Classification of Synthetic and Real Images Using Pattern Features

Myeong Hui Ha [O], Hyun jun Choi, Min Kook Choi, Sang
Chul Lee

myunghee90@hanmail.net, rene650962@gmail.com
mkchoi@inha.edu, sclee@inha.ac.kr

Abstract

Automatic classification of motion pictures has many application areas, and one prominent case is image search based on user query. In this particular application, user's intention for query is difficult to identify in the semantic level because image search algorithm generally exploits only features of image. In order to address this issue, we propose in this paper an automatic image classification algorithm applicable to synthetic and real motion pictures so that user's intention could be fully reflected at image search and classification. Feature-based or edge-based histogram descriptors produce poor results when applied to classification of synthetic and real images because this is inter-class classification. The algorithm proposed in this paper obtains feature vectors of color distribution pattern in motion pictures to classify synthetic and real images. In our experiment, the proposed algorithm was able to classify images with the accuracy around 74%.

1. Introduction

The majority of image classification algorithms are intra-class classification methods and use low-level visual features or text features to distinguish images of different categories, such as shots of mountains, beach, sunset scenery, or urban landscape [1].

Existing work [1, 2] generally do not distinguish similar images with different semantic meanings. Similar images are put into the same categories and it is hard to provide service that fit user query that have different semantic intentions. Hence, it is necessary to implement a classification method tailored for semantic classification. One example of classification within the same category is distinguishing synthetic images and real images. Lienhart and Hertman [3] was among the first to explore this topic; their algorithm distinguished graphical images and photo-like images, which were further classified into true photos and ray-traced/rendered images. Other image classification algorithms such as SIFT (scale invariant feature transform), SURF (speeded-up robust features), BRISK (binary robust invariant scalable keypoints), and GLOH (gradient location and orientation histogram) are feature-based descriptors [4, 5, 6, 7]; HoG (histogram of oriented gradients) exploits edge histogram-based features [8]. Existing feature descriptors show great performance for distinguishing images of different categories but not for images within the same category. This weakness is caused by the fact that image descriptors are based on object edges. here have been attempts to improve on this weakness by analyzing images through CBIR (content-based image retrieval) [9] and Garbarino's work [10] that exploits textual feature. In this paper, we propose an algorithm for synthetic and real images by combining widely used feature descriptors with color patterns and feature distribution analysis. We chose four image classes of apple, human face, landscape, and buildings, all of which can be processed with binary classification. With the proposed algorithm, it is possible to implement a search/classification system that can meet user query in the semantic level.

Figure 1 illustrates the overall structure of the proposed algorithm.
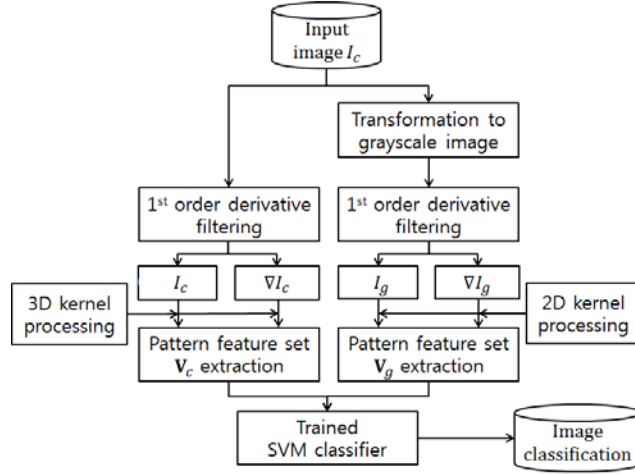
[Figure 1]

2. Pattern Feature Extraction

In this paper, we propose a pattern-based feature extraction algorithm that classify synthetic and real images. Unlike feature descriptors and histogram-based features, pattern features of images considers the distribution of pixel strengths within local regions. Such pattern data is based on the type of color distribution and local texture data, and this produces improved classification results.

2.1 Two-dimensional Pattern Extraction

When the input images has been entered as an $m \times n$ tensor $Ic \in \mathbb{N}m \times n \times 3$, then converting to black-and-white color scheme yields $Ig \in \mathbb{N}m \times n$.

To find the pattern vector $\boldsymbol{Vg}$ for $Ig \in \mathbb{N}m \times n$, a kernel array set $\mathbf{f}w \in \{fw,1, fw,2, \ldots, fw,6\}$ is defined. Here, w is the parameter for the size of the kernel and it can be in four different forms. Equation (1) is an example when w=3 for 2-dimensional pattern extraction kernel.

$$f_{3,1} = \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}, \quad f_{3,2} = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix}$$

$$f_{3,3} = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad f_{3,4} = \begin{bmatrix} -1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

$$f_{3,5} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}, \quad f_{3,6} = \begin{bmatrix} 0 & -2 & -1 \\ 2 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix} \quad (1)$$

In order to reflect a wide range of characteristics in the images, first-order differential image $\triangledown Ig$ for the black-and-white image Ig is calculated.

Let the first-order differential images of Ig in the horizontal and vertical directions be (Ig,x, Ig,y) then the change of pixel at each pixel coordinates (u, v) for (Ig,x, Ig,y) is defined as follows:

$$Ig,x(u,v) = Ig(u,v) - Ig(u+1,v)$$
$$Ig,y(u,v) = Ig(u,v) - Ig(u,v+1)$$

Pixel strength at coordinates (u, v) is denoted by $Ig(u,v)$. Here, the first-order differential image can be defined as:

$$\triangledown Ig = Ig, x + Ig, y.$$

In order to obtain pattern features from two images of Ig and $\triangledown Ig$, $\mathbf{f}w$ is applied, which can be in one of three different forms. The first form is applied to the ith element of $\mathbf{f}w$, or fw,i as a convolution calculation:

$$b1,i = Ig * fw,i \quad (2)$$
$$b2,i = \triangledown Ig * fw,i \quad (3).$$

The term bi is the convolution calculation result for the i th kernel defined within the range of $1 \leq i \leq 6$. Following Equations (2) and (3), the feature vector $\mathbf{v}1,(u,v) = [b1,1,\ldots,b1,s,b2,1,\ldots,b2,s]T$ is extracted on the pixel at coordinates (u,v). With the type of kernel and input image, we obtain s=6. Secondly, $\mathbf{f}w$ is used for extracting the statistical parameters for the ith element fw,i. The term $b3,i(u,v)$ is defined as the weighted average for pixel Ig on coordinates (u,v) in the kernel matrix fw,i. Next, the term $b4,i(u,v)$ is the weighted average for pixel $\triangledown Ig$ on coordinates (u,v) in the kernel matrix fw,i, and it is used for extracting feature vector $\mathbf{v}2,(u,v)$. The third form of pattern extraction for the ith element $f_{w,i}$ in $\mathbf{f}_w$ also is a statistical parametric form The term $b5,i(u,v)$ is defined as the weighted standard deviation for the pixel Ig on the kernel matrix fw,i on coordinates (u,v). We also define $b_{4,i}(u,v)$ as the weighted standard deviation for the pixel $\triangledown I_g$, and this is used to extract the feature vector $\mathbf{v}_{3,(u,v)}$.

In short, the algorithm performs $Ic \rightarrow Ig$ black-and-white conversion for the 2-dimensional input image Ic and produces the first-order differential image $\triangledown Ig$ in order to extract pattern features.

With set operations of $\mathbf{f}w$ kernel matrix on the converted input image Ig and $\triangledown Ig$, pattern vector set $\mathbf{V} \in \{\mathbf{v}1, \mathbf{v}2, \ldots, \mathbf{v}k\}$ is obtained, where k is the total number of feature vectors for the input image IC and the ith element of set $\mathbf{V}$ is $\mathbf{v}i = [b1,b2,\ldots,bs]T$. The letter s denotes the dimension of each feature vector. If we follow the kernel matrix of Equation (1) in this case, a feature vector set $\mathbf{V}$ with k=3 and s=12 is produced.

2.2 Three-dimensional Pattern Feature Extraction

Three-dimensional pattern feature extraction is performed on the color input image Ic. Following the same principles as in the 2-dimensional pattern feature extraction, we define a 3-dimensional kernel trifocal tensor set $\mathbf{f}q' \in \{fq,1', fq,2', \ldots, fq,6'\}$, which is an extension from the 2-dimensional kernel in Equation (1), in order to find the pattern vector set $\mathbf{V}'$ for the input image Ic. The extended 3-dimensional kernel tensor $fq,1'$ as an extension from the kernel matrix $fw,1$ in Equation (1) is defined as follows:

$$f'_{q,1,1} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}, \quad f'_{q,1,2} = \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix},$$

$$f'_{q,1,3} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \quad (2)$$

The index i in the trifocal tensor $f_{q,1,i'}$ denotes the matrix plane within the tensor and it consists of extended $\mathbf{f}w \rightarrow \mathbf{f}q'$ as in Equation (1). The first-order differential image $\triangledown Ic$ can be defined as follows for each color plane Ic,r, Ic,g, Ic,b in the color image Ic:

$$\nabla I_{c,r} = I_{c,r,x} + I_{c,r,y}$$
$$\nabla I_{c,g} = I_{c,g,x} + I_{c,g,y}$$
$$\nabla I_{c,b} = I_{c,b,x} + I_{c,b,y}$$

Here, $\nabla Ic,r$, $\nabla Ic,g$, and $\nabla Ic,b$ denote the first-order differential image on each of the RGB color planes. The equation to extract a pattern feature vector set $\mathbf{V}'$ from 3-dimensional image Ic and ∇Ic extends the equation for 2-dimensional image. Similar to the 2-dimensional case, $\mathbf{v}'4,(u,v) = [b7,1,..,b7,s,b8,1..,b8,s]T$ is found with $b7,i$ and $b8,i$; $\mathbf{v}'5,(u,v) = [b9,1, . . , b9,s, b10,1. . , b10,s]T$ with $b9,i$ and $b10,i$; $\mathbf{v}'6,(u,v) =[b11,1, . . , b11,s, b12,1. . , b12,s]T$ with $b11,i$ and $b12,i$. Following the type of kernel and input image, we obtain $s = 12$. Three-dimensional input image Ic is processed to obtain ∇Ic similarly as in processing 2-dimensional pattern features, and a pattern vector set $\mathbf{V}' \in \{\mathbf{v}1',\mathbf{v}2',...,\mathbf{v}k'\}$ is extracted by performing set operations with $\mathbf{f}q'$ kernel tensor set on the input image Ic and ∇Ic. The i th element in $\mathbf{V}'$, is defined as $\mathbf{v}i' = [b1,b2,..,bs]T$; following the kernel matrix of Equation (1), we obtain a feature vector set $\mathbf{V}'$ where $k=3$ and $s=12$.

## 3. Implementation of Pattern Feature-based Classifier

From the feature vector sets $\mathbf{V}$ and $\mathbf{V}'$ extracted from input image dataset, we implemented an SVM classifier, which is defined as a Lagrange optimization problem as in Equation (3), which finds a hyperplane with maximum margin:

$$\max_{\alpha} L(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i y_i)(\alpha_j y_j) \mathbf{K}(b_i, b_j)$$
$$\text{s.t. } \mathbf{y}\boldsymbol{\alpha} = 0, 0 \leq \alpha_i \leq \frac{1}{n} \quad (3)$$

The class label variable y in this paper satisfies $y \in \{-1,1\}$; b is the feature vector of extracted pattern features.

4. Experiment and Analysis

To test the performance of our proposed algorithm, we downloaded from the Internet images of apples, people, landscapes, and buildings to distinguish real and synthetic images. In [Figure 2], examples of images selected are shown.

(a)　　　　　(b)

(c)　　　　　(d)
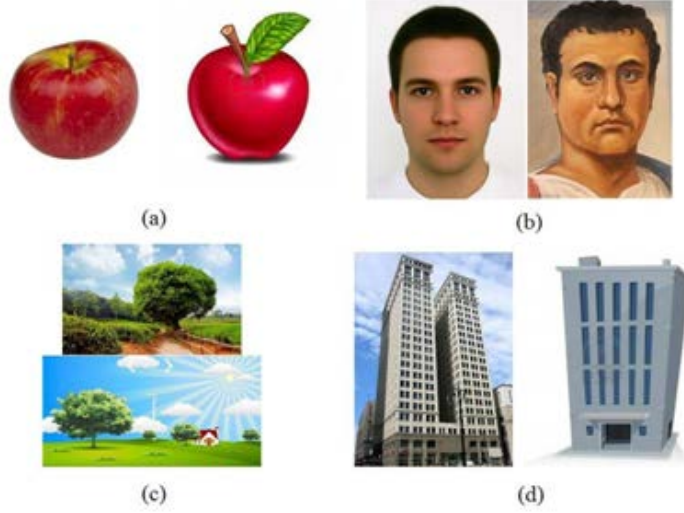
For algorithm evaluation, we collected 100 files for synthetic and real images each and performed 5-fold cross validation. We trained our classifier with 80 synthetic and 80 real images for each category and tested the classifier with 20 synthetic and 20 real images. We made 21 different composition of training and test images for experiment and calculated precision and recall values:

$$precision = \frac{tp}{tp + fp}$$
$$recall = \frac{tp}{tp + fn}$$

apple	real	synthetic	precision
Real	144	66	0.686
synthetic	137	73	0.348
recall	0.512	0.525	0.517
			0.519

face	real	synthetic	precision
real	189	21	0.9
synthetic	89	121	0.576
recall	0.68	0.852	**0.738**
			0.766

landscape	real	synthetic	precision
real	127	83	0.605
synthetic	139	71	0.338
recall	0.477	0.461	0.471
			0.469

building	real	synthetic	precision
real	169	41	0.804
synthetic	124	83	0.41
recall	0.577	0.677	0.607
			0.627

[Table 1]

Here, tp, fp, fn denote true positive, false positive, and false negative, respectively. In [Table 1], precision and recall results for each category are shown.

5. Conclusion

In this paper we proposed a classification method for images in the same category. Exploiting pattern features (texture) showed improved classification accuracy over other classification methods based on points, edges. In our future work, we will advance from the current form of kernels and extract more detailed feature extraction and different types of statistics data for localized areas in images.

Acknowledgements

References

[1] Aditya Vailaya, Mario A. T. Figueiredo, Anil K. Jain, HongJiang Zhang. "Bayesizn framework for semantic classification of
outdoor vacation images." In Proc. Storage and Retrieval for Image and Video Databases, vol. 3656, 1999.

[2] John R. Smith, Shih-Fu Chang. "Image and Video search engine for the World Wide Web." In Proc. Storage and Retrieval for Image and Video Databases, vol. 3022, 1977.

[3] Rainer Lienhart, Alexander Hartmann. "Classifying images on the web automatically." Journal of Electronic Imaging, vol.11, no.4, pp. 445-454, 2002.

[4] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, 2008.

[6] K. Mikolajczyk and C. Schmid "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, 10, 27, pp 1615--1630, 2005.

[7] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints", in Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2011.

[8] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection", In Proc. of CVPR, 2005.

[9] Fei Wang, Min-Yen Kan. "NPIC: Hierarchical Synthetic Image Classification Using Image Search and Generic Features." Image and Video Retrieval Lecture Notes in Computer science, vol. 4071, pp. 473-482, 2006.

[10] Maurizio Garbarino. "Automatic Classification of Natural and Synthetic Images" Master of Science Thesis Stockholm, Sweden, 2008