Multidimensional image selection and classification system based on visual feature extraction and scaling

Francesco Mancusi, Sophie Triantaphillidou and Elizabeth Allen
University of Westminster, Watford Road, HA1 3TP, London, UK

ABSTRACT

Sorting and searching operations used for the selection of test images strongly affect the results of image quality investigations and require a high level of versatility. This paper describes the way that inherent image properties, which are known to have a visual impact on the observer, can be used to provide support and an innovative answer to image selection and classification. The selected image properties are intended to be comprehensive and to correlate with our perception. Results from this work aim to lead to the definition of a set of universal scales of perceived image properties that are relevant to image quality assessments.

The initial prototype built towards these objectives relies on global analysis of low-level image features. A multidimensional system is built, based upon the global image features of: *lightness, contrast, colorfulness, color contrast, dominant hue(s)* and *busyness*. The resulting feature metric values are compared against outcomes from relevant psychophysical investigations to evaluate the success of the employed algorithms in deriving image features that affect the perceived impression of the images.

Keywords: Image analysis, feature extraction, psychophysical scaling, similarity, image quality.

1. INTRODUCTION

This project aims to provide an innovative solution to the increasing need for versatility in the methods by which images are accessed, when considering the rapid growth of the amount of data related to digital images. The main purpose of the sorting and searching operations proposed in this work is to assist in the selection of test images employed in psychophysical investigations related to image quality, since their results are strongly affected by the identification of these test sets[1-4].

This project's methods go beyond traditional text-based systems of "tags", instead organizing images according to inherent image properties that are known to have visual impact on the observer. The resulting software application focuses on the extraction of image information via the quantification of chosen image features. This quantification aims to lead to the definition of a set of universal feature scales that are relevant to image quality assessments. The scaled image features are intended to be comprehensive and to correlate with our perception of images.

Scaled values for each image feature (i.e. feature metrics) are integrated with traditional text descriptors to retrieve images that satisfy specific criteria. This allows, for example, grouping of images with "overall lightness" within a certain range, and/or a specific level of "chromatic information" and/or a chosen "dominant hue". The development of a platform that uses the images' features as browsing variables requires careful consideration with respect to i) the ability of the selected algorithms to produce metrics that match the visual perception of the chosen image features and ii) the weighting that each feature has on the global visual impact of the particular image.

The initial prototype built toward these objectives relies on a multidimensional system that organizes images according to *global lightness, contrast, colorfulness, color contrast, dominant hue(s)* and *busyness*. Each of these features represents a dimension of scaled 0-100 values, subsequently grouped into five categories for easier analysis. Further, provided that the features' scales are calibrated to linearly relate to perception, the Euclidian distance between images (represented as points in this multidimensional space) can be used as an indicator of image similarity.

The extraction of the feature metric values is achieved by employing existing as well as novel algorithms, carried out in a customized space derived from CIELAB color encoding and involving image statistics and segmentation operations. The resulting metric values are compared against outcomes derived from relevant psychophysical investigations to evaluate the success of the algorithms in representing the perceived impressions of images. Such psychophysical investigations aim to develop subjective interval scales for the chosen image features (i.e. perceived image lightness, contrast, etc). The tests are conducted on an internet-based psychophysical display. They involve categorical scaling operations, where observers are asked to place each image sample in one of five categories, according to 'how much' of the specific feature is present in the image.

The multi-dimensional structure prototype displays results for selected scales on a web-based application. The production of scales representing the structure dimensions, which are intended to be calibrated i.e. uniform and linear with perception, may further facilitate the inclusion of 'image quality', or individual quality attributes, such as 'sharpness', 'noisiness', etc. as extra dimensions in the structure.

2. VISUAL FEATURES

The processes that are applied at this first stage are intended to analyze images globally and to extract low-level features, with attention to the visual and structural elements within the image, such as, for example, statistical aspects of the scene representation.

Global image analysis is carried out to derive intensity, chromatic and busyness related information from scenes. Intensity information is retrieved from the CIELAB $L_{ab}*$ channel and leads to the quantification of global lightness and contrast. Color related information is retrieved from the CIELAB $C*$, $a*$ and $b*$ channels, which provide the basis for calculating colorfulness, color contrast and dominant hues. Finally, a descriptor of lines, edges and high frequency information within scenes is derived, since this information is known to play a fundamental role in the observers' judgments of image quality,[4,5]. The busyness metric that was used for the purpose, which separates 'flat' from 'busy' image areas, is described in detail in reference 3. In summary, the multidimensional system is based on the following six image features (and thus has, in its current form, six dimensions):

- Lightness information: (1) global lightness and (2) lightness contrast;
- Chromatic information: (3) colorfulness, (4) color contrast, (5) dominant hue;
- Amount of detail: (6) busyness.

Figure 1 represents the decision making process that was followed for the determination of the algorithms used for feature extraction:

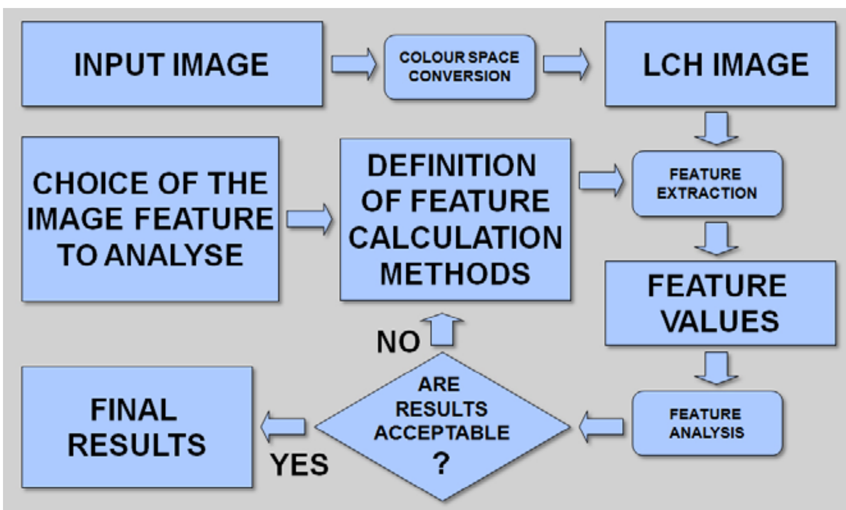

Figure 1. Decision making process used for the determination of the feature extraction algorithms.

2.1 The LCH space

LCH is a space of three dimensions, built to make image calculations more efficient. It is derived from the L_{ab^*}, a^* and b^* color coordinates of the 1976 CIELAB space which is chosen over the RGB space because of its (near) uniformity with human perception[6,7]. The space's C and H dimensions are derived from the CIELAB chroma, C_{ab}^* and from the hue angle h_{ab}^*.

In order to retrieve L, C and H space coordinates, L_{ab}^*, C_{ab}^* and h_{ab}^* are remapped into a [0, 100] range. This choice offers a number of technical advantages, such as carrying out the analysis on the more intuitive descriptor of "percentages". The choice of the limits in the image function co-domain was also taken considering the grouping of the [0-100] values for each individual dimension into five categories (i.e. very low, low, medium, high and very high). These are aimed to represent the 'quantity' of each image feature, providing a flexible solution.

Percentages and categories form a solid tree-structure that leads to the representation of the image in a more computationally efficient way. Further, it is in fact easier for users and observers to group image features using categories than to give precise scale values for each feature[1]. The latter are nevertheless still retrievable in the multidimensional system from the lower plane of image analysis.

2.2 Algorithm implementation

A number of algorithms were implemented to derive feature metric values for the chosen image features. At this first stage of the implementation, the resulting metric values lay on non-calibrated axes. Thus, the magnitudes of the metric values are not related to perceptual magnitudes. They only provide limited information about 'how much' of an image feature is present in each image and thus comparison between images can only be relative. To linearize the metric scales for each dimension of the multidimensional space and provide meaningful metric value differences in terms of JNDs, a number of psychophysical studies need to be conducted. Section 4 describes the first stage of the related investigations.

The algorithms used for the derivation of feature metrics were applied either directly on the pixel values of the image represented in LCH space or their distribution:

- Feature metrics calculated statistically from the distribution of pixel values:

  Lightness, *Contrast*, *Colorfulness*, *Color Contrast*

- Feature metrics calculated as a result of dynamic consideration of probability distribution functions (PDFs):

  Dominant Hues (up to three)

- Feature metrics calculated from segmentation processes applied on the image's pixel values:

  Busyness

Both the mean and the mean and median of the L channel can provide a relevant statistical measure of global image lightness[8,9]. However, it was observed that different pixel value distributions responded differently to these measures, giving in a number of cases misleading results. An example is illustrated in Figure 2. A particularly high-contrast image has in fact an exceedingly low median. On the other hand, it has a mean that falls into the middle range of lightnesses and is not representative when the perceived lightness of the image is compared to the lightness of the entire image set. Had the median been used as a measure of the global lightness the left image in Figure 2, the image would have been rated as too dark. Had the mean been used instead, the image's global lightness would have been found too high. The solution to this is to derive the global image lightness by averaging the results from the two statistical measures:

$$\text{Lightness: } LI = \mu\{\mu_L; med_L\} \tag{1}$$

Where μ_L is the mean and med_L is the median of the L values in the image.

Figure 2. Image "Gilberto" (left) showing how neither the mean nor the median are representative metrics for the global lightness.

The lightness contrast is derived by taking the standard deviation of the L channel, which produces results that matched initial observations. The output values are retrieved after setting the overflow cases to [100], in order to correctly fall into output range. Similarly, the color contrast is derived by calculating the standard deviation of the C channel.

$$\text{Contrast: } LC = 10 \times \sigma_L \tag{2}$$

$$\text{Color contrast: } CC = 10 \times \sigma_C \tag{3}$$

Where σ_L is the standard deviation of the L values and σ_C is the standard deviation of the C values in the image.

Similarly to global lightness, the definition of the global image colorfulness is related to both the mean and the median of the C channel, but also includes color contrast information:

$$\text{Colorfulness: } CI = \mu\{\mu_C; med_C; 10 \times \sigma_C\} \tag{4}$$

Where μ_C is the mean of the C values, med_C is the median of the C values and σ_C the standard deviation of the C values in the image.

The definition of the dominant hue cannot be obtained with standard statistical operations, since the H dimension is expressed in radiant notation and its values have no "direction". The identification of the dominant hue in the image, if there is one, needs a different approach. For this purpose, a customized process is used, based on a dynamic analysis of the probability distribution function of hue angles.

Starting from the assumption that the dominant hue corresponds to the most frequent one, the problem can be shifted to correctly define the "hue". As long as different shades of equally, or similarly perceived colors correspond to different hue angles the dominant hue cannot be represented by a single hue angle. Instead dominant hues are defined using two parameters: the dominant hue angle, represented by the mode of the distribution, and the width of the distribution around this dominant hue:

$$\text{Hue value: } V_H = mod_H \text{ with } P(V_H) > P_{MIN} \tag{5}$$

$$\text{Hue width: } W_H = [V_H\text{-}h,\ V_H{+}h] \text{ with h| h min } \mu(P(V_i)) > P_h \text{ and } V_i \in W_H \tag{6}$$

Where mod_H is the mode of the H-distribution, $P(V_H)$ and $P(V_i)$ are the probabilities of the hue value V_H and V_i.

From this description, a parameter I_H of the visual importance of that particular hue is retrieved from the cumulative probability of the hue angles belonging to the specific hue range.

$$\text{Hue visual importance: } I_H = \sum_{i=V_H-h}^{V_H+h} (P(V_i)) \tag{7}$$

Where mod_H is the mode of the H-distribution and $P(V_i)$ is the probability of the hue value V_i.

To define acceptable values for V_H and W_H two constant thresholds are used. P_h represents the minimum mean value of the probabilities in the W_H range and it is used to derive the value of h. P_{MIN} is referred to the minimum probability needed in order to consider the mode of the hue PDF as a possible dominant.

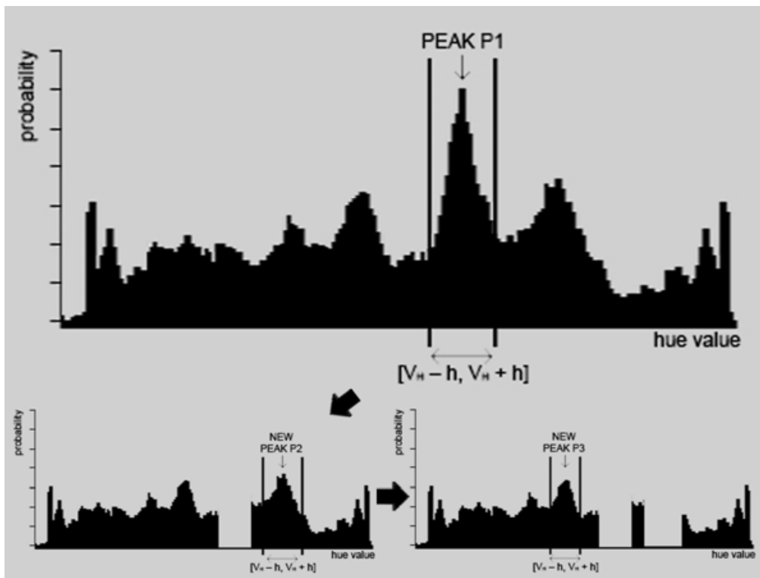

Figure 3. Sequential steps used for the determination of the dominant hue of the image.

The dominant hue in the image is then the most visually important hue derived from Equation 7. Further improvements on this algorithm could take into account the way such hues are spatially distributed, considering for example, if pixels corresponding to specific hues form blocks of uniform color, or if they are mixed together with pixels of different hues - with consequently different visual impact. When using this algorithm dithered images can in fact produce incompatible results with what is visually perceived. For each image, up to three dominant hues can be derived:

- Dominant hue: most visually important hue in the image
- 2nd hue = second most visually important hue in the image
- 3rd hue = third most visually important hue in the image

The last extracted feature is related to the amount of high detail areas in the image in relation to the total image area. This image characteristic is referred to as busyness and plays an important role in terms of our perception of image content. To quantify such feature, the chosen approach is image segmentation, as described by Triantaphillidou *et al.*[3]

Their method is applied to the L channel and follows various steps to separate busy from non-busy areas in the image, depending on certain thresholds. The process involves image blurring and the application of a binary gradient mask to identify edges, the dilatation of such edges and uses a filling operation to fill the ‘holes’ between closely identified edges, since they too are perceived as part of the image detail. Finally erosion is used to remove spurious information. The value *b*, expressed as a percentage, is the metric of scene busyness. It is calculated from the ratio of the number of pixels that correspond to busy image areas to the total number of pixels.

$$\text{Busyness: } b = 100 \times \left(A_{BUSY} / A_{IMAGE} \right) \tag{8}$$

Where A_{BUSY} is the size of the busy part of the image and A_{IMAGE} is the total size of the image.

3. MULTIDIMENSIONAL STRUCTURE

A “Business Intelligence” (BI) structure was employed to withhold and represent the data. The term Business Intelligence, defined in 1989 by Howard Dresner[10] from the Gartner Group, indicates all the processes that are aimed toward the extraction of valuable information from a collection of raw data and to the following analysis of its meanings.

The process of Business Intelligence can be explained through four stages[11]: *sources* (i.e. images)*, storage area, performance* and *access* to the information. The storage area is implemented using a *Data Warehouse* (DW), defined by Inmon[12] as a collection of data that is: i) aimed to give information on a particular subject, ii) retrieved by various sources and merged into a coherent whole, iii) meant to identify a particular time period, iv) stable (data can be integrated but not modified). The DW is implemented using a relational database and based on two dimensional tables[13,14]. The *performance* stage consists of an OLAP AREA, called *Cube*, which is a multidimensional structure of data.

The three operations that allow moving from one stage to the other are:

- **ETL**: extraction, transformation and loading of the data from the sources into the DW.
- **OLAP:** on-line analysis processing of the data in the DW, aimed to build the *Cube*.
- **REPORTING TOOLS:** used to visualize valuable information from the *Cube* (i.e. website)

Central part of this structure resides in the processes of turning the two dimensional table-based content of the *Data Warehouse* into the multidimensional dressing of the *Cube*. To perform this operation, the tables in the DW are separated into *Dimensions* and *Facts*. *Dimensions* contain the descriptors of the features they represent and are represented in Cartesian axes. Points on *Dimensions* define the categories that can be assumed and are called *members*. Depending on the density of its members, a dimension has a certain *granularity*. It is also possible to group members into macro-categories and use them to define hierarchies. On the other hand, *Facts* are the numerical values that can be aggregated together and define the co-domain of the Cartesian space.

3.1 Dimensions

To store the image data, each feature is represented by a dimension of one hundred members that range from 0 to 100. These are further split into five categories (with the exception of “Hue” which is split to nine) for hierarchical purposes. The definition of the category values and objective category boundaries for each feature metric was performed by initial empirical observations, i.e by identifying individual images in the test set that corresponded to a turning point in visual impact for that feature.

The aim for the psychovisual experiment discussed in Section 4 is to calculate feature scaled values for each feature and to identify the perceptual category boundaries that can be used further for the definition of the categories that classify the feature metric values. The categories used to group the values related to hue angles, were not defined empirically, but were derived from a studies on color gamut mapping made by Braun and Fairchild[15,16].

3.2 Facts

The facts of the multidimensional structure are function values retrievable from a set of coordinates in the Cartesian space. These are therefore numerical values that have to be presented with a proper aggregation function, since not all conditions in the coordinates might be specified. The BI structure supports the use of more than one fact in the structure, but the only one taken into account is the *number of images* that satisfy that condition.

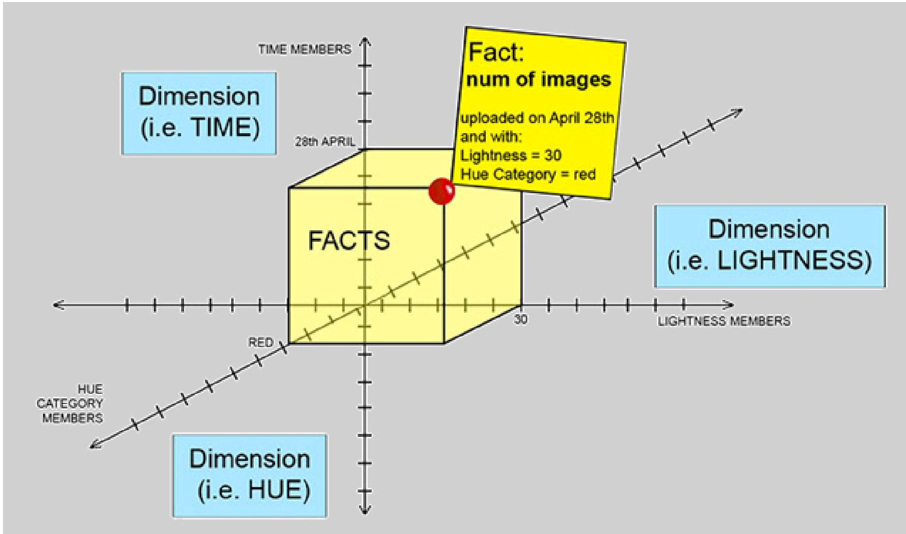

Figure 4. Multidimensional data structure of the CUBE with respect to dimensions and facts.

3.3 Similarity

According to this structure, the quantification of the image features leads to the definition of a set of coordinates that place the images as a point into the Cartesian space. From the relative position of all points, it is possible to retrieve an index for closeness. The Euclidean distance, can represent an index for image similarity, provided that the axes are linear with perception and equally scaled, a condition which, at the current stage of the project, is not true. In other words, the condition for the similarity metric, S in equation 9, to predict with some success the perceived image similarity is that one unit of distance in one dimension is the same distance as in any other dimension, and ideally equal to 1 JND.

$$\text{Similarity } S_{12} = 100 - \frac{\sqrt{\left(\Sigma_{\forall D}\left(D_{i1} - D_{i2}\right)^2\right)}}{\sqrt{D_{\#}}} \tag{9}$$

Where $D_{\#}$ is the number of *Dimensions* and D_{i1} is the value assumed for the ith *Dimension* in image 1.

Equation 9 is used to calculate similarity, S, between two images IM1 and IM2, where $D_{\#}$ is the number of dimensions and D_{i1} represents the value derived for the ith dimension of IM1. Note that, for the hue angle dimension the smallest absolute difference is taken as the index of similarity, because hues are expressed in radians.

Further to the condition described above, the index for similarity will have to account for the visual weighting that each feature (and thus each dimension in the multidimensional space) has on the image. A number of psychophysical studies need to be carried out to define scales and weightings for each dimension. The next section describes the first stage of such investigations.

4. PSYCHOVISUAL TEST

The psychovisual test, built to determine initial interval scales for each of the visual features, was developed for an internet-based user interface. The perception of the image features of global lightness, contrast, colorfulness, color contrast and busyness is rated by observer by choosing amongst five categories, ranging from very low to very high. To evaluate the perception of the dominant hue, the observers are given nine categories from which to choose. These include six categories for the three primaries and the three secondary colors, one category for Caucasian skin tones and one category for neutrals. Lastly, a choice for more than two dominant hues is given in case the observer does not clearly perceive one or two hues that dominate in the image. This last choice is necessary to identify images that have a particularly spread color distribution.

The psychophysical process involves the evaluation of 38 scenes, with a variety of content and feature information, presented randomly to the observers. Three preliminary steps in the test include: an introduction; a set of questions to the observers that retrieve personal information as well as information on his/her computer station and display system; a test on the quality of the perceived test stimuli on their display in terms of brightness[17] (via specification of the black and the white point on the screen), gamma[18] and hue shift[17]. Observers are provided with a clear definition of each image feature which are asked to rate, and a set of instructions on how to conduct the test.

The psychophysical test involves a first step aimed to evaluate the perception of the lightness-based features (lightness, contrast, busyness), and a second one based on the perception of the color-based features (colorfulness, color contrast, dominant hue). Finally, observers are given the choice to submit comments and suggestions for further improvements.

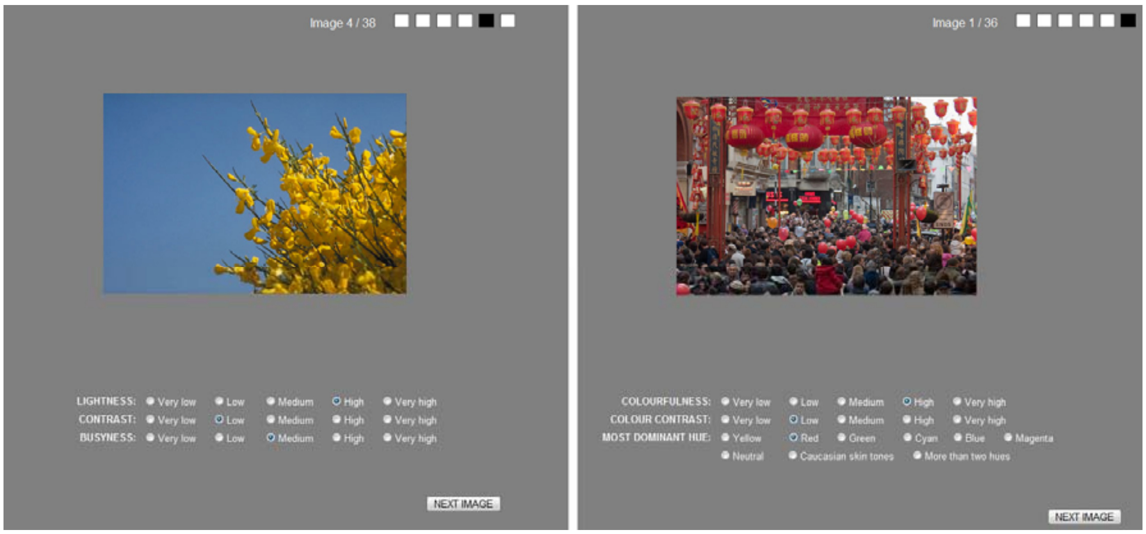

Figure 5. Psychovisual displays of the L-based image features (left) and of the C-based image features (right).

5. RESULTS

Although observers are invited to consider each point in the relative ‘feature scale’ as equally distant from its adjacent point, in the analysis of results obtained from the subjective investigations the data are treated as if perceptual distances between categories are different. Category boundaries along with perceptual sample scale values have to be calculated for creating an interval scale for each image feature[1,19]. The reason for the unequal perceptual differences between categories is that the distribution of the sample judgments along the category scale does not necessarily have equal dispersion. According to the Thurstone’s Law of Comparative Judgment, the difference between the sample scale value and the category boundaries is given by[1,19]:

$$t_g - S_j = z_{jg}\sqrt{\sigma_j^{\,2} + \sigma_g^{\,2} - 2\rho_{gj}\sigma_g\sigma_j} \tag{10}$$

Where t is the g^{th} category boundary with $g \in [1, m]$; S_j is the j^{th} sample with $j \in [1, n]$; z is the vector of z-values; σ is the standard deviation; ρ is the correlation coefficient.

Under the condition D of the Four Conditions of the Law of Categorical Judgment, the standard deviation and the correlation coefficient are constant and independent from the category, or from the sample. Thus, it is possible to simplify Equation 10 and obtain[1,19]:

$$t_g - S_j = z_{jg}\sqrt{1} = z_{jg} \tag{11}$$

Figure 6 illustrates the range of the subjective scale values and category boundaries of the interval scales derived for each visual feature. The dominant hue categories are otherwise retrieved using a different approach, since perceived hues do not follow a monotonic, directional scale. To this purpose, hue related results will be analyzed separately, by considering the dispersions of the clusters around selected dominant hues.

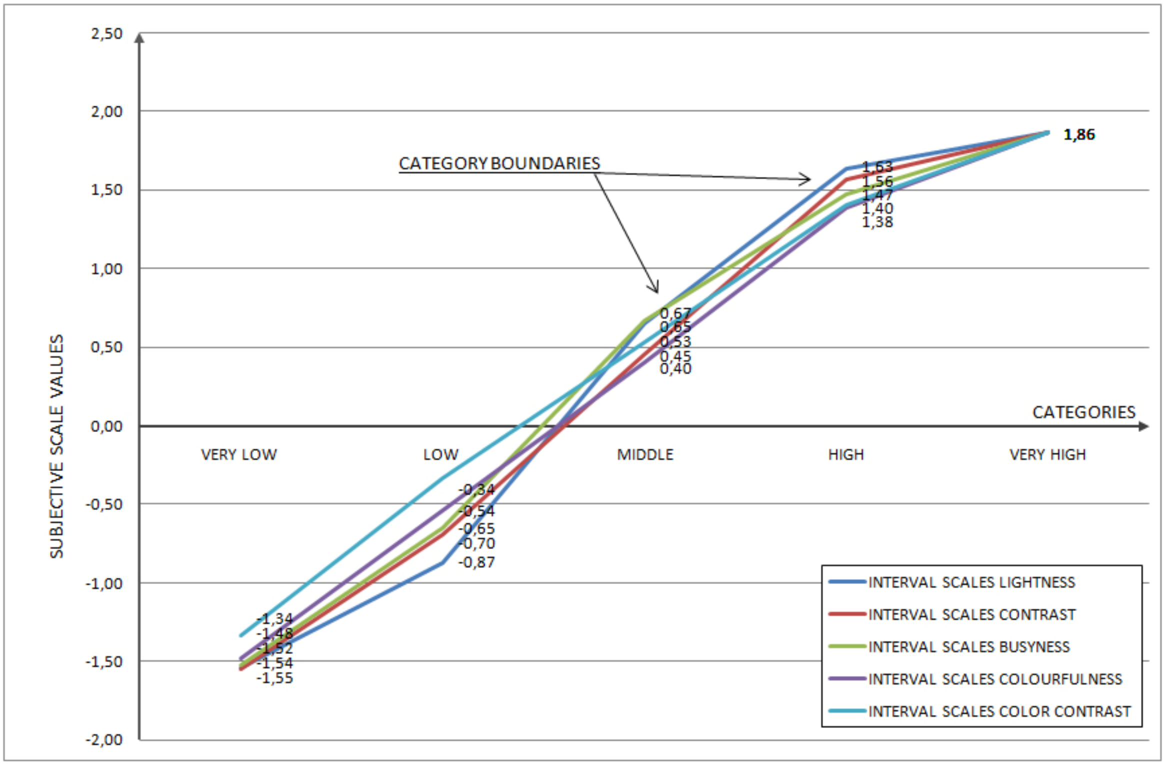

Figure 6. Interval scales retrieved for each of the image features.

6. CONCLUSIONS

This paper discusses the development of a multidimensional image selection and classification system that is based on feature extraction, scaling and weighing of selected image features that that are relevant to image quality assessments. The work toward this project is ongoing and the results presented at this stage are still preliminary. So far initial subjective scales for a number of image features have been derived. Further experimentations will involve the compilation of many more observations and the use of different image sets for creating universal, calibrated scales that represent reliably the perception of each of the selected image features.

The evaluation of the visual contribution that each feature has on the overall impression of the image is also an essential element in this work. Further, a very large number of images need to be accessed and integrated in the multidimensional structure that is used to store image information in order to evaluate the reliability of the similarity index presented here. Moreover, we aim to investigate the use of different approaches, such as local image analysis and the addition of higher level image processes to introduce new dimensions in the multidimensional structure, or ameliorate those which are already there.

Summarizing, the system presented here is intended to provide an innovative tool for images classification based on inherent image features and their objective and subjective quantification. The dimensions in the multidimensional structure are intended to allow easy browsing and retrieval of selected images and image data. Finally, it is possible to derive a parameter for image closeness that could quantify the degree of similarity between images that relate to observed image similarity.

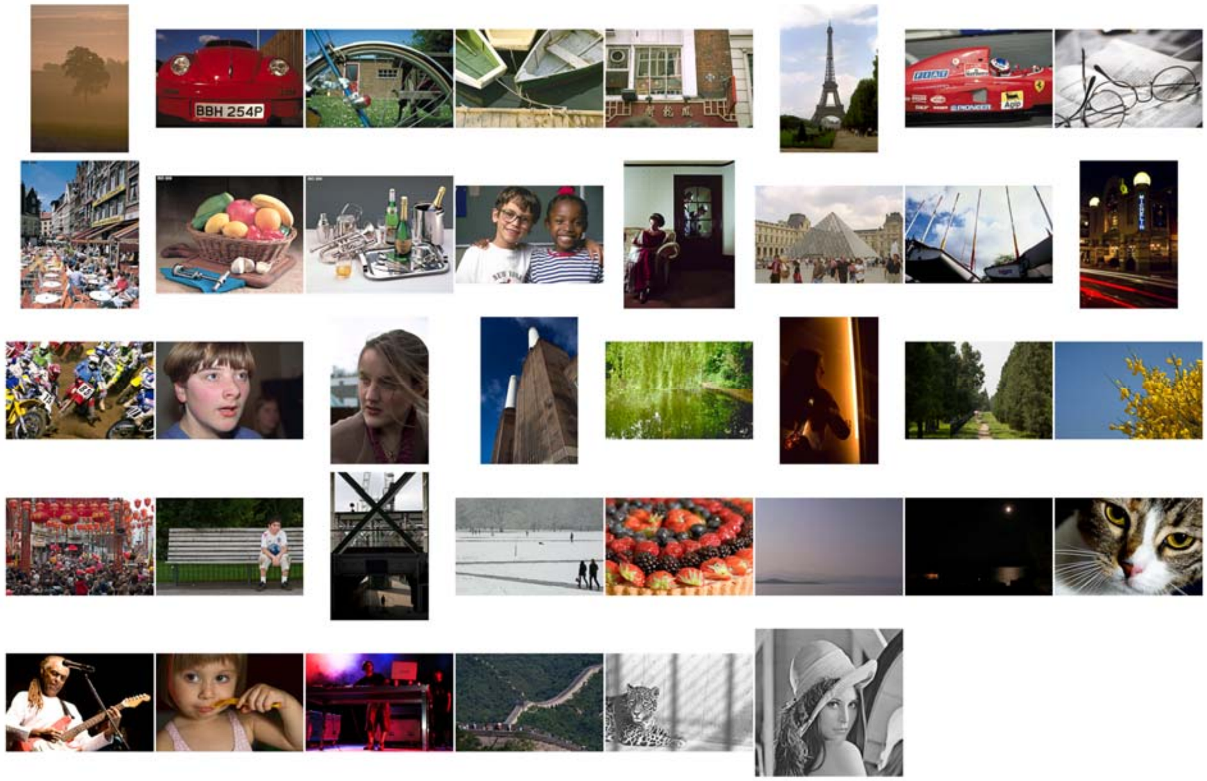

Figure 7. Test images

REFERENCES

[1] Engeldrum P.G., “Psychometric scaling”, Imcotec press, Winchester, MA, USA, (2000).

[2] Triantaphillidou S., Allen E. and Jacobson R. E., “Image Quality Comparison between JPEG and JPEG2000 – I. Psychophysical Investigation”, Journal of Imaging Science and Technology, vol.51, n.3, p.248, (2007).

[3] Triantaphillidou S., Allen E. and Jacobson R. E., “Image Quality Comparison between JPEG and JPEG2000 – II. Scene Dependency, Scene Analysis, and Classification”, Journal of Imaging Science and Technology, vol.51, n.3, p.259, (2007).

[4] Keelan B.W., “Handbook of Image Quality”, Marcel Dekker Inc., New York, NY, USA, (2002).

[5] Oh K. H., Triantaphillidou S., Jacobson R. E., “Perceptual image attribute scales derived from overall image quality” , SPIE/IS&T Electronic Imaging 2009, Image Quality and System Perfomance conference, San Jose, CA, USA, (2009).

[6] Fairchild M.D., “Colour appearance models” Wiley and Sons, 2nd ed., Chichester, England, UK, (2000).

[7] Schanda J., “Colorimetry: Understanding the CIE System”, Wiley-Interscience, Hoboken, NJ, USA, (2007).

[8] Pratt W. K., “Digital Image Processing: PIKS Scientific Inside”, 4th ED, Wiley-Interscience, Hoboken, NJ, USA, (2007).

[9] Nixon M. S., Aguado A. S., “Feature Extraction and Image Processing”, 2nd ED., Academic Press, Oxford, England, UK, (2008).

[10] Dresner H., “The Performance Management Revolution: Business Results Through Insight and Action”, Wiley, Hoboken, NJ, USA, (2008).

[11] Simonetto G. – “Sistemi di Business Intelligence e data mining nella sanità pubblica”, Thesis (Bsc) University of Padova, Italy, (2008).

[12] Inmon W. H., Imhoff C., Sousa R. , “Corporate Information Factory”, II Edition, Wiley, Hoboken, NJ, USA, (2001).
[13] Kimball R., Ross M., “Data Warehouse – la guida completa”, HOEPLI, Milano, Italy, (2003).
[14] Kimball R., Caserta J., “The Data Warehouse ETL toolkit”, Wiley, Hoboken, NJ, USA, (2004).
[15] Braun G. J., Fairchild M. D., “Colour Gamut Mapping in a Hue-Linearized CIELAB Color Space”, Sixth Color Imaging Conference: Color Science, Systems and Applications, The Society for Imaging Science and Technology, Scottsdale, Arizona, USA, (1998).
[16] Hung, P., Berns, R.S., “Determination of Constant Hue Loci for a CRT Gamut and Their Predictions Using Color Appearance Spaces”, Color Research and Application, Wiley Periodicals, Vol. No. 5, pp. 285-295, (1995).
[17] I. Sprow, Z. Baranczuk, T. Stamm, and P. Zolliker, “Web-based psychometric evaluation of image quality”, SPIE/IS&T Electronic Imaging 2009, Image Quality and System Perfomance conference, San Jose, CA, USA, (2009).
[18] Bilissi E., Jacobson R. E. and Attridge G. G., “Just noticeable gamma differences and acceptability of sRGB images displayed on a CRT monitor”, The Imaging Science Journal Vol.56, p.189, (2008).
[19] Agresti A., “Categorical Data Analysis”, 2nd ed., Wiley, Hoboken, NJ, USA, (2002).