

Classification of Synthetic and Real Images Using Pattern Features

Myeong Hui Ha¹, Hyun jun Choi¹, Min Kook Choi², and Sang Chul Lee³

^{1,2,3}Department of Computer Science, Inha University, Incheon, South Korea

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, the users intention for query is difficult to identify at the semantic level because image search algorithms generally exploit only features of images. 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 a users intention can be fully reflected in the 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 standing at around 74%.

I. INTRODUCTION

The majority of image classification algorithms are intra-class classification methods and use low-level visual features or text features to distinguish images in different categories, such as shots of mountains, beaches, sunset scenes or urban landscapes [2].

Existing work [1, 2] generally does not distinguish similar images with different semantic meanings. Similar images are put into the same categories and it is hard to provide a service that fits the user query that has 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, 6, 7, 10]; HoG (histogram of oriented gradients) exploits edge histogram-based features [5]. 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. There have been attempts to improve on this weakness by analyzing images through CBIR (content-based image retrieval) [8] and Garbarinos work [9], which exploits textual feature. In this paper, we propose an algorithm for both synthetic and real images by combining widely used feature descriptors with color patterns and feature distribution analysis. We chose four image classes of an apple, a human face, landscapes 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 the user query at the semantic level.

II. PATTERN FEATURE EXTRACTION

In this paper, we propose a pattern-based feature extraction algorithm that classifies both synthetic

Myeong Hui Ha email address: myunghee90@hanmail.net

Hyun jun Choi email address: rene650962@gmail.com

Min Kook Choi email address: mkchoi@inha.edu

Sang Chul Lee email address: sclee@inha.ac.kr

*This paper is a translation from the original, written in Korean. The two papers have the same content.

and real images. Unlike feature descriptors and histogram-based features, pattern features of images consider 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.

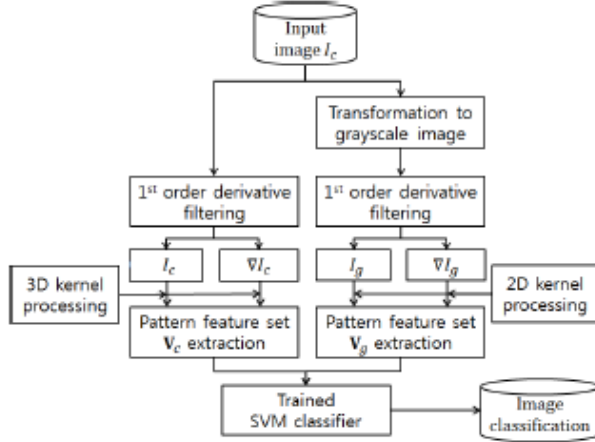


FIG 1: An illustration of the overall structure of the proposed algorithm.

A. Two-dimensional Pattern Extraction

The input images have been entered as an $m \times n$ tensor $I_c \in \mathbb{N}^{m \times n \times 3}$, then converted to black-and-white color scheme yields $I_g \in \mathbb{N}^{m \times n}$. To find the pattern vector V_g for $I_g \in \mathbb{N}^{m \times n}$ a kernel array set $f_w \in \{f_{w,1}, f_{w,2}, \dots, f_{w,6}\}$ 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.

$$\begin{aligned} f_{3,1} &= \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} & f_{3,2} &= \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \\ f_{3,3} &= \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} & f_{3,4} &= \begin{bmatrix} -1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \\ f_{3,5} &= \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} & f_{3,6} &= \begin{bmatrix} 0 & -2 & -1 \\ 2 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix} \quad (1) \end{aligned}$$

In order to reflect a wide range of characteristics in the images, first-order differential image ∇I_g for the black-and-white image I_g is calculated. Let the first-order differential images of I_g in the horizontal and vertical directions be $(I_{g,x}, I_{g,y})$ then the change of pixel at each pixel coordinates (u, v) for $(I_{g,x}, I_{g,y})$ is defined as follows:

$$\begin{aligned} I_{g,x}(u, v) &= I_g(u, v) - I_g(u + 1, v) \\ I_{g,y}(u, v) &= I_g(u, v) - I_g(u, v + 1) \end{aligned}$$

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

$$\nabla I_g = I_{g,x} + I_{g,y}$$

In order to obtain pattern features from two images of I_g and ∇I_g , f_w is applied, which can be in one of three different forms. The first form is applied to the i th element of f_w , or $f_{w,i}$ as a convolution calculation:

$$b_{1,i} = I_g * f_{w,i} \quad (2)$$

$$b_{2,i} = \nabla I_g * f_{w,i} \quad (3)$$

The term b_i 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 $v_1(u, v) = [b_{1,1}, \dots, b_{1,s}, b_{2,1}, \dots, b_{2,s}]^T$ is extracted on the pixel at coordinates (u, v) . With the type of kernel and input image, we obtain $s=6$. Secondly, f_w is used for extracting the statistical parameters for the i th element $f_{w,i}$. The term $b_{3,i}(u, v)$ is defined as the weighted average for pixel I_g on coordinates (u, v) in the kernel matrix $f_{w,i}$. Next, the term $b_{4,i}(u, v)$ is the weighted average for pixel ∇I_g on coordinates (u, v) in the kernel matrix $f_{w,i}$, and it is used for extracting feature vector $v_2(u, v)$. The third form of pattern extraction for the i th element $f_{w,i}$ in f_w is also a statistical parametric form. The term $b_{5,i}(u, v)$ is defined as the weighted standard deviation for the pixel I_g on the kernel matrix $f_{w,i}$ on coordinates (u, v) . We also define $b_{4,i}(u, v)$ as the weighted standard deviation for the pixel ∇I_g , and this is used to extract the feature vector $v_3(u, v)$. In short, the algorithm performs $I_c \rightarrow I_g$ black-and-white conversion for the two-dimensional input image I_c and produces the first-order differential image ∇I_g in order to extract pattern features.

With set operations of f_w kernel matrix on the converted input image I_g and ∇I_g , pattern vector set $V \in \{v_1, v_2, \dots, v_k\}$ is obtained, where k is the total number of feature vectors for the input image I_c and the i th element of set V is $v_i = [b_1, b_2, \dots, b_s]^T$. 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 V with $k=3$ and $s=12$ is produced.

B. Three-dimensional Pattern Feature Extraction

The three-dimensional pattern feature extraction is performed on the color input image I_c . Following the same principles as in the two-dimensional pattern feature extraction, we define a three-dimensional kernel trifocal tensor set $f'_q \in \{f'_{q,1}, f'_{q,2}, \dots, f'_{q,6}\}$, which is an extension from the two-dimensional kernel in equation (1), in order to find the pattern vector set V' for the input image I_c . The extended three-dimensional kernel tensor $f'_{q,1}$ as an extension from the kernel matrix $f_{w,1}$ 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 $f_w \rightarrow f'_q$ as in equation (1). The first-order differential image ∇I_c can be defined as follows for each color plane $I_{c,r}$, $I_{c,g}$, $I_{c,b}$ in the color image I_c :

$$\begin{aligned} \nabla I_{c,r} &= I_{c,r,x} + I_{c,r,y} \\ \nabla I_{c,g} &= I_{c,g,x} + I_{c,g,y} \\ \nabla I_{c,b} &= I_{c,b,x} + I_{c,b,y} \end{aligned}$$

Here, $\nabla I_{c,r}$, $\nabla I_{c,g}$, and $\nabla I_{c,b}$ denote the first-order differential image on each of the RGB color planes. The equation to extract a pattern feature vector set V' from a three-dimensional image I_c and ∇I_c extends the equation for the two-dimensional image. Similar to the two-dimensional case, $v'_4(u, v) = [b_{7,1}, \dots, b_{7,s}, b_{8,1}, \dots, b_{8,s}]^T$ is found with $b_{7,i}$ and $b_{8,i}$; $v'_5(u, v) = [b_{9,1}, \dots, b_{9,s}, b_{10,1}, \dots, b_{10,s}]^T$ with $b_{9,i}$ and $b_{10,i}$; $v'_6(u, v) = [b_{11,1}, \dots, b_{11,s}, b_{12,1}, \dots, b_{12,s}]^T$ with $b_{11,i}$ and $b_{12,i}$. Following the type of kernel and input image, we obtain $s = 12$. The three-dimensional input image I_c is processed to obtain ∇I_c similarly as in processing two-dimensional pattern features, and a pattern vector set $V' \in v'_1, v'_2, \dots, v'_k$ is extracted by performing set operations with f'_q kernel tensor set on the input image I_c and ∇I_c . The i th element in V' , is

defined as $v'_i = [b_1, b_2, \dots, b_s]^T$; following the kernel matrix of equation (1), we obtain a feature vector set V' where $k=3$ and $s=12$.

III. IMPLEMENTATION OF PATTERN FEATURE-BASED CLASSIFIER

From the feature vector sets V and V' extracted from the 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 margins:

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

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

IV. 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 synthetic and real images. In [Figure 2], examples of the images selected are shown.

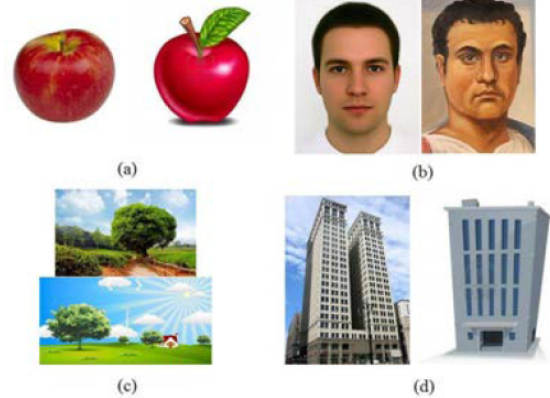


FIG 2: four kinds of synthetic/real images for image classification (a) apple, (b) face, (c) landscape, and (d) building

For algorithm evaluation, we collected 100 files for each category of synthetic and real images each and performed a 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 compositions of training and test images

for experiment and calculated precision and recall values:

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

$$\text{precision} = \frac{tp}{tp+fp}$$

$$\text{recall} = \frac{tp}{tp+fn}$$

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 I: precision-recall by category.

V. 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 and 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.

VI. ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) and the grant was funded by the Korea government(MSIP) (No. 10047078).

REFERENCES

- [1] John R Smith and Shih-Fu Chang. “Image and video search engine for the world wide web”. In: *Storage and Retrieval for Image and Video Databases V*. Vol. 3022. International Society for Optics and Photonics. 1997, pp. 84–95.
- [2] Aditya Vailaya et al. “Bayesian framework for semantic classification of outdoor vacation images”. In: *Storage and Retrieval for Image and Video Databases VII*. Vol. 3656. International Society for Optics and Photonics. 1998, pp. 415–426.
- [3] Rainer W Lienhart and Alexander Hartmann. “Classifying images on the web automatically”. In: *Journal of electronic imaging* 11.4 (2002), pp. 445–455.
- [4] David G Lowe. “Distinctive image features from scale-invariant keypoints”. In: *International journal of computer vision* 60.2 (2004), pp. 91–110.
- [5] Navneet Dalal and Bill Triggs. “Histograms of oriented gradients for human detection”. In: 2005.
- [6] Krystian Mikołajczyk and Cordelia Schmid. “A performance evaluation of local descriptors”. In: (2005).
- [7] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. “Surf: Speeded up robust features”. In: *European conference on computer vision*. Springer. 2006, pp. 404–417.
- [8] Fei Wang and Min-Yen Kan. “NPIC: Hierarchical synthetic image classification using image search and generic features”. In: *International Conference on Image and Video Retrieval*. Springer. 2006, pp. 473–482.
- [9] Maurizio Garbarino. *Automatic classification of natural and synthetic images*. Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2008.
- [10] Stefan Leutenegger, Margarita Chli, and Roland Siegwart. “BRISK: Binary robust invariant scalable keypoints”. In: *2011 IEEE international conference on computer vision (ICCV)*. Ieee. 2011, pp. 2548–2555.

패턴 특징을 활용한 가상 영상 분류 알고리즘 연구

하명희[○], 최현준, 최민국, 이상철

인하대학교 컴퓨터정보공학과

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

요 약

자동화된 영상물 분류 알고리즘은 많은 응용분야에서 활용되고 있다. 대표적인 응용분야 중 하나로 사용자 질의에 의한 영상 검색 및 분류가 있는데, 이 경우 대부분 영상의 특징 기반의 검색을 수행하기 때문에 의미적인 단계에서의 사용자 질의 의도를 파악하기 힘들다는 단점이 있다. 본 논문은 이러한 영상 검색 및 분류 문제에서 사용자가 원하는 의미적 단계의 분류를 달성하기 위해 가상 영상과 실제 영상을 자동화된 방법으로 분류하는 알고리즘을 제안한다. 이러한 가상 영상과 실제 영상의 분류는 기존의 카테고리 별 분류와 다르게 동일한 물체 간 카테고리 간 분류(inter class classification)로 특징점 기반의 기술자나, 에지 기반의 히스토그램 기술자로 좋은 성능을 얻기가 어렵다. 그러므로 본 논문에서는 영상 내에 존재하는 컬러의 분포 패턴을 통해 얻어진 특징 값들로 구성된 특징벡터를 제안하여 가상 영상과 실제 영상의 분류를 달성하고 있다. 본 논문에서 제안한 패턴 기반의 영상 특징으로 약 74% 분류 정확도를 달성하였다.

1. 서론

기존의 영상물 분류 연구는 저 수준 시각 특징 (low-level visual feature)이나 텍스처 특징 (texture feature)을 이용한 서로 다른 카테고리간의 분류(intra class classification)에 집중되어 있다. 예를 들면 산, 바다, 노을 지는 풍경, 도시 등등의 서로 다른 카테고리의 사진들을 분류하는 연구[1]가 그것이다.

카테고리 간 분류에 집중되어 있는 기존의 영상물 분류 연구[1, 2] 방법으로는 서로 다른 의미를 갖고 있는 비슷한 영상을 분류하기 어렵다. 비슷한 영상은 같은 카테고리 분류되기 때문에 사용자 질의에 부합하는 의미적 단계의 서비스는 제공하기 어렵다. 그러므로 좀 더 의미적인 검색 및 분류를 달성하기 위해서는 좀 더 의미적 단계에서의 분류 과업(classification task)를 달성할 필요성이 있다.

이러한 의미적 단계에서의 카테고리 내 분류의 대표적인 예가 가상영상과 실제 영상의 분류이다. 가상영상과 실제영상의 분류의 대표적인 초기 연구는 Lienhart 와 Hertman 의 연구[3]로써, 먼저 가상영상 (graphical image)들과 실제 사진과 유사한 영상 (photo-like image)들을 구분해 내고, 그 다음 단계에서 실제 사진과 유사한 영상(photo-like image)들 중에서도 실제 영상(true photo)와 광선추적영상/렌더링된 영상(ray-traced/rendered image)들을 분류하는 알고리즘을 제안하였다.

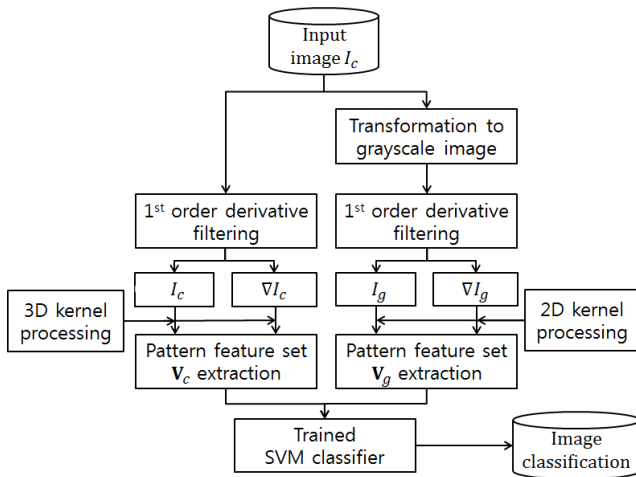
과거에는 일반적인 카테고리 간 영상물 분류를 위해서는 SIFT (Scale Invariant Feature transform),

SURF (speeded up robust features), BRISK (binary robust invariant scalable keypoints), GLOH (gradient location and orientation histogram) 등의 특징점 기반의 기술자[4, 5, 6, 7]나 HoG(Histogram of oriented gradients)등의 에지 히스토그램 기반의 특징[8]들을 사용해 왔다.

기존의 특징 기술자들은 카테고리 간 분류에서는 훌륭한 성능을 나타내지만 알고리즘의 내재적 한계로 카테고리 내 분류에 있어서는 취약하다. 그 이유는 카테고리 간 영상물 분류에 적합한 특징 기술자들은 대부분 물체의 에지를 기반으로 추출하기 때문에 유사한 물체 간의 분류 정확도가 낮아질 수밖에 없다.

이러한 문제점을 극복하기 위해 가상/실제 영상 분류를 위한 연구들로는 이미지 분석을 통한 CBIR (Contents based image retrieval) 을 위한 연구[9]와 textual feature 를 사용한 Garbarino 의 연구[10] 등이 제안되었다.

본 연구에서는 카테고리 내 분류의 대표적인 응용분야인 가상영상/실제영상 분류를 위해 기존에 널리 사용되는 특징 기술자와 다른 영상 내 컬러 패턴과 분포 특징 분석을 통한 분류 알고리즘을 제안하고 있다. 4 개의 클래스는 각각 사과, 사람 얼굴, 풍경, 건물로 구성되어 있으며, 각각은 이진분류 (binary classification) 문제로 생각할 수 있다. 제안된 알고리즘을 통해 의미적 단계에서 사용자 질의를 만족할 수 있는 시스템 구축이 가능하고, 좀 더 지능적인 서비스를 제공할 수 있다. 다음 [그림 1]은 제안하는 알고리즘의 흐름도를 나타내고 있다.



[그림 1] 가상/실제 영상 분류를 위한 알고리즘의 순서도

2. 패턴 특징 추출 기법

본 논문에서는 가상/실제 영상 분류를 위해 패턴 기반의 특징 추출 알고리즘을 제안한다. 영상의 패턴 특징은 특징점 기술자 및 히스토그램 기반의 특징과 다르게 영상의 특정한 지역적 영역 안에서 픽셀 강도의 분포가 어떤 형태를 이루는가에 대한 정보를 기반으로 한다. 이러한 패턴 정보는 컬러 분포의 형태와 특정 영역 내의 텍스처 정보를 기반으로 하기 때문에 클래스 내 분류에 있어서 보다 나은 성능을 보여줄 수 있다.

2.1 2차원 패턴 추출

입력 영상이 컬러 정보를 담은 $m \times n$ 크기의 텐서 $I_c \in \mathbb{N}^{m \times n \times 3}$ 로 들어왔을 경우 흑백 영상 변환을 통해 $I_g \in \mathbb{N}^{m \times n}$ 를 얻어 낼 수 있다.

먼저 $I_g \in \mathbb{N}^{m \times n}$ 의 경우에 대한 패턴 벡터 집합 V_g 를 얻기 위해 다음과 같은 형태의 커널 행렬 집합 $f_w \in \{f_{w,1}, f_{w,2}, \dots, f_{w,6}\}$ 을 정의한다. w 는 커널의 크기에 대한 매개변수이며, 총 4개의 형태로 이루어져 있다. 식 (1)은 2차원 패턴 추출을 위한 커널의 $w = 3$ 인 경우 예시를 보여주고 있다.

$$\begin{aligned} f_{3,1} &= \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}, & f_{3,2} &= \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix} \\ f_{3,3} &= \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, & f_{3,4} &= \begin{bmatrix} -1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \\ f_{3,5} &= \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}, & f_{3,6} &= \begin{bmatrix} 0 & -2 & -1 \\ 2 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix} \quad (1) \end{aligned}$$

또한 영상 내에 다양한 특성을 반영하기 위해 흑백영상 I_g 에 대한 1차 미분영상 ∇I_g 를 연산한다. I_g 의 수평/수직 방향의 1차 미분 영상을 $(I_{g,x}, I_{g,y})$ 로 나타냈을 때, $(I_{g,x}, I_{g,y})$ 의 각 픽셀 위치 (u, v) 에서의

픽셀 변화량 값은 다음과 같이 정의한다.

$$\begin{aligned} I_{g,x}(u, v) &= I_g(u, v) - I_g(u + 1, v) \\ I_{g,y}(u, v) &= I_g(u, v) - I_g(u, v + 1) \end{aligned}$$

$I_g(u, v)$ 은 특정한 픽셀 위치 (u, v) 에서의 픽셀 강도 값을 의미한다.

이때 1차 미분 영상은 다음과 같이 정의된다.

$$\nabla I_g = I_{g,x} + I_{g,y}$$

두 영상 I_g 와 ∇I_g 로부터 패턴 특징을 얻기 위해 f_w 를 적용한다. f_w 는 총 세 가지 형태로 적용된다.

첫 번째는 f_w 의 i 번째 원소 $f_{w,i}$ 에 대한 컨볼루션 연산으로 적용된다.

$$\begin{aligned} b_{1,i} &= I_g * f_{w,i} \quad (2) \\ b_{2,i} &= \nabla I_g * f_{w,i} \quad (3) \end{aligned}$$

b_i 는 i 번째 커널에 대한 컨볼루션 연산이 적용된 결과 행렬로 $1 \leq i \leq 6$ 범위에서 정의된다. 식 (2)와 (3)에 의해 특정한 픽셀 위치 (u, v) 상에서 특징 벡터 $\mathbf{v}_{1,(u,v)} = [b_{1,1}, \dots, b_{1,s}, b_{2,1}, \dots, b_{2,s}]^T$ 를 추출하게 되고 커널 종류와 입력 이미지에 따라 $s = 6$ 의 값을 갖게 된다.

두 번째로 f_w 의 i 번째 원소 $f_{w,i}$ 에 대한 통계적 파라미터 추출의 형태로 적용된다. $b_{3,i}(u, v)$ 은 (u, v) 상에서 $f_{w,i}$ 커널 행렬의 I_g 픽셀 값에 대한 가중치 평균(weighted average)으로 정의된다. 또한 $b_{4,i}(u, v)$ 은 (u, v) 상에서 $f_{w,i}$ 커널 행렬의 ∇I_g 픽셀 값에 대한 가중치평균으로 정의되며, 특징 벡터 $\mathbf{v}_{2,(u,v)}$ 를 추출하게 된다.

f_w 의 i 번째 원소 $f_{w,i}$ 에 대한 세 번째 형태의 패턴 추출 또한 통계적 파라미터 추출의 형태로 적용된다. $b_{5,i}(u, v)$ 은 (u, v) 상에서 $f_{w,i}$ 커널 행렬의 I_g 픽셀 값에 대한 가중치표준편차(weighted standard deviation) 값으로 정의된다. $b_{4,i}(u, v)$ 은 (u, v) 상에서 $f_{w,i}$ 커널 행렬의 ∇I_g 픽셀 값에 대한 가중치표준편차 값으로 정의되며, 특징 벡터 $\mathbf{v}_{3,(u,v)}$ 을 추출하게 된다.

결과적으로 패턴 특징 추출을 위해 2차원 입력 영상 I_c 에 대해 $I_c \rightarrow I_g$ 흑백 변환을 수행하고, I_g 에 대한 1차 미분영상 ∇I_g 를 얻게 된다. 변환된 입력 영상 I_g 와 ∇I_g 에 대한 f_w 커널 행렬 집합의 연산으로 패턴 벡터 집합 $\mathbf{V} \in \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ 를 추출하게 된다. 여기서 k 는 입력 영상 I_c 에 대해 추출된 총 특징 벡터의 수이다. 그리고 \mathbf{V} 의 i 번째 원소 \mathbf{v}_i 는 $\mathbf{v}_i = [b_1, b_2, \dots, b_s]^T$ 로 정의되며, 여기서 s 는 각 특징 벡터의 차원이다. 이 때, 식 (1)의 커널 행렬을 따를 경우 $k = 3$ 이고 $s = 12$ 의 크기를 갖는 특징 벡터 집합 \mathbf{V} 를 추출하게 된다.

$k = 3$ 이고 $s = 12$ 일 경우, 픽셀 위치 (u, v) 에 대해

여 특징 벡터 집합 $\mathbf{V}_{(u,v)} \in \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ 을 추출하게 되고, 총 원소의 수 $ks = 36$ 을 갖게 된다.

2.2.3 차원 패턴 특징 추출

3 차원 패턴 특징 추출은 컬러 입력 영상 I_c 에 대해서 수행한다. 2 차원 패턴 추출과 같은 원리로 I_c 에 대한 패턴 벡터 집합 \mathbf{V}' 를 얻기 위해 식 (1)의 2 차원 커널이 확장된 3 차원 커널 트라이포컬 텐서 (trifocal tensor) 집합 $\mathbf{f}'_q \in \{f'_{q,1}, f'_{q,2}, \dots, f'_{q,6}\}$ 을 정의한다. 식 (1)에서 나타난 커널 행렬 $f_{w,1}$ 에 대해 확장된 3 차원 커널 텐서 $f'_{q,1}$ 는 다음과 같이 정의된다.

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

트라이포컬 텐서 $f'_{q,i}$ 의 인덱스 i 는 텐서 내의 행렬 평면을 의미한다. 식 (1)과 같은 형태의 확장 $\mathbf{f}_w \rightarrow \mathbf{f}'_q$ 이루어진다.

컬러영상 I_c 는 $m \times n$ 각 컬러 평면 $I_{c,r}, I_{c,g}, I_{c,b}$ 에 대한 1 차 미분영상 ∇I_c 를 연산을 다음과 같이 정의한다.

$$\begin{aligned} \nabla I_{c,r} &= I_{c,r,x} + I_{c,r,y} \\ \nabla I_{c,g} &= I_{c,g,x} + I_{c,g,y} \\ \nabla I_{c,b} &= I_{c,b,x} + I_{c,b,y} \end{aligned}$$

$\nabla I_{c,r}, \nabla I_{c,g}, \nabla I_{c,b}$ 는 각각 RGB 컬러 평면에서의 1 차 미분 영상을 의미한다.

3 차원 영상 I_c 와 ∇I_c 를 입력으로 패턴 특징 벡터 집합 \mathbf{V}' 를 추출하기 위한 수식은 2 차원 영상의 경우의 차원을 확장한 형태와 같다.

2 차원의 경우와 같이 $b_{7,i}$ 과 $b_{8,i}$ 에 의해 $\mathbf{v}'_{4,(u,v)} = [b_{7,1}, \dots, b_{7,s}, b_{8,1}, \dots, b_{8,s}]^T$ 를 얻게 되고, $b_{9,i}$ 과 $b_{10,i}$ 에 의해 $\mathbf{v}'_{5,(u,v)} = [b_{9,1}, \dots, b_{9,s}, b_{10,1}, \dots, b_{10,s}]^T$ 를 얻으며 $b_{11,i}$ 과 $b_{12,i}$ 에 의해 $\mathbf{v}'_{6,(u,v)} = [b_{11,1}, \dots, b_{11,s}, b_{12,1}, \dots, b_{12,s}]^T$ 를 얻게 된다. 커널 종류와 입력 이미지에 따라 $s = 12$ 의 값을 갖게 된다.

2 차 패턴 특징과 유사하게 3 차원 입력 영상 I_c 에 대해 ∇I_c 를 얻게 되고, 입력 영상 I_c 와 ∇I_c 에 대한 \mathbf{f}'_q 커널 텐서 집합의 연산으로 패턴 벡터 집합 $\mathbf{V}' \in \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_k\}$ 를 추출하게 된다. \mathbf{V}' 의 i 번째 원소 \mathbf{v}'_i 는 $\mathbf{v}'_i = [b_1, b_2, \dots, b_s]^T$ 로 정의되며, 식 (10)의 커널 행렬을 따를 경우 $k = 3$ 이고 $s = 12$ 의 크기를 갖는 특징 벡터 집합 \mathbf{V}' 를 추출하게 된다.

3. 패턴 특징 기반의 분류기 구축

입력 영상 데이터 셋으로부터 추출된 특징 벡터의 집합 \mathbf{V} 와 \mathbf{V}' 로부터 SVM 분류기를 구축한다. 최대

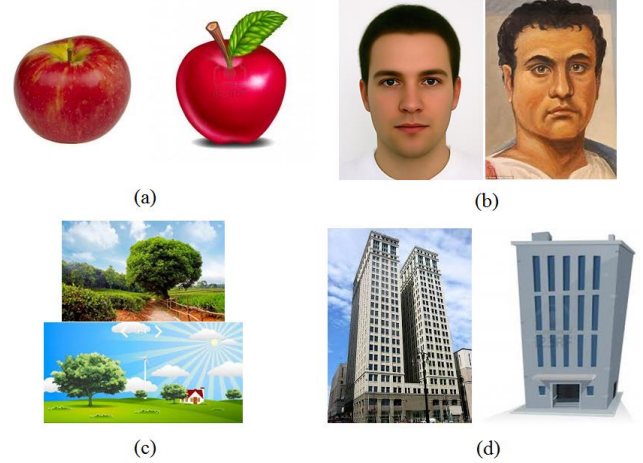
마진(margin)을 만족하는 초평면(hyper-plane)을 찾는 듀얼 형태의 라그랑지 최적화 문제 식 (18)로 정의된다.

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

y 는 부류(class) 레이블 변수로 본 연구에서는 $y \in \{-1, 1\}$ 값을 갖으며, b 는 추출된 패턴 특징의 특징 벡터이다.

4. 실험 결과 및 분석

제안한 알고리즘의 성능 검증을 위해 웹에 존재하는 가상/실제 영상 데이터에 대하여 총 4 종류의 부류(사과, 사람, 풍경, 건물)를 선별하였다. [그림 2]는 가상/실제 영상 분류를 위한 데이터 셋의 예시를 보여주고 있다.



[그림 2] 4 종류의 부류의 가상/실제 영상물 분류를 위한 영상 데이터 (a) 사과 (b) 얼굴 (c) 풍경 (d) 건물

알고리즘의 성능 평가를 위해 각 부류 별 총 100 장이 가상/실제 영상을 수집하여 5-fold cross validation 을 수행하였다. 총 80 장에 해당하는 가상/실제 영상으로부터 분류기를 구축하고, 20 장의 가상/실제 영상으로 분류를 수행하였다. 학습과 테스트하는 영상의 내부 구성을 변경하여 총 21 번의 실험을 수행하였다. 수행된 실험 결과로부터 precision 과 recall 값을 연산하였다.

$$\begin{aligned} \text{precision} &= \frac{tp}{tp + fp} \\ \text{recall} &= \frac{tp}{tp + fn} \end{aligned}$$

이때 tp , fp , fn 는 각각 true positive, false positive, false negative 를 의미한다. [표 1]은 각 부류에 대한 분류 결과의 precision 과 recall 값을 나타내고 있다.

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

[표 1] 각 부류별 분류 precision-recall.

5. 결론

본 논문에서는 영상에 대한 부류 내 분류를 위한 커널 기반의 패턴 특징을 제안하였다. 패턴 특징은 기존의 특징점 및 에지 기반의 특징과 다르게 클래스 내 분류 정확도를 향상 시키기 위한 텍스처 기반의 특징으로 기존의 알고리즘에 비해 상대적으로 높은 분류 정확도를 달성하였다.

앞으로 연구에서는 제안하는 커널의 형태뿐 아니라, 확장된 패턴을 활용한 특징의 추출이 가능하고, 특정 지역적 영역 내에서 활용할 수 있는 다른 통계적 특성을 모색할 수 있을 것이다.

감사의 글

본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업원천기술개발사업(정보통신)의 일환으로 수행하였음. [10047078, 다시점 블랙박스 영상을 이용한 교통 사고현장 3차원 재구성 기술개발].

참고문헌

[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.