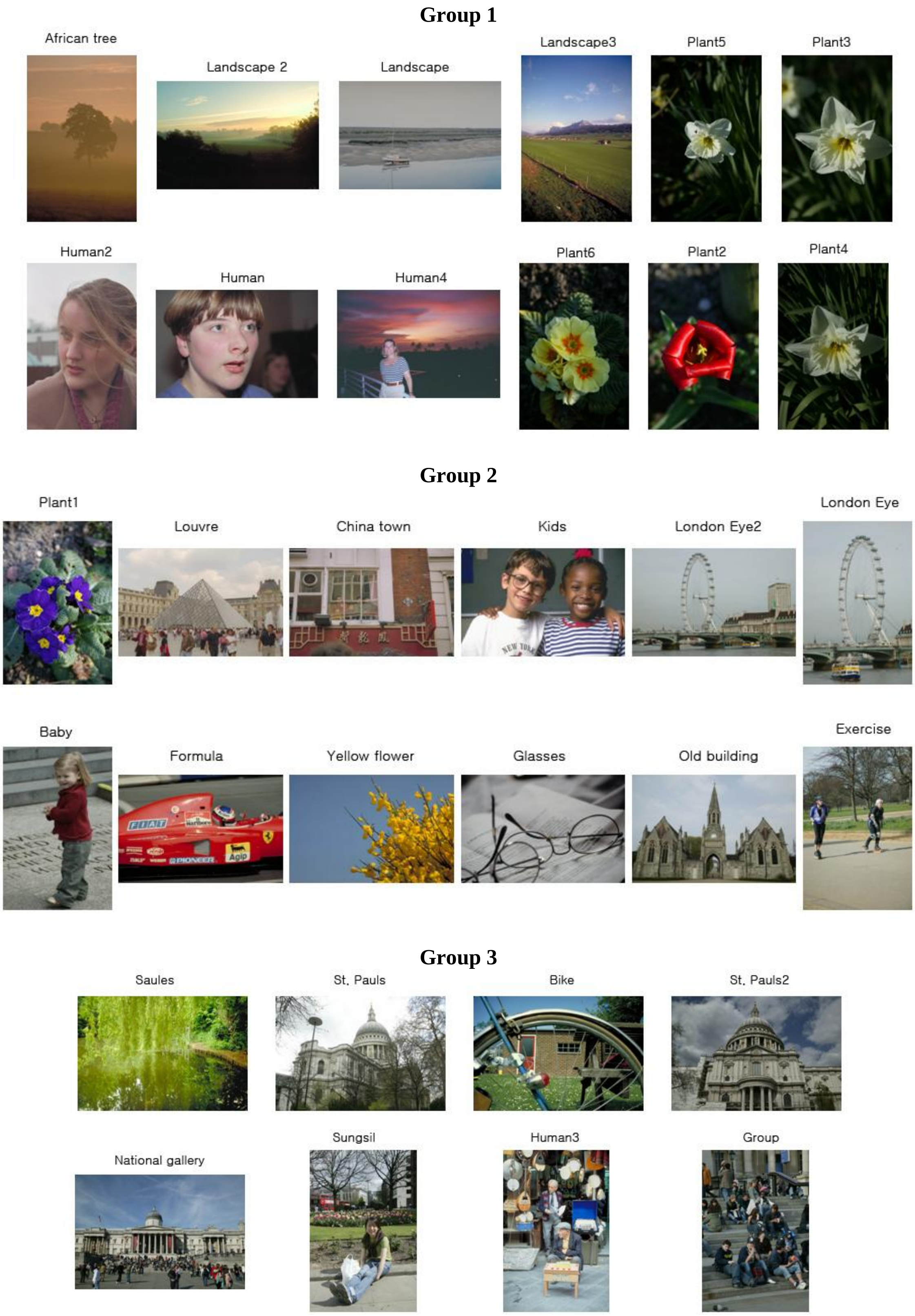

Figure 7. Images in three clusters (groups)

REFERENCES

[1] Keelan, B. [Handbook of Image Quality: Scene and Observer Variability], Marcel Dekker INC, New York, ch 10 (2003).
[2] Oh, K. H, Triantaphillidou, S. and Jacobson, R. E. “Perceptual image attribute scales derived from overall image quality,” Proc: SPIE 7242-11 (2009).
[3] Triantaphillidou, S. Allen, E. and Jacobson, R. E. “Image Quality of JPEG vs JPEG 2000, Part 2: Scene Dependency, Scene Analysis and Classification,” Jornal of Imaging Science and Technology, Vol. 51(3), 259-270 (2007).
[4] ISO 20462-3. [Photography-psychophysical experimental methods for estimating image quality], International Organization : ISO (2005).
[5] Keelan, B. [Handbook of Image Quality: A detailed example of objective metric design], Marcel Dekker INC, New York, ch 14 (2003).
[6] Westen, S. J. P, Lagendijk, R. L. and Biemond, J. “Perceptual image quality based on a multiple channel HVS mode,” Proceedings of the 1995 International Conference on Acoustics, Speech and Signal processing. 4 (1995).
[7] Keelana, B. and Jin, E. W. “Weighting of field heights for sharpness and noisiness,” Proc: SPIE 7242 (2009).
[8] Kadir, T. and Brady, M. “Scale, Saliency and Image Description,” International Journal of Computer Vision. Vol 45(2), 83-105 (2001).
[9] MathWorks. [MATLAB 2007a], The MathWorks Inc (2007).
[10] Pratt, W. K. [Digital image processing: Image feature extraction], John Wiley & Sons, New York & Chichester, 557-596 (1991).
[11] Gonzalez, R. C. and Woods, R. E. [Digital Image Processing: Image segmentation], Prentice Hall, London, ch 10 (2002).
[12] Othman, A. and Martinez, K. “Colour appearance descriptors for image browsing and retrieval,” Proc: SPIE 6820 (2008).
[13] Siegel, F. A. [Statistics and data analysis], John Wiley & Sons, New York & Chichester (1996).
[14] Yendrikhovskij, S. [Colour Image Science: Image Quality and Colour Categorisation], John Wiley & Sons. London, ch.16 (2002).
[15] Catherine, A. S. and Gareth, M. J. "Finding the number of clusters in a data set: An information theoretic approach," Journal of the American Statistical Association. Vol 98 (January), 750–763 (2003).
[16] Keelan, B. [Handbook of Image Quality: Monte Carlo Simulation of System Performance], Marcel Dekker INC, New York, ch 28 (2003).
[17] SPSS, [SPSS 14.0], SPSS Inc (2005).