Exploiting Depth Information for Indoor-Outdoor Scene Classification

Ignazio Pillai, Riccardo Satta, Giorgio Fumera, and Fabio Roli

Deparment of Electrical and Electronic Engineering,
Univ. of Cagliari Piazza d'Armi, 09123 Cagliari, Italy
{pillai,riccardo.satta,fumera,roli}@diee.unica.it

Abstract. A rapid diffusion of stereoscopic image acquisition devices is expected in the next years. Among the different potential applications that depth information can enable, in this paper we focus on its exploitation as a novel information source in the task of scene classification, and in particular to discriminate between indoor and outdoor images. This issue has not been addressed so far in the literature, probably because the extraction of depth information from two-dimensional images is a computationally demanding task. However, new-generation stereo cameras will allow a very fast computation of depth maps. We experimentally show that depth information alone provides a discriminant capability between indoor and outdoor images close to state-of-the art methods based on colour, edge and texture information, and that it allows to improve their performance, when it is used as an additional information source.

Keywords: scene classification, depth map, indoor-outdoor.

1 Introduction

Scene classification is a challenging topic in computer vision. Discriminating between indoor and outdoor images is a particular instance of scene classification which has been addressed by many authors, since it is often at the root of scene classification taxonomies [8,9,21,24]. Its applications range in various fields, including personal photo tagging, image retrieval [25], colour constancy [3], and robotics [6]. Several methods have been proposed so far to face this problem. Most of them are based on extracting low-level information embedded in the image pixels, both in the spatial domain (like colour moments and histograms) and in the frequency domain (through wavelets and the discrete cosine transform). Some recent works exploit meta-data as another source of information.

We argue that a further information source which can convey some discriminant capability between indoor and outdoor scenes is the depth map of the image, namely a map that associates each pixel of the image with its distance computed with respect to the observer.

To our knowledge, depth information was exploited to improve performance in tasks related to the more general field of scene understanding only in [11]. Here different modules (scene classification, image segmentation, object detection,

G. Maino and G.L. Foresti (Eds.): ICIAP 2011, Part II, LNCS 6979, pp. 130–139, 2011.

Fig. 1. Example of indoor (left) and outdoor (right) images and corresponding relative depth maps, estimated through [15]. Darker colours correspond to higher depth values.

and depth estimation) were combined to improve the performance of each one, by providing the output of other modules as an additional input. Instead, the usefulness of the depth information in discriminating among different scenes has not been investigated yet, although suggested in [23].

A depth map can be easily obtained from a stereo image pair (i.e., two images of the same scene taken from slightly different viewpoints) while its estimation from a single image is computationally demanding. Stereo acquisition devices are not yet widespread and this is perhaps one of the reasons why the use of depth information for scene classification has not been explored. However, thanks to the growing interest in 3D imaging, many manufacturers have presented (or are planning to present) devices able to take stereo pictures and videos.

Intuitively, depth maps of outdoor scenes are likely to exhibit higher depth values than indoor ones. However, absolute depth values, corresponding to real distances, can not be obtained unless all the parameters of the stereo camera system are exactly known. Moreover, there is always a practical limit to the maximum depth value that can be measured, which generally lies around 30 meters [26]. Still, in any case a *relative* depth map can be computed, whose values are relative ranking of pixels based on their depth. Our intuition is that even relative depth maps embed in their "structure" enough information to discriminate between indoor and outdoor scenes (see Fig. 1).

Based on the above motivations, in this paper we experimentally investigate the usefulness of relative depth maps as a novel information source for indoor-outdoor scene classification. To this aim, we propose three possible feature sets extracted from the relative depth map, based on the analysis of the pixel depth distribution in two publicly available image data sets. We then evaluate the discriminant capability of these feature sets, when they are used as the only information source for a classification algorithm, as well as when they are used as additional information sources to improve existing methods. Since no suitable stereoscopic image data set is available, we carried out experiments on two publicly available corpora of single images, estimating their relative depth maps using a recently proposed method. These depth maps can be seen as an approximation of those obtainable in real application scenarios (relative depth maps estimated by stereo pairs).

Our results show that proposed feature sets exhibit a good discriminant capability for the indoor-outdoor problem. Furthermore, we show that depth information allows to improve the performance of state-of-the art methods, due to

its complementarity with pixel-based information. Since these results have been obtained in an unfavourable setting (approximated depth maps), performances in real scenarios should be even better.

In Sect. 2 we survey previous works on indoor-outdoor scene classification. Basic information on depth map computation, including the method used in this work, is provided in Sect. 3. The features we devised based on depth maps are described in Sect. 4. In Sect. 5, we report experimental results. Conclusions and further research directions are summarised in Sect. 6.

2 Related Works on Indoor-Outdoor Classification

Most of the works on indoor-outdoor scene classification rely on low-level features based on colour, like histograms [7,17,18,20,21], and moments [10,16]. Many approaches are also based on textures [2,10,16,17,18,21,22] and/or on edges [7,14,16,21]. In a few works more complex features are also used, for example based on entropy of the pixel values [21], or shape [10]. Features are extracted either from the whole image [21], or from regions obtained by a predefined rectangular block subdivision [2,14,17,18,20] or by image segmentation [7,10]. Other works are based on bags of visual words, e.g. [2]. Recently, some authors proposed to combine pixel-level image information with the meta-data often associated to images, like *EXIF* camera informations [4,12,19] and user-generated tags [12].

In the following we focus on the methods in [17,21], which can be considered representative of the pixel-based works mentioned above, are claimed to attain a high accuracy, and provide as well enough information for their implementation. We also consider two feature sets that have proven to attain a high performance in various scene classification tasks, including the indoor-outdoor problem: the *Gist* descriptor [13] and the *Centrist* descriptor [27]. These methods will be used in the experiments of Sect. 5, and are described in the rest of this section.

The approach of [17] is based on two different feature sets, extracted from 4×4 rectangular image sub-blocks of identical size: colour features (histograms on the LST colour space), and texture features (energy of the sub-bands of a two-level wavelet decomposition). Each sub-block is classified by two SVM classifiers, based respectively on colour and texture features. The sum of the classifier outputs over all sub-blocks is computed separately for each feature set; the two resulting values are then fed to another SVM classifier, which provides the final label. In a subsequent work the authors evaluated the performance improvement obtainable through sky and grass detectors [18]. However, only by using ground truth information they obtained significant improvements in respect to [17]. For this reason, we considered [18] not profitable in a realistic classification scenario, and thus chose to implement [17].

In [21], a two-stage approach as in [17] was proposed. The input image is represented using seven different feature set, based on colour distribution, wavelet decomposition, entropy and edge directions. Each feature vector is fed to a distinct neural network. The outputs of the seven first-stage classifiers are then combined by another neural network, which provides the label of the image.

In [13] an image *Gist* descriptor was proposed, based on a set of holistic properties capable to represent the spatial structure of a scene. Such properties were estimated by means of spectral and coarsely localized information.

The *Centrist* descriptor proposed in [27] is based on the Census Transform, a per-pixel transform originally designed for matching among local patches. Histograms of the transformed image are computed at different scales by exploiting spatial pyramids.

3 Depth Map Estimation

The depth map of a scene from a stereo image pair can be computed using several methods, which exhibit a low computational time, suitable even for real time applications. A comprehensive survey is reported in [26]. The simplest way to compute the depth map is to exploit binocular disparity. Given a plane parallel to the ones where the pair of two-dimensional images lie, the distance of a given point in the scene from that plane can be computed by using triangulation [26].

If all the parameters of the stereo system are exactly known, a depth map of absolute distances can be computed. If not, only relative depths can be obtained. However, even in the former case, practical issues limit the range of depths which can be measured, which is generally between 0 to around 30 meters [26].

Given the lack of data sets of stereoscopic images suitable for scene classification, in this paper depth maps were estimated using a method based on single images. We point out that such methods provide only relative depth maps, and exhibit two main drawbacks with respect to methods based on stereo image pairs: a higher computational cost which makes them unsuitable in most real application scenarios, and a lower accuracy. Nevertheless, they are suitable to the purposes of this work, namely to investigate whether depth map information can be useful to discriminate between indoor and outdoor images. Among the different approaches proposed so far to estimate the depth map from a single image [26], we used the one in [15]. It is based on extracting several small image segments, and in inferring their 3D-orientation and 3D-location using Markov Random Fields. This method provides a depth map with a resolution of 55×305 pixels, independent of the original image resolution. It is able to estimate the depth map in 1-2 minutes, depending on the size of the input image.

4 Feature Extraction from Depth Maps

In this section we propose three possible feature sets which can be extracted from depth maps, to discriminate between indoor and outdoor images. We also discuss how such features can be combined with the ones proposed in other works (see Sect. 2), to improve their discriminant capability.

As explained in Sect.3, we consider relative depth maps obtained by the method of [15]. An example is given in Fig. 1.

To define the feature sets, we first analysed the average histogram of relative depth values computed over all images of the two data sets described in Sect. 5,

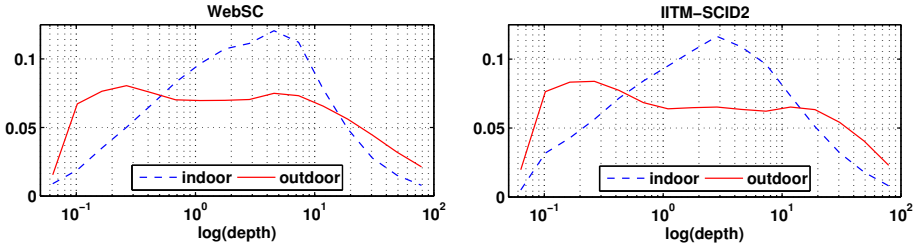

Fig. 2. Average distribution of depth values of images of the WebSC (left) and IITM-SCID2 (right) data set

separately for indoor and outdoor images. These distributions are reported in Fig. 2. It can be seen that indoor and outdoor images exhibit a clearly different behaviour, especially at lower depth values (which are emphasised by the log-scale). Depths of an indoor scene are likely to lie at medium values (see Fig. 1, left), while in outdoor scenes they are distributed more uniformly (see Fig. 1, right). Interestingly, these distributions are very similar over the two image data sets, despite the corresponding images strongly differ in terms of image quality, size, and acquisition device.

The above analysis suggests that a simple set of features potentially exhibiting a discriminant capability between indoor and outdoor images is the histogram of the logarithm of relative depth values of a given image. We denoted this feature set as $3D_{\mathrm{H}}$; its size equals the number of histogram bins.

A drawback of the histogram computed over the whole image is that it does not retain any information about the spatial distribution of depth values. To address this problem, a possible solution is to subdivide an image into $N \times N$ sub-blocks, and to compute the average logarithm of depth values of each sub-block. This feature set is denoted as $3D_{\mathrm{B}}$, and its size is equal to N^2.

Like any other 2D signal, the depth map can be represented in terms of frequency and phase values. Intuitively, outdoor scenes should exhibit an higher contribution at lower frequencies than indoor scenes. Indeed, lower frequencies are likely to correspond to larger homogeneous areas like sky, sea or sandy beach. Based on this intuition, we define a third feature set made up of the average DCT coefficients, computed by a $K \times K$ window sliding over the image. The size of the resulting feature set, named $3D_{\mathrm{D}}$, is K^2.

5 Experimental Evaluation

In this section we experimentally assess the discriminant capability of the feature sets proposed in Sect. 4. We first compare their discriminant capability with the reference methods mentioned in Sect. 2 [13,17,21,27]. We then investigate whether the performance of each reference method can be improved, by using each of the proposed feature sets as additional information source.

5.1 Experimental Setup

Experiments were carried on two benchmark data sets of indoor and outdoor images. Depth maps were generated using the method in [15]. The first data set is IITM-SCID2.[1] It was used in [10,21], and is made up of 907 images (442 indoor, 465 outdoor), subdivided into 393 training and 514 test images. We removed one test image which was too small to be processed by the depth map estimation method. The second data set, denoted as "Web Scene Collection" (WebSC), is made up of 1917 images (955 indoor, 962 outdoor) collected by the authors from the Web and manually labelled. Both data sets, together with the corresponding depth maps, are available at `http://prag.diee.unica.it/public/datasets`.

The characteristics of the images in the two data sets are rather different. The WebSC corpus contains mainly good-quality, high resolution images. IITM-SCID2 is instead mostly made up of low-quality, low-resolution images, which are often out of focus, and exhibit chromatic aberrations. IITM-SCID2 is thus more challenging than WebSC.

The methods in [17,21] were implemented as follows. They both adopt a two-stage classification scheme. In [17] SVM classifiers with RBF kernel were used at both stages, while in [21] neural networks were used. However, we adopted SVMs with a RBF kernel also for the latter method, as their performance was better than the one of neural networks. To combine the feature sets proposed in this work to the reference methods, we simply added another SVM classifier with a RBF kernel to the first stage, with our features as input.

We used a SVM with a RBF kernel also for the *Gist* [13] and *Centrist* [27] feature sets. To add our feature sets, a two stage classifier similar to the previous ones was built, using again SVMs with a RBF kernel.

Our $3D_{\mathrm{H}}$ feature set was computed as a 16-bin histogram, while for the $3D_{\mathrm{B}}$ feature set images were subdivided into 4×4 sub-blocks. For the $3D_{\mathrm{D}}$ feature set, we chose a sliding window of 8×8 pixel.

We trained the second-stage classifiers of all the methods using the scores provided by first-stage classifiers on training images, obtained through a 5-fold cross-validation. The C parameter of the SVM learning algorithm and the σ parameter of the RBF kernel $K(x_i, x_j) = exp(-\|x_i - x_j\|/(2\sigma))$ were estimated by a 3-fold cross validation on training data. SVMs were implemented using the LibSVM software library [5].

Classification performance was measured as overall and per-class accuracy. Concerning the IITM-SCID2 corpus, we kept the original subdivision into training and testing sets, to allow a direct comparison with the results reported in [10,21]. For WebSC we adopted the 5×2 cross-validation approach of [1], to evaluate also statistical significance. In this case the classifier parameters were estimated separately on each training fold.

As a preliminary evaluation of the degree of complementarity of our feature sets with the ones of the reference methods, we analysed the joint distribution of the score values provided by the corresponding classifiers. Fig 3 shows the joint score distribution for $3D_{\mathrm{B}}$ and *Centrist*, on WebSC. The scores exhibit a high

[1] `http://www.cse.iitm.ac.in/~sdas/vplab/SCID/`

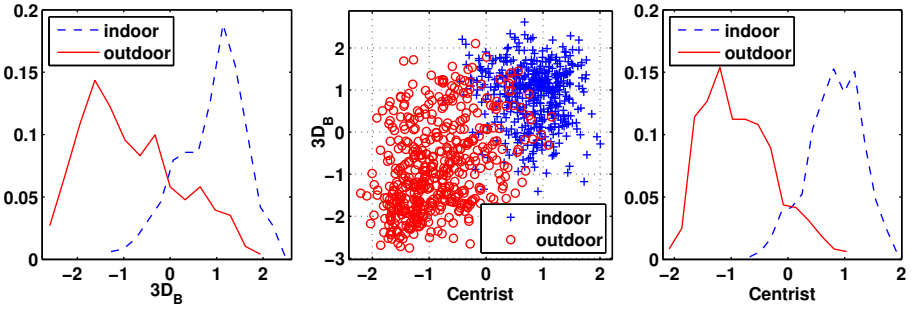

Fig. 3. Joint distributions of the classifier scores obtained by the *Centrist* and $3D_{\mathrm{B}}$ feature sets on WebSC. Each point in the middle plot corresponds to a single image.

degree of complementarity, which suggests that combining depth information with pixel-based information could improve classification performance. A similar behaviour was observed using all the other feature sets.

5.2 Results and Discussion

Table 1 shows the classification accuracy attained by our proposed features sets, by the four reference feature sets, and by all the possible combinations of one of our feature sets with one of the reference methods.

Our feature sets attained an overall classification accuracy between about 70% and 80%. This supports our intuition that depth map provides useful information to discriminate between indoor and outdoor images. Among the proposed features, $3D_{\mathrm{B}}$ attained the best performance on both data sets. This suggests that taking into account also the spatial depth distribution (as in the $3D_{\mathrm{B}}$ features) is beneficial for the considered classification task.

Classifiers based on our features attained however a lower performance than each of the reference methods. Nevertheless, it is worth pointing out that their performance turned out to be comparable to the average accuracy attained by the *individual* feature sets used in [17,21]. In particular, the colour-based features of [17] exhibited an overall classification accuracy of 0.77 and 0.83 respectively on the IITM-SCID2 and WebSC data set, while texture-based features attained an overall accuracy on the same data sets respectively equal to 0.75 and 0.77. The accuracy of the seven feature sets of [21] were between 0.68 and 0.80 for the WebSC data set, and between 0.52 and 0.78 for the IITM-SCID2 data set.

Despite the overall classification accuracy attained by the four reference methods was higher than the one attained by classifiers based on our feature sets, Table 1 shows that the performance of the reference methods was almost always improved when the corresponding classifiers were combined with the ones based on our features. The only exception can be observed for the combination of the *Centrist* and $3D_{\mathrm{D}}$ features, on the IITM-SCID2 data set.

Table 1. Classification accuracy attained by our proposed features sets (top three rows), by the reference methods (top row of each subsequent group of rows), and by all the possible combinations of each reference method with each of our feature sets (remaining rows). For the WebSC data set, the average accuracy and the standard deviation over the 5×2 cross-val. procedure is reported; $*$ and $**$ denote, respectively, results significant with 90% and 95% confidence, with respect to the f-test.

Method	IITM-SCID2			WebSC			
	Indoor	Outdoor	Total	Indoor	Outdoor	Total	
$3D_{\mathrm{B}}$	0.803	0.742	0.772	0.857 ± 0.011	0.750 ± 0.018	0.803 ± 0.010	
$3D_{\mathrm{H}}$	0.635	0.788	0.713	0.835 ± 0.043	0.745 ± 0.018	0.790 ± 0.018	
$3D_{\mathrm{D}}$	0.695	0.773	0.735	0.813 ± 0.022	0.758 ± 0.019	0.785 ± 0.015	
Centrist [27]	0.960	0.875	0.916	0.932 ± 0.016	0.907 ± 0.007	0.920 ± 0.007	
Centrsti + $3D_{\mathrm{B}}$	0.944	0.909	0.926	0.960 ± 0.009	0.917 ± 0.012	0.938 ± 0.005	**
Centrist + $3D_{\mathrm{H}}$	0.940	0.909	0.924	0.956 ± 0.007	0.910 ± 0.010	0.933 ± 0.006	**
Centrist + $3D_{\mathrm{D}}$	0.920	0.902	0.910	0.954 ± 0.011	0.890 ± 0.021	0.922 ± 0.006	
Tao [21]	0.896	0.811	0.852	0.936 ± 0.007	0.906 ± 0.017	0.921 ± 0.009	
Tao + $3D_{\mathrm{B}}$	0.908	0.864	0.885	0.941 ± 0.011	0.912 ± 0.008	0.927 ± 0.003	*
Tao + $3D_{\mathrm{H}}$	0.863	0.879	0.871	0.945 ± 0.008	0.907 ± 0.010	0.926 ± 0.006	
Tao + $3D_{\mathrm{D}}$	0.871	0.886	0.879	0.938 ± 0.012	0.910 ± 0.010	0.924 ± 0.005	
Gist [13]	0.847	0.852	0.850	0.924 ± 0.011	0.876 ± 0.025	0.900 ± 0.014	
Gist + $3D_{\mathrm{B}}$	0.884	0.860	0.871	0.938 ± 0.011	0.891 ± 0.015	0.914 ± 0.009	*
Gist + $3D_{\mathrm{H}}$	0.851	0.883	0.867	0.941 ± 0.013	0.885 ± 0.022	0.913 ± 0.014	*
Gist + $3D_{\mathrm{D}}$	0.847	0.883	0.865	0.944 ± 0.009	0.862 ± 0.021	0.903 ± 0.010	
Serrano [17]	0.871	0.837	0.854	0.871 ± 0.014	0.866 ± 0.010	0.868 ± 0.005	
Serrano + $3D_{\mathrm{B}}$	0.876	0.883	0.879	0.912 ± 0.013	0.870 ± 0.012	0.891 ± 0.009	**
Serrano + $3D_{\mathrm{H}}$	0.855	0.883	0.869	0.908 ± 0.018	0.872 ± 0.012	0.890 ± 0.012	**
Serrano + $3D_{\mathrm{D}}$	0.827	0.894	0.862	0.896 ± 0.010	0.879 ± 0.017	0.887 ± 0.007	**

To assess the statistical significance of the observed accuracy improvements on the WebSC data set, we performed the f-test of [1]. We did not apply this test on the IITM-SCID2 data set, as a single run of the experiments was carried out on the predefined subdivision into a training and a testing set. Table 1 shows that the accuracy improvements were found to be statistically significant at the 90% or 95% confidence level, in eight out of twelve cases. In particular, the improvements attained by using the $3D_{\mathrm{B}}$ feature set turned out to be always statistically significant with a confidence level of at least 90%. These results provide evidence that, although image depth information may exhibit a lower discriminant capability than other information sources for the indoor/outdoor image classification task, it also provides *complementary* information, as suggested by the results reported at the end of Sect. 5.1. This can allow to improve the discriminant capability of other information sources, by combining them with image depth information.

6 Conclusions and Future Work

In this work we investigated the usefulness of image depth map as an information source in the task of scene classification, and in particular to discriminate between indoor and outdoor images. Our interest is motivated by the rapid diffusion of stereoscopic image acquisition devices which is expected in the next years. We provided evidences that relative depth maps embed in their "structure" useful information for discriminating between indoor and outdoor images. Moreover, we showed that such information exhibits a complementariness to other information sources used by state-of-the-art methods, and that the discriminant capability of the latter can be improved by combining their feature sets with features extracted from depth maps.

We point out that our experiments were carried out by estimating relative depth maps from single images, due to the current lack of data set of stereo images. Since estimated depth maps are less accurate than the ones which can be obtained from stereo images, the latter can be expected to provide an even higher discriminant capability than the one observed in our experiments.

An interesting follow-up of this work is to devise more informative features based on depth maps, as the ones considered in this work are only a preliminary attempt. To this aim, the information about spatial distribution of depth values could be further investigated, since it exhibited the highest discriminant capability among the considered features. It is also interesting to investigate the usefulness of depth information for other, more complex scene classification tasks. Finally, it will be clearly useful to construct a data set of stereo images, representative of a real application scenario.

Acknowledgements. This work was partly supported by a grant from Regione Autonoma della Sardegna awarded to Ignazio Pillai, PO Sardegna FSE 2007-2013, L.R.7/2007 "Promotion of the scientific research and technological innovation in Sardinia".

References

1. Alpaydin, E.: Combined 5 x 2 cv F test for comparing supervised classification learning algorithms. Neural Computation 11(8), 1885–1892 (1999)
2. Battiato, S., Farinella, G.M., Gallo, G., Ravi, D.: Exploiting Textons Distributions on Spatial Hierarchy for Scene Classification. EURASIP Journal on Image and Video Processing (2010)
3. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Improving color constancy using indoor-outdoor image classification. IEEE Trans. on Image Processing 17(12), 2381–2392 (2008)
4. Boutell, M., Luo, J.: Beyond pixels: Exploiting camera metadata for photo classification. Pattern Recognition 38(6), 935–946 (2005)
5. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines (2001), `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

6. Collier, J., Ramirez-Serrano, A.: Environment classification for indoor/outdoor robotic mapping. In: Canadian Conference on Computer and Robot Vision, CRV 2009, pp. 276–283 (May 2009)
7. Deng, D., Zhang, J.: Combining multiple precision-boosted classifiers for indoor-outdoor scene classification. In: Int. Conf. on Information Technology and Applications, vol. 2, pp. 720–725 (2005)
8. Ehinger, K.A., Torralba, A., Oliva, A.: A taxonomy of visual scenes: Typicality ratings and hierarchical classification. Journal of Vision 10(7), 1237 (2010)
9. Fei-Fei, L., Iyer, A., Koch, C., Perona, P.: What do we perceive in a glance of a real-world scene? Journal of Vision 7(1) (2007)
10. Gupta, L., Pathangay, V., Patra, A., Dyana, A., Das, S.: Indoor versus outdoor scene classification using probabilistic neural network. EURASIP Journ. on Adv. in Signal Processing, Special Issue on Image Perception 2007(1), 123–123 (2007)
11. Heitz, G., Gould, S., Saxena, A., Koller, D.: Cascaded classification models: Combining models for holistic scene understanding. In: NIPS (2008)
12. Lee, B.N., Chen, W.Y., Chang, E.Y.: A scalable service for photo annotation, sharing, and search. In: Proc. of the 14th Annual ACM Int. Conf. on Multimedia, pp. 699–702 (2006)
13. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. Jour. of Computer Vision 42, 145–175 (2001)
14. Payne, A., Singh, S.: Indoor vs. outdoor scene classification in digital photographs. Pattern Recognition 38(10), 1533–1545 (2005)
15. Saxena, A., Sun, M., Ng, A.: Make3d: Depth perception from a single still image. In: Proc. of The AAAI Conf. on Artificial Intelligence, pp. 1571–1576 (2008)
16. Schettini, R., Brambilla, C., Cusano, C., Ciocca, G.: Automatic classification of digital photographs based on decision forests. International Journal of Pattern Recognition and Artificial Intelligence 18(5), 819–845 (2004)
17. Serrano, N., Savakis, A., Luo, J.: A computationally efficient approach to indoor/outdoor scene classification. In: ICPR, vol. 4, pp. 146–149 (2002)
18. Serrano, N., Savakis, A.E., Luo, J.: Improved scene classification using efficient low-level features and semantic cues. Pattern Recognition 37(9), 1773–1784 (2004)
19. Sinha, P., Jain, R.: Classification and annotation of digital photos using optical context data. In: Proc. of The 2008 Int. Conf. on Content-Based Image and Video Retrieval, New York, NY, USA (2008)
20. Szummer, M., Picard, R.W.: Indoor-outdoor image classification. In: Proc. of IEEE Int. Workshop on Content-Based Access of Image and Video Database, pp. 42–51 (1998)
21. Tao, L., Kim, Y.H., Kim, Y.T.: An efficient neural network based indoor-outdoor scene classification algorithm. In: Int. Conf. on Consumer Electronics (ICCE). Digest of Technical Papers, pp. 317–318 (2010)
22. Torralba, A., Oliva, A.: Semantic organization of scenes using discriminant structural templates. In: Int. Conf. on Computer Vision, pp. 1253–1258 (1999)
23. Torralba, A., Oliva, A.: Depth estimation from image structure. IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (2002)
24. Tversky, B., Hemenway, K.: Categories of environmental scenes. Cognitive Psychology 15(1), 121–149 (1983)
25. Vailaya, A., Figueiredo, M., Jain, A., Zhang, H.J.: Image classification for content-based indexing. IEEE Trans. on Image Processing 10(1), 117–130 (2001)
26. Wei, Q.Q.: Converting 2d to 3d: A survey (2005)
27. Wu, J., Rehg, J.M.: Centrist: A visual descriptor for scene categorization. IEEE Trans. on Pattern Analysis and Machine Intelligence 99 (2010)