Real-time Image Scene Classification and Segmentation System

Sheng-Fuu Lin
Dept. of Electrical Engineering National Chiao Tung University
Hsinchu, Taiwan (R.O.C)
sflin@mail.nctu.edu.tw

Huang-Tsun Chen*
Dept. of Electrical Engineering National Chiao Tung University
Hsinchu, Taiwan (R.O.C)
guro@maruco.url.com.tw

Shih-Hao Shih
Dept. of Electrical Engineering National Chiao Tung University
Hsinchu, Taiwan (R.O.C)
Shinhao.ece93g@nctu.edu.tw

***Abstract*—An original approach is proposed. Called the Pre-Segmented Region of Interest Classification Scene System (PSROI), this system is able both to classify a scene during a digital camera's pre-capture phase, and to determine image processor parameters that will determine the quality of the final picture: white balancing, exposure, and focus. Additionally, it also integrates with the focusing system. An image in the classification scene system is given a value, called the focus weight. The focus weight ensures that the selected classification scene is more suitable for user vision. In the scene classification system, three scenes are set: portrait, landscape, and beach/snow. In order to perform the classification more quickly, the operation region of the image shrinks, and less computation is applied to build the system. Experimental results show that the proposed system is capable of effectively classifying scenes into the right categories within 0.2 s..**

Keywords- scene classification, recognition, support vector machines, fuzzy

I. INTRODUCTION

With the digital still camera development maturing and becoming more competitive, users are requesting more from the devices in terms of cost, appearance, and function. In order to fulfill such requests, manufacturers tend to produce light and beautiful cameras with greater functionality and more scene modes. However, including yet more scene modes invites trouble. For example, digital still cameras need to have a bigger mode dial in order to contain these alternative modes, which goes against the mechanical design of a light digital camera. Moreover, the need to constantly switch between scenes to ensure that photos are taken with the best possible exposure can be a frustrating experience for users.

In order to let users operate digital still cameras conveniently, one possible solution is auto scene. Auto scene mode incorporates all of, or a subset of, a camera's scene modes, provided by the camera manufacturer, into a single scene mode. When photos are taken with this mode, scene classification identifies the scene and sets up the relevant parameters automatically. It does away with the inconvenience of switching between modes, and in doing so provides manufacturers with more flexibility in their choice of mechanical design.

Current studies related to scene classification are in the field of image retrieval, image annotation, and computer vision [1]. These studies are applied to image classification, identification, or retrieval in the backend. There are no studies related to applying scene classification during the process of taking photos. The scene classification system that this paper studies is one that allows users take photos without changing the scene mode.

In recent years, scene classification has become a popular research topic. However, even though much research has been done before, classifying photos into semantic types of scene (e.g., portrait, landscape) is still a difficult problem. The popular method for classifying scenes is to use low-level features (e.g., texture, color). Therefore, image classification can be achieved either by only using low-level features, or by integrating the low- and high-level features [2]. A new classification framework for digital cameras, which employs low-level features, is proposed in this paper. The RGB color space is the most common color format in digital images. Lu et al. [3] used the RGB color space to represent the color of an image patch and proposed a two-level approach to scene recognition for both indoor-outdoor and multiple photo categories. Naccari et al. [4] used the method of

modulated color enhancement to classify natural images. Serrano et al. [5] proposed a set of low-dimensional, computationally efficient low-level features that are extracted from LST color space and wavelet texture features. Here, LST is a luminance-chrominance color space. It consists of one luminance component L, and two chrominance components S and L. In general, we can describe a color photo by using its color features. Color histograms are a popular method of representing the image [6]–[8]. Another popular low-level feature is texture. Texture is the structure arrangement of the surface of an object, which assists users in recognizing objects or regions. Textures have been expressed using several methods [9]–[14], such as Gabor filters, wavelets, local binary patterns (LBP), and so on.

The semantic information of an image carries the meaning of that image. It is trivial for the human eye to extract semantic information from photos. However, for a computer, it is difficult to identify the semantic features of high-level images in photos. Therefore, if a computer can be made to correctly identify the semantic features of objects in photos, it will enhance the image identification rate. Some researchers use the power spectrum to correlate spectral features with the semantic contents of an image [15][16]. Another spectrum-related system used is a harmonic analysis tool called the ridgelet transform [17]. In this system, the Fourier energy spectra are used as an efficient way to describe the semantic information of an image. Boutell et al. [18][19] trained a single-label classifier for each label, using all single-label documents and only the multi-label documents with that label. In order to make systems user-friendly, many researches use low-level features to infer the semantic features [20]. Integration of low-level features and semantic features to enhance the image classification system has also been developed widely [2][3][5]. When classifying the scene of an image, the most important element is to extract the features of the image and then use those features to both train the system and classify the image. It is important to recognize that the content of an image is composed of objects in the scene. In order to extract the features, a scene-classification system needs to be sure which objects are in the image before classifying the scene. The common method of doing this is to segment the objects in the image and then identify the low-level features and semantic concepts. There are two methods of segmenting the image to look for the objects. One is block-based [21]–[23], and the other is region-based [3][24]. The block-based method simply segments the image into several rectangles. The region-based method also segments the image, but the objects are more meaningful to the human eye. The most common approach for Exemplar-based systems use low-level features and statistical pattern recognition techniques. Such systems rely on learning patterns from a training set [25]. Bayesian classifiers are the most popular classifiers. Many researchers propose different Bayesian frameworks to improve the rate of image classification. Vailaya et al. [26] cast the image-classification problem in a Bayesian framework. Luo et al. [27][28] presented a unified image-understanding framework based on a Bayesian network, where both low-level and semantic features are integrated to improve performance. Tsin et al. [29] proposed a Bayesian approach to classifying a color image of an outdoor scene. In some research, metadata is used to improve classification performance, applying a Bayesian network to integrate metadata cues in a robust framework [30][31]. Much research has also been done on problems of scene classification [32]–[37]. The majority of these systems employed a probabilistic-model approach based on low-level features derived exclusively from scene content. Several content-based methods have been proposed for image classification and image retrieval. Vailaya et al. [38] attempted to capture high-level concepts from low-level image features under the constraint that the test image belongs to one of the classes. They additionally used a hierarchical architecture to classify vacation images. Vogel and Schiele [39] presented an image representation that renders it possible to access natural scenes by local semantic description. They modeled the semantic content of an image and used this model to classify local image regions into semantic concept classes. Torralba and Oliva [40] studied the statistical properties of natural images belonging to different categories and their relevance for scene- and object-categorization tasks. Mode-based systems rely upon the configuration of the scene components. Luo and Stephen [41] proposed a mode-based approach to detecting sky. This approach consists of color classification, region extraction, and physics-motivated sky-signature validation. Lipson et al. [42] at MIT used an approach they call “configural recognition,” using relative spatial and color relationships between pixels in low-resolution images to match the images with class models. Some researchers believe that a segmentation of images into regions can provide more semantic information than the usual global image features, and have proposed a scene-recognition approach to identify the image-region types in a given scene, along with a classification scheme designed to separate images in the descriptor space [24][43]. Kaick and Mori [44] also segmented images into different regions based on different low-level features, and compared this to a method that partitions the image into regular blocks. The quality of the results does not differ significantly whether regular blocks or segments are used as the regions for similarity computation. Our paper will use the block-based method, and will give the relevant semantic information for each block. The support vector machine (SVM) classifier is a new method of parameterization of functions. It has been shown to have better error rates [6]. Therefore, SVM is chosen to classify low-level features as semantic features in this paper. A more complete description of SVM can be found in Ref. [45]. Since fuzzy set theory was proposed in 1965, the related study of fuzzy rule-based systems (FRBSs) has developed continuously [46]. An FRBS is used in this paper to classify images into one of three scenes: portrait, landscape, or beach/snow.

This paper is structured as follows: Section II gives an overview of the system architecture. In Section III, details of the system, including the training and testing modules, are given. Section IV presents experimental results of the implementation of our proposed algorithm. We deliver a conclusion in Section V.

II. SYSTEM ARCHITECTURE

Fig. 1 is a flow chart representing PSROI. The figure shows that the PSROI system has three main portions. The first portion is image-features extraction. In this step, low-level features are those that can be extracted from the image, such as color, texture, and edges. The second portion is the semantic features. Most classification

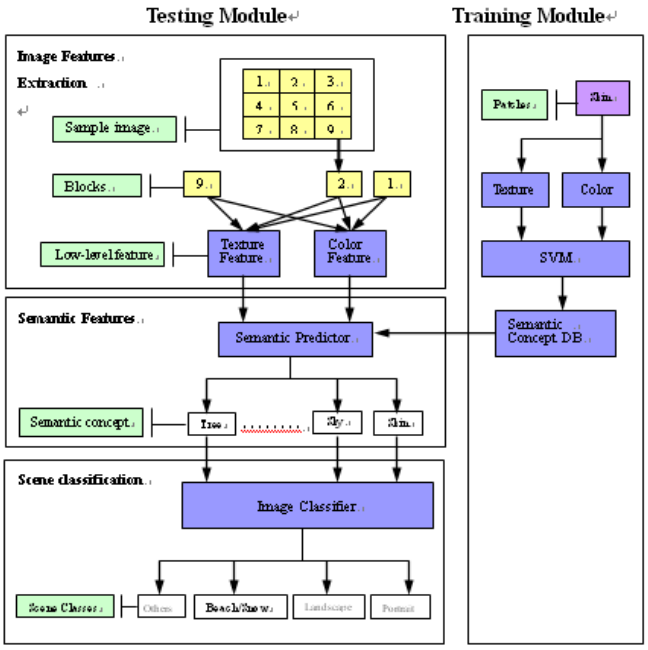

Fig. 1. System architecture

systems use semantic concepts to present the object in the image [2][5][15][16][39], which is the best method for users to classify the image. The third portion is image-scene classification. The image is classified into portrait, landscape, beach/snow, or others by using a rule-based classifier.

III. SYSTEM DETAILS

A. *Training Module*

The general approach of a training system is to compute feature vectors and then develop classifiers for these features. This paper performs supervised feature classification, which can control the structure of representations. The role of the training module is to train the image patches of seven predefined materials that are segmented by human semantic concepts. These seven materials are skin, sky, grass, water, snow, sand, and tree.

Image regions for each material are cropped manually into the database to prepare the training data. These image patches are cropped in the shape of rectangles of 80×60 or 60×80 pixels based on testing the image direction. There are a total of 3,150 image patches for all of the seven materials. 1,400 image patches out of 3,150 are aligned in the vertical direction.

In the training module, the first step is to extract low-level features from the image patch of each material. Two kinds of low-level features are extracted — color and texture feature. Color features are extracted from LST color space and texture features are extracted via wavelet transform, which will be introduced in the next section. After the extraction of color and texture features, image patches of all materials are trained via SVM to build a semantic concept database. And the training mode file will be used in the testing procedure.

B. *Testing Module*

1) *Image Feature Extraction*

To extract image features, it is necessary to firstly segment the region of interest from the sample image. The most important information in photos from digital cameras is in the middle of the image, so we set the middle

area as the region of interest (ROI). We then segment into 3×3 blocks the region of interest, and extract the low-level features from each block going from left to right and up to down.

Generally speaking, the image is described by using low- and high-level features. Low-level features are those features such as color, texture, and edges — the most common image features for classifying images — which are collected via image processing. High-level features are those features that match human semantic concepts. Color and texture are chosen as the image features in this paper. The following is an elaboration of color and texture.

Color features are extracted from an LST color space. This color space had been shown to be suitable for image classification [47]. The transformation equation of RGB to LST is:

$$\begin{aligned} L &= \frac{1}{\sqrt{3}}(R+G+B) \\ S &= \frac{1}{\sqrt{2}}(R-B) \\ T &= \frac{1}{\sqrt{6}}(R-2G+B). \end{aligned} \tag{1}$$

The equations in (1) consists of one luminance component L, and two chrominance components S and T. The two chrominance components do not vary with light-source intensity changes; the S component approximately represents the illumination variation (daylight to tungsten light). This is an important point, because it can be used to distinguish different illuminants, whether natural to artificial, which allows the classification system to more accurately classify the scene category.

In this paper, a color histogram is used to analyze the color-distribution feature because it can resist the changing of rotation and shift no matter whether photos are taken horizontally or vertically. Moreover, it is simple and fast. Quantized color histograms are suitable for those features that are computationally simple and have been shown to be useful for image classification [2][5][21][22]. Let $\{q_L(x,y), q_S(x,y), q_T(x,y)\} \in [0, n_C - 1]$ be quantized representations of the L, S, and T image channels, respectively. The quantized color histogram features are then simply

$$\begin{aligned} H_L(k) &= \Pr\{q_L(x,y) = k\} \\ H_S(k) &= \Pr\{q_S(x,y) = k\} \\ H_T(k) &= \Pr\{q_T(x,y) = k\}, \end{aligned} \tag{2}$$

where k = 0, …, $n_C - 1$, and n_C is the number of bins per color histogram. Then we get color feature x_{cj}, where j = 1, 2, …, 9, via combining H_L, H_S, and H_T. The nine color features are extracted from nine blocks in

the region of interest as shown in Fig. 2. This color-feature extraction approach is similar to [5]. The number

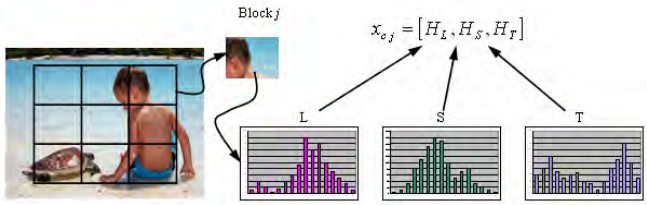

Fig. 2. Color-feature extraction

of bins per color was set to n_C = 16, and so the color-feature dimension is $3n_C$ = 48.

Various methods of texture analysis have been proposed in recent years, such as filters, wavelets, and so on. Of these, it has been shown that wavelet transform can describe both local and global information efficiently and has comparatively better computational efficiency [5][12][14]. Because this paper needs to describe which local information belongs to which semantic concept, the wavelet transform is chosen to extract texture features. Considering both computational tractability and optimal texture characterization, the 5/3 wavelet filter pair is used, where $h_0(k)$ = [-1/8 1/4 3/4 1/4 -1/8] and $h_1(k)$ = [-1/2 1 -1/2] are the low- and high-pass filters, respectively. The wavelet coefficients can be derived from:

$$C_{ij}^{l}(x,y)\begin{cases}\sum_{m}\sum_{n} L(m,n)h_i(2x-m)h_j(2y-n) & l=1\\ \sum_{m}\sum_{n} C_{00}^{l-1}(m,n)h_i(2x-m)h_j(2y-n) & l\geq 2\end{cases} \tag{3}$$

where L(m, n) is the image-luminance information and $C_{ij}^{l}(x,y)$ are the wavelet coefficients obtained by applying filters $h_i(k)$ and $h_j(k)$ along the rows and columns of the signal at decomposition level l. Here, we implement a two-level wavelet transform to get the wavelet coefficients of seven sets from each block.

Texture features are extracted from the wavelet coefficients of the two-level decomposition described above. However, the lowest frequency coefficients c_{00}^{2} are not useful for texture analysis. A high-pass filter is therefore used to transform the coefficients c_{00}^{2} to get high-frequency signal information. The transformation equation is:

$$c_{00}^{2}(x,y)=\sum_{m}\sum_{n} c_{00}^{2}(m,n)h_{hp}(x-m,y-n) \tag{4}$$

and

$$h_{hp}(x,y)=\begin{bmatrix}-1 & -1 & -1\\ -1 & 8 & -1\\ -1 & -1 & -1\end{bmatrix} \tag{5}$$

After the transformation, the texture features are derived by computing the sub-band energy of all wavelet coefficients:

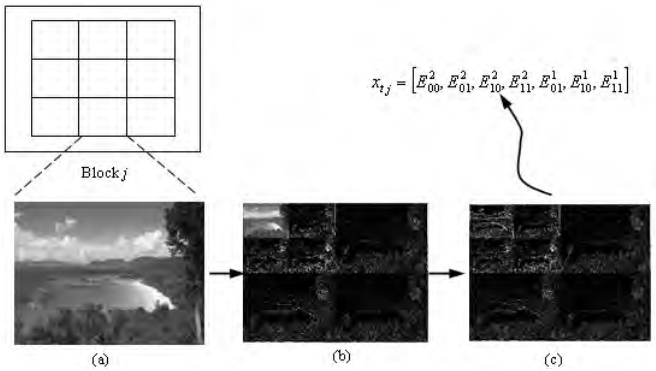

Fig. 3. Texture-feature extraction. (a) is original image ($L(x, y)$).
(b) is a two-level 5/3 wavelet transform. (c) coefficients c_{00}^{2} are transformed by high-filter.

$$E_{ij}^{l}=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|c_{ij}^{l}(m,n)^2\right| \tag{6}$$

where M and N are the dimensions of coefficients $c_{ij}^{l}(x,y)$. The seven-set wavelet texture features are extracted from a 3×3 tessellation segmented from the image as shown in Fig. 3. Texture features $x_{tj}=\left[E_{00}^{2},E_{01}^{2},E_{10}^{2},E_{11}^{2},E_{01}^{1},E_{10}^{1},E_{11}^{1}\right]$ are computed from Fig. 3(c), where j is the number of blocks in the region of interest.

2) Image Feature Extraction

Semantic features are those features most relevant to human vision. Therefore, in order to classify the image into the correct class of scene, we use SVM as a semantic predictor to transform the color and texture feature vectors x_{cj} and x_{tj}, which are extracted from nine block. Each block is labeled by its relative semantic concept by using an SVM semantic predictor. The input data of the SVM semantic predictor is x_c and x_t, and the output

data of the SVM is the class of testing block. Finally, the scene classifier gathers these semantic concepts to decide which scene class the image belongs to.

Although a semantic predictor can be used to establish a relationship between an original image and its semantic scene content, not all semantic-concept information is useful in determining whether or not a given image is a portrait or landscape scene. Consequently, this paper considers seven kinds of semantic concept: skin,

TABLE 1
COLOR CONFUSION MATRIX OF SVM SEMANTIC FEATURES CLASSIFICATION (Γ= 0.25,C=32)

classes	skin	sky	grass	water	snow	sand	tree
skin	88.9%	0	0	0	0	11.1%	0
sky	0	91.2%	0	5.8%	3.0%	0	0
grass	0	0	94.6%	0	0	0	5.4%
water	0	13.9%	0	86.1%	0	0	0
snow	0	0	0	0	44.4%	55.6%	0
sand	9.2%	0	0	0	0	90.8%	0
tree	0	0	8.3%	0	0	0	91.7%

TABLE 2
TEXTURE CONFUSION MATRIX OF SVM SEMANTIC FEATURES CLASSIFICATION (Γ= 2,C= 32768)

classes	skin	sky	grass	water	snow	sand	tree
skin	51.8%	35.2%	0	0	0	13.1%	0
sky	0	69.4%	0	0	3.5%	27.1%	0
grass	0	0	63.9%	0	0	8.3%	27.8%
water	0	0	0	80.6%	0	19.4%	0
snow	0	14.9%	4.9%	2.6%	43.8%	33.8%	0
sand	10.2%	0	0	0	3.4%	86.4%	0
tree	0	0	43.9%	0	0	0	56.1%

sky, grass, water, snow, sand, and tree. Those semantic features can be detected by using the color and wavelet texture features described in earlier sections. SVM is not the only classifier that can detect these semantic features (others classifier can be trained to infer the semantic information), but this paper uses SVM as it has been shown that it gives the best classification results [39]. The distance-measure method's classification results are not good enough. Although the event though distance measure has a better computational efficiency than SVM, SVM is nonetheless chosen as the semantic predictor. This paper uses the LIBSVM package [48] as the semantic-feature classifier. The package offers an efficient multi-class support by using internally a one-against-one approach [49]. The color and texture semantic-feature classification result from using the LIBSVM package is shown in Table 1 and Table 2, respectively.

3) Scene Classification

Digital cameras offer various useful modes, which are optimized for specific scenes and environmental conditions. These modes are preprogrammed by the manufacturer to automatically give the best exposure and settings for each scene. The classification system used by this paper tries to classify an image into three classes:

A) Portrait: A human is the subject and the background is not important.

B) Landscape: Images with a wide view. The camera automatically focuses on a distant object. The object can be sky, grass, water, and tree.

C) Beach/Snow: Photographs of beach or snow scenes. When the object is sand or snow, this image belongs to the beach or snow scene. Exposure and white balance are set to help prevent the scene from looking washed out.

If an image does not belong to three classes it will be classified as belonging to the "others" class.

After feature extraction and semantic-features analysis is run on an image, the semantic concept of each block is extracted. Although it is easy for humans to recognize semantic concepts, it is difficult for the machines to integrate the semantic-concept information from nine blocks and then classify the scene category accordingly. Therefore, if there were a method that could integrate the natural ability of humans to classify the category of an image, the accuracy rate of classification would be increased. The fuzzy rule-based classification system can fulfill this need.

First, we construct four fuzzy rules for PSROI classification system.

Rule R_1: IF one of the nine blocks is skin THEN the image is Portrait.

Rule R_2: IF one of the nine blocks is snow or sand THEN the image is Beach/Snow.

Rule R_3: IF one of the nine blocks is sky, grass, water, or tree THEN the image is Landscape.

Rule R_4: IF none of the blocks belong to any semantic-concept feature described in above THEN the image is Others.

Second, infer the scene category of the sample scene by following fuzzy rules to integrate the semantic concepts (color and texture have same weight) with the color and texture features from the nine blocks.

C. Integrate Focus Weight to Classification System

Focus systems in digital cameras are useful for users because they provide the option to focus on the object of interest. Generally, the object users focus on is the area users want to take photos of. The image of the object focused upon is the clearest area.

In order to make the classifying result meet users' needs, the result of the focus system is integrated into the nine blocks in the classification system. Assuming there are nine focus points in the focus system and each point is given a weight, called the Focus Weight, w_j, where $j = 1,2,...,9$. For example, if the object the user focus on is center w_5, w_5 gets more weight and others get the same weight. Then, each semantic feature from each block is multiplied by the focus weight. Finally, we get the classification result by applying the fuzzy rule-based classification system.

IV. EXPERIMENTAL RESULT AND ANALYSES

We implemented our classification system with a PC using an Intel Pentium M 1.7 GHz and 512 MB RAM. The software used is Borland C++ Builder 6.0 on Windows XP. In the training module, 3,150 image patches for all materials are trained by SVM for semantic-concept detection. In the testing module, 1,200 images of 320×240 pixels are collected from the Internet as sample images for the purpose of testing. The images are considered at two different angles, 0° (horizontal) and 90° (vertical). The classification result will be introduced in the following sections.

A. Image Database

Image regions for each material are cropped manually into the database to prepare the training data. These image patches are cropped in the shape of a rectangle of 80×60 pixels. Altogether, we have 3,150 image patches for all materials. There are seven different materials that are predefined. Each material is segmented by human semantic concepts. These seven classes are skin, sky, grass, water, snow, sand, and tree.

The testing set in the database includes 1,200 images collected from the Internet. These include 800 horizontal images and 400 vertical images. The images in the database are composed of portrait, landscape, and beach and snow scenes. We then subsample all images into 320×240 pixels for increased computational efficiency. Before we start testing the image, we need to divide all the images into four independent sets and put them in a different folder.

B. Horizontal Image Classification Result

In this section, we describe the horizontal image classification results using FRBS. A 320×240 horizontal image is segmented into 3×3 tessellation (ROI). The size of each block is 80×60. The classification accuracy of portraits is 94%. This result is given by combining color and texture features. As the result, the classification accuracy of color is better than that of texture in the portrait class. This is because the texture-classification result of skin material is easily confused with sky and sand as shown in Table 2. The portrait-classification success rate is very important because most people use digital cameras to picture children or other family members. The classification accuracy rates of landscape and beach/snow scenes are 86% and 87.5%, respectively. The overall accuracy rate is 89.17%. Although the classification accuracy, 89.17%, is not good enough to compete with the 90.1% result achieved in [2], this proposed classification system has a lower classification time. We test 200 images in each class, except beach/snow, and calculate the average time taken. For the class of beach/snow, the average time is 183 ms after testing 400 images. The overall computation time is 186.1 ms. This result is good, because most digital camera manufacturers limit the time of digital camera processing focus (S1) to less than 200 ms.

C. Vertical Image Classification Result

Vertical images are tested with vertical image patches. A 3×3 tessellation (ROI) of the vertical image is segmented prior to testing.

As experimental results, the color differs only slightly between vertical and horizontal images. In order to improve the texture-classification rate, 700 vertical image patches were used to train for the classification of vertical images. Although the ability of resisting the rotation image for wavelet texture transforms is not satisfactory, the classification rate of vertical images can still reach 66.67%. The overall computation time of vertical image is 183.6 ms.

Because of the impact of texture classification, the overall rate of classification accuracy for vertical images, 86.67%, is lower than that of horizontal images. The biggest difference is in the landscape class. This is because the texture of water is easy to confuse with snow and sand, and the number of vertical image patches are not enough.

D. Comparison With Alternative Approach

The proposed system is used in [5] to compare with PSROI. The main reason is all of them use low-level

TABLE 3
HORIZONTAL/ VERTICAL IMAGE CLASSIFICATION PRECISION OF [5] AND PSROI

Classification system		Portrait	Landscape	Beach /Snow	Overall	Classification time (ms)
horizontal	[5]	92%	79.5%	90%	87.2%	292.2
	PSROI	94%	86%	87.5%	89.2%	186.1
vertical	[5]	94%	78.6%	82%	84.8%	288
	PSROI	92%	82%	86%	86.7%	183.6

features (color and texture) to transform to relative semantic concepts. The system in [5], however, is used to classify images into indoor/outdoor classes. The system in [5] is modified and implemented to classify the images into portrait, landscape, or beach/snow. Of course, before testing these images, a training procedure is needed; this need is fulfilled by using the PSROI training database. The results of horizontal and vertical images of [5] and PSROI are listed in Table 3.

As shown in Table 3, the overall classification precision of PSROI is better than [5]. Additionally, the classification time in [5] requires 292.2 ms (horizontal image) while PSROI requires only 186.1 ms (horizontal image). The main reason is that the operation region of PSROI only covers a 56.25% area of the whole image, and thus needs less calculation time to achieve a better classification rate. The results in Table 3 show that PSROI not only reduces operation time but also achieves a good classification precision.

V. CONCLUSIONS

Current studies related to scene classification are all applied for image classification, identification, or retrieval from the backend. Few studies have investigated the application of scene classification in the process of taking photos. The scene-classification system considered herein is intended to allow users to take photos without changing the scene mode. It can integrate the focus result with the classification system to better fit a user's intention. Three pre-defined classes are set: portrait, landscape, and beach/snow. Experimental results show that the proposed system is capable of effectively classifying scenes into the correct categories within 0.2 s.

VI. REFERENCES

[1] S. Antani, R. Kasturi, and R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video," Pattern Recognition, vol. 35, issue 4, pp. 945-965, April 2002.

[2] J. Luo and A. Savakis, "Indoor vs outdoor classification of consumer photographs using low-level and semantic features," Proceedings of the 2001 International Conference On Image Processing (ICIP 01), Thessaloniki, Greece, vol. 2, pp. 745-748, Oct. 2001.

[3] L. Lu, K. Toyama and G. D. Hager, "A two level approach for scene recognition," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 20-25, Pp. 688-695, June 2005.

[4] F. Naccari, S. Battiato, A. Bruna, A. Capra, and A. Castorina, "Natural scenes classification for color enhancement," IEEE Transactions on Consumer Electronics, vol. 51, issue 1, pp. 234-239, Feb. 2005.

[5] N. Serrano, A. Savakis, and J. Luo, "Improved scene classification using efficient low-level features and semantic cues," Pattern Recognition 37(9), pp.1773-1784, Sep. 2004.

[6] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," IEEE transactions on neural networks, vol. 10, issue 5, pp. 1055-1064, Sep. 1999.

[7] L. Cinque, S. Levialdi, K. A. Olsen, and A. Pellicano, "Color-based image retrieval using spatial-chromatic histograms," IEEE International Conference on Multimedia Computing Systems, Florence, Italy, vol. 2, issue 5, pp. 969-973, 1999.

[8] A. Vailaya, A. Jain, and H. J. Zhang, "On image classification: city vs. landscape," Proceedings of the 1998 IEEE Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, CA, USA, pp. 3-8, June 1998.

[9] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), pp. 837-842, Aug. 1996.

[10] M. Turtinen and M. Pietikaine, "Visual training and classification of textured scene images," the 3rd International Workshop on Texture Analysis and Synthesis (Texture 2003), Nice, France, pp. 101-106. Oct. 2003.

[11] L. W. Renninger and J. Malik, "When is scene identification just texture recognition?" Vision Research, 44, pp. 2301- 2311, June 2004.

[12] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," Pattern recognition letters, vol. 24, pp. 1513-1521, 2003.

[13] J. Luo and A. E. Savakis, "Two-stage texture segmentation using complementary features," Proceedings of the 2000 International Conference on Image Processing, Vancouver, BC, vol. 3, pp. 564-567, Sept. 2000.

[14] K. M. Rajpoot, and N. M Rajpoot, "Wavelets and support vector machines for texture classification," Proceedings of 8th International Multitopic Conference, pp. 328-333, Dec. 2004.

[15] A. Oliva, A. Torralba, A. G. Dugue, and J. Herault, "Global semantic classification of scenes using power spectrum templates," Challenge of Image Retrieval (CIR99), Electronic Workshops in Computing Series, Sringer-Verlag, Newcastle, 1999.

[16] A. Torralba and A. Oliva, "Semantic organization of scenes using discriminant structural templates," The Proceedings of the Seventh International Conference on Computer Vision (ICCV99), Kerkyra, pp. 1253-1258, 1999.

[17] S. Foucher, V. Gouaillier, and L. Gagnon, "Global semantic classification of scenes using Ridgelet transform," Human vision and electronic imaging, Conference No9, San Jose CA , vol. 5292, pp. 402-413, Jan. 2004.

[18] M. Boutell, X. Shen, J. Luo, and C. Brown, "Multi-label semantic scene classification," Tech. Rep. 813, University of Rochester, Rochester, NY, Sept. 2003.

[19] X. Shen, M. Boutell, J. Luo, and C. Brown, "Multi-label machine learning and its application to semantic scene classification," International Symposium on Electronic Imaging, San Jose, CA, Jan. 2004.

[20] J. Luo and M. Boutell, "A probabilistic approach to image orientation detection via confidence-based integration of low-level and semantic cues," 4th International Workshop on Multimedia Data and Document Engineering (in conjunction with CVPR2004), Washington, DC, July 2004.

[21] G. H. Hu, J. J. Bu, and C. Chen, "A novel Bayesian framework for indoor-outdoor image classification," IEEE International Conference on Machine Learning and Cybernetics, vol. 5, pp. 3028-3032, Nov. 2003.

[22] M. Szummer and R.W. Picard, "Indoor-outdoor image classification," Proceedings of IEEE International Workshop on Content-based Access of Image and Video Databases, Bombay, India, pp. 42-51, 1998.

[23] C. Ko, H. S. Lee, and H. Byun, "Image retrieval using flexible image subblocks," Proceedings of the 2000 ACM symposium on Applied computing 2000, pp.574-578, March 2000.

[24] B. L. Saux and G. Amato, "Image classifiers for scene analysis," International Conference on Computer Vision and Graphics 2004.

[25] J. Luo, M. Boutell, R. T. Gray, and C. Brown, "Image transform bootstrapping and its applications to semantic scene classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 35, No. 3, June 2005.

[26] Vailaya, M. Figueiredo, A. Jain, and H. J. Zhang, "A Bayesian framework for semantic classification of outdoor vacation images," Proc. SPIE Storage Retrieval Image Video Databases VII, San Jose, CA, vol. 3656, pp. 415-426, Jan. 1999.

[27] J. Luo, A. E. Savakis, and A. Singhal, "A Bayesian network-based framework for semantic image understanding," Pattern Recognition, vol. 38, No. 6, pp. 919-934, June 2005.

[28] J. Luo, A. E. Aavakis, S. P. Etz, and A. Singhal, "On the application of Bayes networks to semantic understanding of consumer photographs," Proceedings of the 2000 International Conference on Image Processing, Vancouver, BC, vol. 3, pp. 512-515, Sept. 2000.

[29] Y. Tsin, R. T. Collins, V. Ramesh, and T. Kanade, "Bayesian color constancy for outdoor object recognition," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), vol. 1, pp. I-1132-I-1139, Dec. 2001.

[30] M. Boutell and J. Luo, "Bayesian fusion of camera metadata cues in semantic scene classification," IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, vol. 2, pp. 623-630, June 2004.

[31] M. Boutell and J. Luo, "Photo classification by integrating image content and camera metadata," Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, pp. 901-904, Aug. 2004.

[32] A. Singhal and J. Luo, "Probabilistic spatial context models for scene content understanding," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-235-I-241, June 2003.
[33] T. Ehtiati and J. J. Clark, "A strongly coupled architecture for contextual object and scene identification," Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 3, pp. 69-72, Aug. 2004.
[34] S. Kumar, A. C. Loui, and M. Hebert, "Probabilistic classification of image regions using an observation-constrained generative approach," ECCV Workshop on Generative Models based Vision (GMBV), pp. 91-99, 2002.
[35] M. Boutell and J. Luo, "A generalized temporal context model for semantic scene classification," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04), June 2004.
[36] M. Boutell and J. Luo, "Incorporating temporal context with content for classifying image collections," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04), vol. 2, pp. 947-950, June 2004.
[37] A. Bosch, X. Munoz, A. Olivr, and R. Marti, "Object and scene classification: what does a supervised approach provide us?" Proceeding of the 18th International Conference on Pattern Recognition (ICPR 2006), vol. 1, pp. 773-777, Aug. 2006.
[38] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H. J. Zhang, "Image classification for content-based indexing," IEEE Transactions on Image Processing, vol. 10, issue 1, pp. 117-130, Jan. 2001.
[39] J. Vogel and B. Schiele, "Semantic modeling of natural scenes for content-based image retrieval," International Journal of Computer Vision, 2004.
[40] A. Torralba and A. Oliva, "Statistics of natural image categories," Network, vol. 14, pp. 391-412, 2003.
[41] J. Luo and S. P. Etz, "A physical model-based approach to detecting sky in photographic images," IEEE Transactions of Image Processing, vol. 11, issue 3, pp. 201-212, Mar. 2002.
[42] P. Lipson, E. Grimson, and P. Sinha, "Configuration based scene classification and image indexing," Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1007-1013, Jun. 1997.
[43] B. L. Saux and G. Amato, "Image recognition for digital libraries," Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pp. 91-98, 2004.
[44] O. V. Kaick and G. Mori, "Automatic classification of outdoor images by region matching," The 3rd Canadian Conference on Computer and Robot Vision, June 2006.
[45] C. Nello and S. T. John, An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge, New York, 2000.
[46] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," IEEE Transactions on Fuzzy Systems, vol. 13, no. 4, pp. 428-435, Aug. 2005.
[47] J. Luo, R. T. Gray, and H. C. Lee, "Towards physics-based segmentation of photographic color images," Proceedings of the 1997 International Conference on Image Processing, vol. 3, pp. 58-61, Oct. 1997.
[48] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001, Software available at: http://www.csie.ntu.edu.tw.
[49] C. W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines," IEEE Transactions on Neural Networks, vol. 13(2), pp. 415-425, 2002.

AUTHORS PROFILE

Sheng-Fuu Lin (S'84–M'88) was born in Taiwan, R.O.C., in 1954. He received the B.S. and M.S. degrees in mathematics from National Taiwan Normal University in 1976 and 1979, respectively, the M.S. degree in computer science from the University of Maryland, College Park, in 1985, and the Ph.D. degree in electrical engineering from the University of Illinois, Champaign, in 1988. Since 1988, he has been on the faculty of the Department of Electrical and Control Engineering at National Chiao Tung University, Hsinchu, Taiwan, where he is currently a Professor. His research interests include image processing, image recognition, fuzzy theory, automatic target recognition, and scheduling

Huang-Tsun Chen was born in Taipei, Taiwan, R.O.C., in 1969. He received the M.S degree in institute of mechatronic engineering from National Taipei University of Technology, Taipei, Taiwan, R.O.C, in 199 He is currently pursuing the Ph. D. degree in the Department of Electrical and Control Engineering, the National Chiao Tung University, Hsinchu, Taiwan. His current research interests include image recognition, medicine image processing, machine vision.