Journal of Engineering Science and Technology
Vol. 13, No. 3 (2018) 739 - 754

INDOOR-OUTDOOR IMAGE CLASSIFICATION USING DICHROMATIC REFLECTION MODEL AND HARALICK FEATURES

A. NADIAN-GHOMSHEH

Cyber Space Research Institute, Shahid Beheshti University
Danshjoo Blv. Evin. Tehran, Iran
E-mail: a_nadian@sbu.ac.ir

Abstract

The problem of indoor-outdoor image classification using supervised learning is addressed in this paper. Conventional indoor-outdoor image classification methods, partition an image into predefined sub-blocks for feature extraction. However in this paper, we use a simple color segmentation stage to acquire meaningful regions from the image for feature extraction. The features that are used to describe an image are color correlated temperature, Haralick features, segment area and segment position. For the classification phase, an MLP was trained and tested using a dataset of 800 images. A classification accuracy of 94% compared with the result of other state of the art indoor-outdoor image classification methods showed the efficiency of the proposed method.

Keywords: Indoor-outdoor image classification, color correlated temperature, Haralick feature, co-occurrence matrix.

1. Introduction

Scene classification is a problem domain which deals with categorization of images into sematic groups. Expansion of online data storage, and demands for better organization and accurate image retrieval from large databases requires scene classification tools [1]. Labelling images as indoor or outdoor, allows for high-level processing systems to improve performance by using approaches in accordance with the scene class [2]. Further, indoor-outdoor classification has application in the field of robotics [3], smart phones [4], and color correction algorithms [5]. Image labelling algorithms could also benefit highly from high-level scene classification [6].

Conventional approaches for indoor-outdoor image classification focus on

Nomenclatures

b	Blue color channel
BX_{Chn}	Binary mask of an image segment
C_b	Body reflectance component
C_h	Color segment
Cs	Surface reflectance component
E	Entropy
FV	Total feature vector
Fv_{area}	Area feature vector
FV_{cct}	CCT feature vector
FV_h	Haralick feature vector
$Fv_{position}$	Position feature vector
g	Green color channel
H	Number of hidden neurons
I	Image
I_m	Horizontal image size
I_n	Vertical image size
h	Hue color channel
L	Scene radiation
$Mask$	Image mask
M_b	Body component scale factor
M_c	Surface component scale factor
N	Number of image segments
P	Co-occurrence matrix
r	Red color channel
s	Saturation color channel
S	Size of Co-occurrence matrix
T^{opt}_{blue}	Threshold on blue channel
T^{opt}	Optimum threshold
T^{opt}_{green}	Threshold on green channel
T^{opt}_{red}	Threshold on red channel
v	Value color channel
Greek symbols	
σ	Standard deviation
μ	mean value
λ	Wavelength
Abbreviation	
CCT	Color Correlated Temperature
DRM	Dichromatic Reflection Model
FPR	False Positive Rate
MLP	Multi-Layer Perceptron
TPR	True Positive Rate

classification of low-level features, extracted from image sub-blocks. Most frequent features that are considered describe the color and edge properties of images [7-9].

In a pioneer study, Szummer and Picard [9] divided the input image into same-sized patches. They used color and texture information from each patch to

extract the image feature vector for indoor-outdoor classification. Ohta color space [10] and multi-resolution simultaneous autoregressive model (MSAR) was used for describing color and texture. Serrano et al. [8] replaced MSAR with wavelets for describing texture in order to reduce the dimensionality of the feature space. They used SVM for image classification as it is a well-known binary classifier. The confidence values resulted from the SVM was used for deciding on the image class. Gupta et al. [11] performed unsupervised image segmentation using fuzzy c-means clustering. From each segment, color, texture, and shape features were extracted and Probabilistic Neural Network (PNN) was used for image classification. Payne and Singh [12] noted indoor images have a greater proportion of straight edges compared to outdoor images. They performed edge analysis by using multi-resolution estimates on edge straightness.

Luo and Savakis [13] integrated semantic and low-level features using a Bayesian network. They showed that semantic features extracted from ground truth sky and grass regions improves the result of indoor-outdoor image classification. They later showed that including camera metadata such as exposure time, flash fired and subject distance improves the result of indoor-outdoor image classification [14]. However, the camera metadata are not always available for all applications. Kim et al. [15] focused on the orientation of low-level features. The edge and color orientation histogram (ECOH) descriptors were defined to represent each block efficiently. The image feature vector was fed into the SVM classifier for the indoor–outdoor scene classification. Since color has shown to be a strong feature for indoor-outdoor image classification, some research focused on using multiple color channels for describing color [16, 17]. In a recent work, Cvetkovic et al. [18] used MPEG7 descriptors for extracting low-level features of indoor-outdoor images. They used SVM form image classification. Chen et al. [19] used Expert Decision Fusion (EDF) for indoor-outdoor image classification. They partitioned an image into small patches and extract several features including color, texture, and dark channel information for describing indoor and outdoor images. They applied data grouping and decision stacking for image classification.

Conventional indoor-outdoor detection methods have two main disadvantages. First, dividing images into fixed size sub-blocks will produce image partitions with mixed content that could affect the color feature extracted from each block. Secondly, simple statistical texture features do not provide information about the edge information relative to pixel positions in the image which could decrease the indoor-outdoor image classification accuracy.

Ghomsheh and Talebpour [20] showed that the Color Correlated Temperature (CCT) was a distinguishing feature for classifying indoor-outdoor images. However other features were not considered for indoor-outdoor image classification.

In this study, to improve the classification result of indoor-outdoor image detection, we propose the use of following features: simplified version of the CCT in order to decrease to computation complexity, area and position of each image segments, and Haralick texture features.

Dichromatic Reflection Model (DRM), describes how light is radiated from the surface objects [21]. Using this model it can be shown that objects have different apparent color when scene illumination changes. Correlated color

temperature (CCT) is a measure of light source color appearance defined by the proximity of the light source's chromaticity coordinates to the blackbody locus [22]. To calculate the CCT value, a special color averaging process is conducted, and the average value is used to find the CCT [20]. We simplify this process to reduce the computation time, by limiting the CCT calculation up to the point that the average color value in each image segment is found. We call the average value the CCT since it holds the same information as the actual CCT value. Haralick features have shown to provide strong features for statistically describing textual features in an image [23] and are used to describe the image texture. The Haralick features are extracted from the image segment co-occurrence matrix. For evaluation of the method, a dataset of 800 images, divided into two equal sets of indoor and outdoor images were collected. The classification stage was performed using MLP. The contribution of the paper could be summarized as below:

- Color segmentation was performed to extract features from global regions compared to small size patches.
- The color feature takes the apparent color of objects into account by using the CCT. This is an important accept as the color difference between many indoor-outdoor objects is the result of the light source.
- Previous studies show that for outdoor images, texture is random and could not be explained with specific patterns. Thus, we propose using the Haralick features for describing the texture of the image statistically.

The rest of the paper is organized as follows: Section two explains the proposed method for indoor outdoor image classification. Results are shown in section three and the paper is concluded in section four.

2. The proposed method

This section explains the proposed method for indoor-outdoor image classification. The presented method is based on two observations: The color and texture details of objects in the scene. Objects in indoor environments usually have strong straight edges, while components of outdoor images consist of fine texture details. Area and position of color regions is also notable for indoor-outdoor image classification. For example, large blue regions on the top region of the image is more likely to indicate sky.

Figure 1 shows sample indoor and outdoor images. Simple observation reveal that these two classes of images are different in terms of low-level features such as edge information, color and lighting conditions. As the samples show, straight edges occur frequently in indoor images. Outdoor images show lots of fine edges with no specific orientation. Also the presence of green and blue regions in expected regions of outdoor images is notable. Such differences between low-level features of indoor and outdoor images allows for accurate indoor-outdoor image classification. Precise indoor-outdoor image classification improves scene classification and allows such processing systems to improve the performance by taking different methods based on the scene class [5].

Fig. 1. Sample indoor (top row) and outdoor (bottom row) images.

The proposed algorithm for indoor-outdoor image classification is based on the following features:

- Color is an important feature for indoor and outdoor images as shown in Fig. 1. Further, apparent color of objects is influenced by scene illumination, thus, it was also considered in the presented color feature. Color is described based on CCT where the apparent color of an object can be described based on the scene illumination.
- Co-occurrence matrix, known as a state of the art approach for describing texture was used to describe textual features of indoor-outdoor images.
- Spatial features including area and position of homogenous color regions of the image were used to describe the spatial distribution of objects in the scene.

Further, color segmentation was considered for detecting homogenous color regions. For example, if sky is present in an image, it would be segmented into a single region and features are extracted for the entire sky region. Previous methods use small size image blocks for feature extraction. If different color regions are present in a single block, then the feature vector holds the combined information of those regions and this will reduce the performance of indoor-outdoor image classification.

The steps for indoor-outdoor image classification presented in this paper are shown in Fig. 2. To extract the image feature vector, each image is segmented into *N* color channels. From each segment the following features were extracted: CCT, position, area, and texture data. For image classification, MLP was considered as it has shown to be a strong binary classifier [24]. The process of feature extraction is explained in the next sub-sections.

2.1. Color segmentation

In order to divide an image into regions that are more relevant to the image content, color segmentation using HSV color space was performed. HSV is a perceptual color space where color is presented in its Hue component, independent of the illumination information of the image. This characteristic of

the HSV color space makes it suitable for color segmentation [25]. HSV color space was obtained from RGB information using the following equations:

$$h=\begin{cases}0 & if\ \max=\min\\ \left(60^\circ\times\dfrac{g-b}{\max-\min}\right)\bmod 360^\circ & if\ \max=r\\ 60^\circ\times\dfrac{b-r}{\max-\min}+120^\circ & if\ \max=g\\ 60^\circ\times\dfrac{r-g}{\max-\min}+240^\circ & if\ \max=b\end{cases}\quad s=\begin{cases}0 & if\ \max=0\\ 1-\min/\max & otherwise\end{cases}\quad v=\max(R,G,B) \tag{1}$$

In this equation, *r*, *g*, and *b* represent the red, green, and blue color channels of the RGB color space. *Min* and *Max* are the minimum and maximum values of the *r*, *g*, and *b* color components for each pixel respectively.

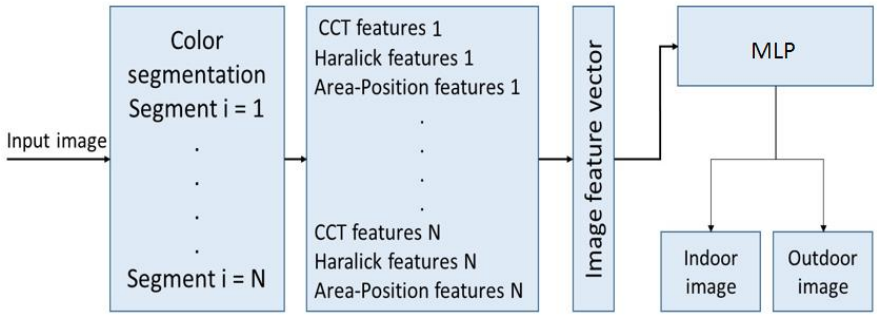


Fig. 2. The overall diagram of the proposed method.

2.2. Calculating CCT

The light captured from a scene is the result of scene illumination. The light radiated from the surface of objects can be modelled with Dichromatic Reflection Model (DRM). In this model, the scene radiation (*L*) for wavelength λ has two components:

$$L(\lambda)=m_C c_b(\lambda)+m_S c_S(\lambda) \tag{2}$$

c_b is the body reflected component and c_s is the surface reflected component. m_s and m_b are scale factors depending on illumination, view direction and surface orientation. These components are shown in Fig. 3.

By inspecting Eq. (1) it can be implied that the light entering the camera is dependent on the power spectrum of the light source. Therefore, for the same object, different apparent color is perceived under different illuminations. The perceived color can be represented by CCT. In order to calculate the CCT, the black body locus has to be found. Having the black body locus or the Planckian, it is possible to find a color temperature for each chromaticity pair which results in a unique CCT value. The CCT value does not hold additional information when compared with the chromaticity value and CCT calculation was limited up to the point where the chromaticity value has been found.

The algorithm used for calculation CCT for each color channel has the following steps: first, using *v*, discard pixels that their value is smaller than 10% of

maximum range of *v* for each *ch*, and then, take average of the remaining pixels in RGB color space to discard surface reflectance through an iterative process.

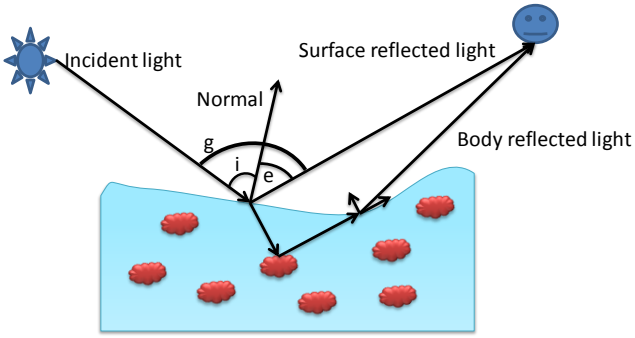


Fig. 3. Surface and Body reflectance components.

The dark pixels are discarded because they either do not reflect any light, or light does not reach them. In the second stage, an adaptive step is applied to find the average chromaticity value for each color channel. In this step, the bright pixels are discarded through an adaptive averaging step, since their information mostly reflects the properties of the illuminant, not the color of an object. The aim of averaging each channel is to discard pixels that have been exposed to direct light, or where the reflection from object's surface is in line with camera lenses. The value of such pixels may vary with respect to the surface reflectance and lighting condition. The flow chart to implement the proposed algorithm is shown in Fig. 4.

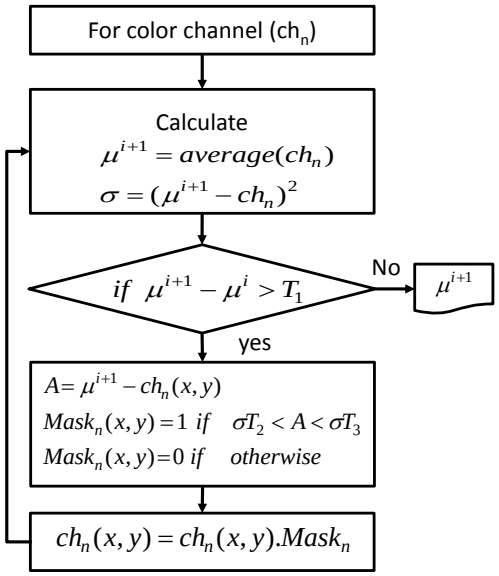


Fig. 4. Extracting the body reflectance chromaticity value.

In this process, the average value and the standard deviation for each color channel are calculated. If the difference of the mean values in two successive iterations is less than a predefined threshold, the current mean value is considered

as the body color of the channel being processed. The pixels that their difference with the mean value is less than σT_2 or greater than σT_3 are discarded from the averaging process. Discarding the pixels is done by updating the binary mask of each color channel, represented with $Mask_n$ in each iteration and then discard unwanted pixels by performing the point by point product between the mask image and the original image. In order to decide which pixels take part in the averaging process in each iteration, thresholds T_1, T_2, and T_3 should be determined. The threshold matrix was adopted from [20]:

$$\boldsymbol{T^{opt}} = \begin{bmatrix} 2 & 1 & 1 \\ 1.1 & 1.2 & 1.1 \\ 1.4 & 1.4 & 1.3 \end{bmatrix} \tag{3}$$

where T^{opt} is

$$T^{obt}(t) = \left\{ T^{opt}{}_{red}, T^{opt}{}_{Green}, T^{opt}{}_{blue} \right\} \tag{4}$$

and T^{opt} is the set of best thresholds for the red, green and blue channels of the RGB color space respectively. After calculating the average chromaticity value μ, it is possible to find the CCT of each Ch_n. To calculate the CCT of a chromaticity value $\mu(x,y)$, it is transformed into Luv color space [25], where pixel μ is represented by coordinates (u,v). These values will be stored as the body chromaticity feature. After extraction of the color channels, a total of $N\times2$ features are added to the feature space. Detailed explanation of CCT calculation is given in [20]. The CCT color feature is represented as FV_{CCT} with two values for each color channel.

2.3. Co-occurrence matrix and Harlick features

A simple way to describe texture is to use statistical moments of the grey level histogram of an image. The use of raw information alone will only lead to measure of texture that has information about the distribution of intensity. Hence, no information about the relative position of pixels with respect to each other is obtained. Co-occurrence matrix provides valuable information about the intensity of a grey image considering the relative position of pixels in an image. Give an image I, of size $I_n \times I_m$, the co-occurrence, matrix P can be defined as:

$$P(i,j) = \sum_{x=1}^{I_n} \sum_{y=1}^{I_m} \begin{cases} 1 & if \;\; I(x,y) = i \;\; and \;\; I(x+\Delta x, y+\Delta y) = j \\ 0 & otherwise \end{cases} \tag{5}$$

(Δx, Δy) represents the offset specifying the distance between the pixel being processed and its neighbours. With this notation the co-occurrence matrix is shift variant and for different choices of (Δx, Δy) different outcome is achieved. This problem will be neglected by using a set of offsets sweeping through 180 degrees at the same distance parameter Δ to achieve a degree of rotational invariance (i.e., [0 Δ] for 0◦ : $\boldsymbol{P_H}$ horizontal, [-Δ, Δ] for 45◦ : $\boldsymbol{P_R}$ right diagonal, [-Δ 0] for 90◦ : $\boldsymbol{P_V}$ vertical, and [-Δ -Δ] for 135◦ : $\boldsymbol{P_L}$ left diagonal). An illustration of the process of obtaining co-occurrence matrix for P_H and P_V is shown in Fig. 5.

In order to find the Haralick features, represented with F_h, the following notation can be considered: let $P(i,j)$ be the $(i,j)^{th}$ entry in the normalized co-

occurrence matrix and S represent the dimension of the square co-occurrence matrix. The marginal probabilities $P_x(i,j)$ and $P_y(i,j)$ will be defined as:

$$P_x(i,j) = \sum_{i=1}^{S} P(i,j) \tag{6}$$

and

$$P_y(i,j) = \sum_{i=1}^{S} P(i,j) \tag{7}$$

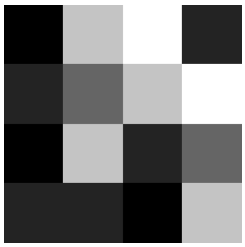
μ and σ represent the mean and standard deviation values respectively. Entropy of P_x, E_x is defined as:

$$E_x = -\sum_{i=1}^{S} P_x(i) \log\{P_x(i)\} \tag{8}$$

and entropy of P_y, E_y is defined as:

$$E_y = -\sum_{j=1}^{S} P_y(i) \log\{P_y(i)\} \tag{9}$$

By these definitions the Haralick features can be formulated as shown in Appendix1. These 14 features are extracted for each co-occurrence vector and placed in the final feature vector adding 52 elements to the feature space.



Input image

	1	2	3	4
1	0	33	43	6
2	6	17	33	43
3	0	33	6	17
4	6	17	0	33

Intensity matrix

	1	2	3	4	5
1	0	1	0	3	0
2	1	0	3	1	0
3	0	3	0	1	1
4	3	1	1	0	2
5	0	0	1	2	0

P_H

	1	2	3	4	5
1	0	0	4	0	0
2	0	0	0	4	1
3	4	0	0	1	1
4	0	4	1	0	1
5	0	1	1	1	0

P_V

Fig. 5. Co-occurrence matrix for P_H and P_V.

2.4. Area and position features

Area and position of color channels can be effective for classification of indoor-outdoor images. For example, a large blue region positioned on top region of an image can be an indication of sky, or large green areas in the bottom part of an

image can indicate the presence of trees. For extraction of area and position features, the binary mask of each color channel was used. The area feature (FV_{area}) is simply defined as:

$$Area = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} im(i,j) \mid \{i,j \in Ch_n\}}{\sum_{i=1}^{m}\sum_{j=1}^{n} im(i,j)} \tag{10}$$

image, $pos[pos_x(ch_n), pos_y(ch_n)]$ is calculated as:

$$pos_x(Ch_n) = \frac{\sum_{i=1}^{n} i \sum_{j=1}^{m} BX_{ch_n}(i,j)}{\sum\limits_{i,j \in BX_{Ch_n}} \sum BX_{Ch_n}(i,j)} \tag{11}$$

$$pos_y(Ch_n) = \frac{\sum_{j=1}^{m} j \sum_{i=1}^{n} BX_{ch_n}(i,j)}{\sum\limits_{i,j \in BX_{Ch_n}} \sum BX_{Ch_n}(i,j)} \tag{12}$$

This formulation is used to find the center of gravity for each color channel using the binary mask (BX_{Chn}) of that channel. For an image partitioned into N color channels a total of $N \times 2$ features are added to the image feature vector. The final feature vector of the image, FV is denoted as:

$$FV = [FV_{CCT}, FV_h, FV_{area}, FV_{pos}] \tag{13}$$

In this study we use the MLP network for classification of indoor-outdoor images. MLPs represent a prominent class of ANNs in classification, implementing a feed forward, supervised and hetero-associative paradigm. MLPs consist of several layers of nodes, interconnected through weighted cyclic arcs from each preceding layer to the next, without cross or feedback connections. By choosing sufficient number of hidden neurons, an MLP can approximate any complex decision boundary to separate the input training data into distinguished regions.

3. Experimental results

In order to evaluate the performance of the proposed method we use an image set consisting of 400 indoor and 400 outdoor images. 40% of the images were used for the training phase and 60% were used in the test stage. The data set was created using the WWW and can be downloaded from the following DOI: https://dx.doi.org/10.6084/m9.figshare.4595323.v1.

Image classification was tested for N = 4, 8, 16 and H =5, 10, 15, 20, where, H is the number of the hidden neurons in the MLP. Quantitative comparisons were maid based on Recursive Operative Curves (ROC). Each ROC curve is illustrated using True Positive Rate (TPR) and False Positive Rate (FPR). TPR is defined as the rate of classification for images that were correctly selected as indoor or outdoor and FPR is defined as the rate of images that were incorrectly detected as indoor or outdoor.

For the case of N = 4 the ROC curves are shown in Fig. 6. In this figure, TPR and FPR are normalized. Hence, a TPR of one is equivalent to 100. The resulted ROC curves illustrate that regardless of the number of neurons chosen, the ROC curves do not provide sufficient classification rates as the ROC curves deviate from the TPR axis as TPR approaches 100%. For quantitative analysis of the

result, Table 1 shows the FPR values for each test case when TPR is 90% and 95% respectively. As shown in the table the best classification rate is 91 %.

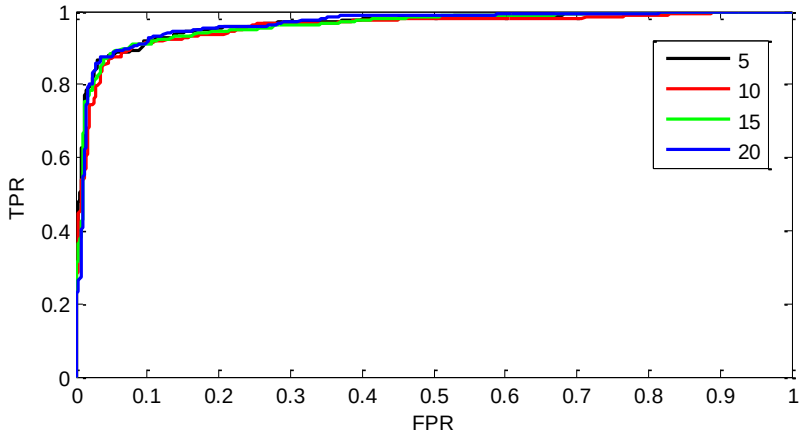


Fig. 6. Classification ROC curves with *N*=4.

Table 1. FPR values for TPR = 90 % and FPR = 95% with *N* = 4.

Number of neurons	TPR (%)	FPR (%)	Accuracy (%)
5	90	12	89.0
	95	15	90.0
10	90	17	86.5
	95	19	88.0
15	90	07	91.5
	95	22	86.5
20	90	08	91.0
	95	13	91.0

For the case were each image was divided into 8 color channels, the ROC curves shown in Fig. 7 were obtained. In this setup, the ROC curves show much better classification results as they keep near the TPR axis even for large TPR values. In these experiments, the MLP with 15 and 20 hidden layers show to be more efficient in image classification. Table 2 shows numerical results obtained from ROC curves of Fig. 7. As the table shows using 20 neurons in the hidden layer of the MLP, an accuracy 94% percent with TPR = 95% and FPR = 7% was obtained. When *H*=15 was used for training the MLP, the resulted accuracy was 93%.

The classification ROC curves for the case of $N = 16$ are shown in Fig. 8. In this case, the ROC curves also showed a good performance of the MLP for indoor-outdoor image classification. Table 3 shows the numerical results obtained from ROC curves of Fig. 7. For $N = 16$ and $H = 20$ the accuracy of classification was 94% when TPR = 95%. The results of classification for $N = 8$ and $N = 16$ are nearly very similar. However when $N = 8$ less calculation is required and this choice is more desirable.

Top row in Fig. 9 shows some sample images that were correctly classified as indoor or outdoor. The bottom row of this figure shows some false positive classification results.

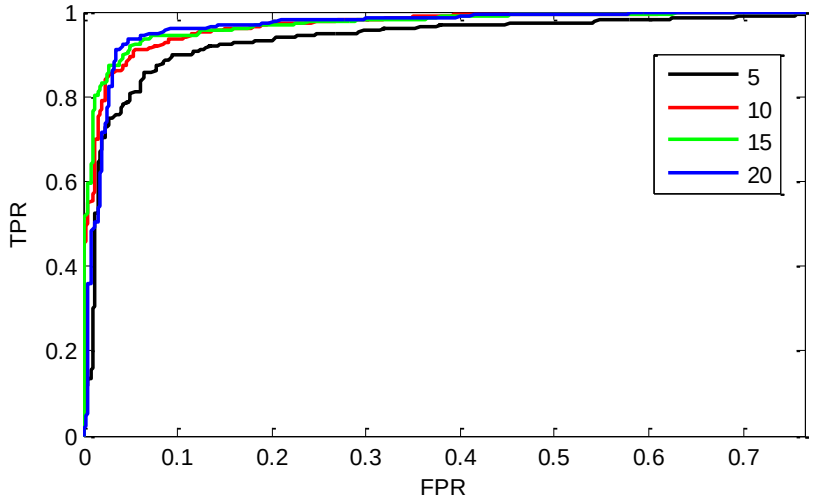


Fig. 7. Classification ROC curves with *N* = 8.

Table 2. FPR values for TPR = 90 % and FPR = 95% with *N* = 8.

Number of neurons	TPR (%)	FPR (%)	Accuracy
5	90	12	89.0
	95	23	86.0
10	90	05	92.5
	95	12	91.5
15	90	04	93.0
	95	11	92.0
20	90	03	93.5
	95	07	94.0

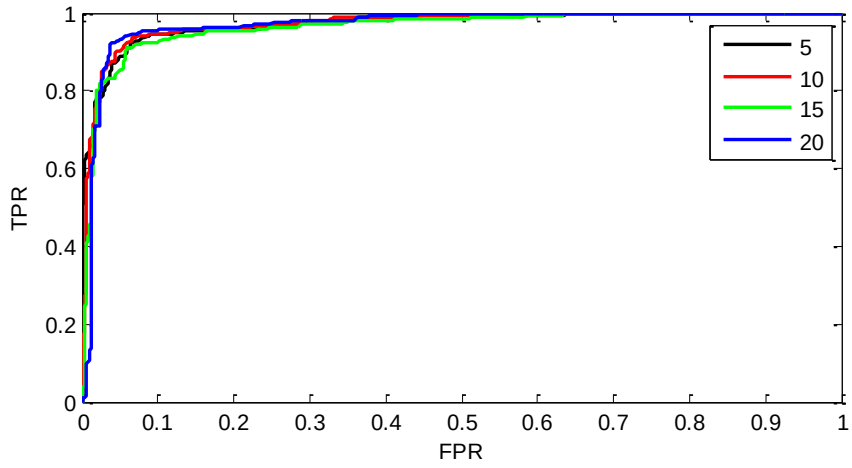


Fig. 8. Classification ROC curves with *N* = 8.

Table 3. FPR values foe TPR = 90 % and FPR = 95% with N = 8.

Number of neurons	TPR (%)	FPR (%)	Accuracy (%)
5	90	06	92.0
	95	22	86.5
10	90	05	92.5
	95	15	90.0
15	90	05	92.5
	95	18	88.5
20	90	04	93.0
	95	07	94.0

Fig. 9. Indoor-outdoor images correctly classified (Top row). Sample images, incorrectly classified (Bottom row).

The comparison between the result of classification using the proposed method and methods of Szummer and Picard [9], Kim et al. [15], and [20] are shown in Table 4. In this table Precision_x, x = 90 and 95 is the precision obtained for the cases of TPR = 90 and 95% respectively. Based on result of this table, adding Haralick features along with area and position features enhanced the classification rate of [20] by 10%, for H = 20 and TPR = 95%. This shows that adding texture and spatial features, improved indoor-outdoor image classification in comparison to [20] work. Compared with the work of [15], the presented method showed an improvement of 6% in classification precision for TPR = 95% and H = 20. The results of [9] showed that partitioning an image into small size blocks is not much effective for indoor-outdoor image classification. The best precision obtained by this method on the tested dataset was 0.78.

Table 4. Comparison of results between the proposed method and methods of [9, 15, 20].

H	TPR	[20]	[15]	[9]	Proposed method
10	90	0.85	0.90	0.78	0.94
	95	0.80	0.85	0.74	0.88
15	90	0.88	0.91	0.77	0.95
	95	0.82	0.84	0.75	0.89
20	90	0.89	0.90	0.78	0.96
	95	0.84	0.87	0.78	0.93

5. Conclusion

This paper presented a new method for indoor-outdoor image classification. The overall process included a color image segmentation step where Haralick features, CCT, and color and area features were extracted for image description. For the classification phase an MLP was used for training and test purposes. The experimental results of the presented method showed that adding texture along with area and position features for each color segment increases the accuracy of indoor-outdoor image classification rate when compared with [20] Further in comparison with the work of [15] and [9], the proposed method showed an improvement of 3.25% and 10.5% percent respectively. High classification rate of the presented method makes it suitable for applications that require to distinguishing between indoor and outdoor images.

References

1. Yu, H.; and Grimson, W.E.L. (2013). Pairwise constraints based multiview features fusion for scene classification. *Pattern Recognition*, 46(2), 483-496.
2. Tao, L.; Kim, Y.H.; and Kim, Y.T. (2010). An efficient neural network based indoor-outdoor scene classification algorithm. *International Conference on Consumer Electronics*. NV, USA, 317-318
3. De Cristóforis, P.; Nitsche, M.; Krajník, T.; Pire, T.; and Mejail, M. (2014). Hybrid vision-based navigation for mobile robots in mixed indoor/outdoor environments. *Pattern Recognition Letters*, 53(10), 118-128.
4. Zhou, P.; Zheng, Y.; Li, Z.; Li, M.; and Shen, G. (2012). IODetector: A generic service for indoor outdoor detection. *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. NY, USA, 113-126.
5. Bianco, S.; Ciocca, G.; Cusano, C.; and Schettini, R. (2008), Improving Color Constancy Using Indoor–Outdoor Image Classification. *IEEE transaction on Image Processing*, 17(12), 2381-2392.
6. Ren, Y.; Chen, C.; Li, S.; and Kuo, C-C J. (2017), *GAL: A global-attributes assisted labeling system for outdoor scenes.* Journal of Visual Communication and Image Representation, 42(2)**,** 192-206.
7. Miene, A.; Hermes, T.; Ioannidis, G.; Fathi, R.; and Herzog, O. (2003), Automatic shot boundary detection and classification of indoor and outdoor scenes, in *Information Technology, 11th Text Retrieval Conference*. Maryland, USA, 615-620.
8. Serrano, N.; Savakis, A.; and Luo, J. (2002). A computationally efficient approach to indoor/outdoor scene classification. *International Conference on Pattern Recognition*. QC, Canada, 146-149.
9. Szummer, M.; and Picard, R.W. (1998). Indoor-outdoor image classification. *Workshop on Content-Based Access of Image and Video Database*. Bombay, India, 42-51.
10. Ohta, Y.I.; Kanade, T.; and Sakai, T. (1980). Color information for region segmentation. *Computer Graphics and Image Processing*, 13(3), 222-241.
11. Gupta, L.; Pathangay, V.; Patra, A.; Dyana, A.; and Das, S. (2007), Indoor versus outdoor scene classification using probabilistic neural network. *Eurasip Journal on Advances in Signal Processing*, 2007(1), 123-133.

12. Payne, A.; and S. Singh, (2005). Indoor vs. outdoor scene classification in digital photographs. *Pattern Recognition*, 38(10), 1533-1545.
13. Luo, J.; and Savakis, A. (2001). Indoor vs outdoor classification of consumer photographs using low-level and semantic features. *International Conference on Image Processing (ICIP)*. Thessaloniki, Greece. 745-748.
14. Boutell, M.; and Luo, J. (2004). Bayesian fusion of camera metadata cues in semantic scene classification. *Computer Vision and Pattern Recognition*. Washington, USA, 623-630.
15. Kim, W.; Park, J.; and Kim, C. (2010). A Novel Method for Efficient Indoor–Outdoor Image Classification. *Signal Processing Systems*, 61(3), 251-258.
16. Collier, J.; and A. Ramirez-Serrano. (2009). Environment Classification for Indoor/Outdoor Robotic Mapping. *Canadian Conference on Computer and Robot Vision*, BC, Canada, 276-283.
17. Qiu, G.; Feng, X.; and Fang, J. (2004). Compressing histogram representations for automatic colour photo categorization. *Pattern Recognition*, 37(11), 2177-2193.
18. Cvetkovic, S.S.; Nikolić, S.V.; and Ilic, S. (2014). Effective combining of color and texture descriptors for indoor-outdoor image classification. *Facta Universitatis, Series: Electronics and Energetics*, 27(3), 399-410.
19. Chen, C.; Ren, Y.; and Kuo, C.-C.J. (2014). Large-Scale Indoor/Outdoor Image Classification via Expert Decision Fusion. *Asian Conference on Computer Vision*. Singapore, 426-442.
20. Ghomsheh, A.N.; and Talebpour, A. (2012). A new method for indoor-outdoor image classification using color correlated temperature. *Int. J. Image Process*, 6(3), 167-181.
21. Tominaga, S., (2014). Dichromatic Reflection Model. *Computer Vision: A Reference Guide*, 191-193.
22. Boyce, P.R., (2014). Human factors in lighting. Third Edition. Crc Press.
23. Haralick, R.M.; Shanmugam, K.; and Dinstein, I.H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3(6), 610-621.
24. Guo, W.; Wei, H.; Zhao, J.; and Zhang, K. (2015). Theoretical and numerical analysis of learning dynamics near singularity in multilayer perceptrons. *Neurocomputing*, 151(1), 390-400.
25. Cheng, HD.; Jiang, X-H.; Sun, Y.; and Wang, J.C. (2001). Color image segmentation: advances and prospects. *Pattern recognition*, 34(12), 2259-2281.

Appendix A

Definition of Harlick features

Feature name	Definition
Angular second moment	$f_1 = \sum_i \sum_j p(i,j)^2$
Contrast:	$f_2 = -\sum_{n=0}^{S-1} n^2 \left\{ \sum_{i=1}^{S} \sum_{j=1}^{S} p(i,j) \right\}_{\lvert i-j \rvert = n}$
Correlation:	$f_3 = \dfrac{\sum_i \sum_j (i,j)\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Variance:	$f4 = \sum_i \sum_j (i-\mu)^2\, p(i,j)$
Inverse different moment:	$f_5 = \sum_i \sum_j \dfrac{1}{1+(i-j)^2}\, p(i,j)$
Sum average*:	$f_6 = \sum_{i=2}^{2S} i p_{x+y}(i)$
Sum variance:	$f_7 = \sum_{i=2}^{2S} i(i-f_s)^2\, p_{x+y}(i)$
Sum Entropy:	$f_8 = -\sum_{i=2}^{2S} p_{x+y}(i) \log\{P_{x+y}(i)\} = f_s$
Entropy:	$f_9 = -\sum_i \sum_j p(i,j) \log(p(i,j))$
Difference variance:	$f_{10} = \sum_{i=2}^{2S} i^2\, p_{x-y}(i)$
Difference entropy:	$f_{11} = -\sum_{i=0}^{S-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$
Measure of correlation 1:	$f_{12} = \dfrac{HXY - HXY1}{\max\{HX, HY\}}$
Measure of correlation 2:	$f_{13} = (1 - \exp(-2(HXY2 - HXY))^{0.5}$
Maximum correlation coefficient	$f_{14} = \arg\max_{(i,j)} \left(\sum_k \dfrac{p(i,k)\, p(j,k)}{p_x(i)\, p_y(k)} \right)^{0.5}$

*where x and y are the coordinates of entry in the co-occurrence matrix and $p_{x+y}(i)$ is the probability of the co-occurrence matrix coordinates summing to $x+y$

** where

$$HXY1 = -\sum_i \sum_j p(i,j) \log(p(i,j)), \quad HXY2 = -\sum_i \sum_j p_x(i)\, p_y(j) \log(p_x(i)\, p_y(j))$$