

Extraction of a Central Object in a Color Image Based on Significant Colors

SungYoung Kim[†], Eunkyong Lim^{**}, MinHwan Kim^{***}

ABSTRACT

A method of extracting central objects in color images without any prior-knowledge is proposed in this paper, which uses basically information of significant color distribution. A central object in an image is defined as a set of regions that lie around center of the image and have significant color distribution against the other surround (or background) regions. Significant colors in an image are first defined as the colors that are distributed more densely around center of the image than near borders. Then core object regions (CORs) are selected as the regions a lot of pixels of which have the significant colors. Finally, the adjacent regions to the CORs are iteratively merged if they are similar to the CORs but not to the background regions in color distribution. The merging result is accepted as the central object that may include differently color-characterized regions and/or two or more objects of interest. Usefulness of the significant colors in extracting the central object was verified through experiments on several kinds of test images. We expect that central objects shall be used usefully in image retrieval applications.

특이 칼라에 기반한 칼라 영상에서의 중심 객체 추출

김성영[†], 임은경^{**}, 김민환^{***}

요 약

본 논문에서는 특이 칼라 분포에 대한 정보를 활용함으로써 어떠한 사전 지식없이 칼라 영상으로부터 중심 객체를 추출하는 방법에 대해 제안한다. 중심 객체는 영상 중심 부근에 위치하면서 특이 칼라 분포를 갖는 영역들의 집합으로 정의한다. 특이 칼라는 영상 경계 주변에 비해 영상의 중심 위치에서 보다 높은 밀도로 존재하는 칼라로 정의한다. 중심 객체 추출을 위해 우선 특이 칼라 정보를 사용하여 영상 분할된 영역 중에서 객체의 특징을 대표하는 영역들의 집합을 핵심객체영역을 선택한다. 핵심객체영역에 인접하며 이와 높은 칼라 유사도를 갖고 또한 배경이 아닌 영역들을 반복적으로 핵심객체영역에 병합하여 핵심객체영역을 확장함으로써 생성된 최종 병합 결과를 중심 객체로 추출한다. 따라서 중심 객체는 상이한 칼라 특징을 갖는 영역으로 구성될 수 있으며 상호 연결되어 있을 경우에는 두개 이상의 객체가 중심 객체에 포함될 수 있다. 제안된 방법의 타당성 및 중요 칼라의 유용성은 다양한 실험 영상을 통해 확인하였다. 본 논문에서 제안된 방법으로 추출된 중심 객체는 영상 검색 응용 분야에 유용하게 사용될 수 있을 것으로 기대한다.

Key words: Central Object(중심 객체), Object Extraction(객체 추출), Color Image Segmentation(칼라 영상 분할), Region of Interest(관심 영역), Object of Interest(관심 객체), Content Based Image Retrieval(내용기반영상검색), Background Analysis(배경 분석)

※ 교신저자(Corresponding Author): 김성영, 주소: 경북 구미시 신평동 188(730-701), 전화: 054)467-4466, FAX: 054)467-4473, E-mail: sykim@kumoh.ac.kr

접수일: 2003년 4월 22일, 완료일: 2003년 11월 24일

[†] 준회원, 금오공과대학교 컴퓨터공학부 교수

^{**} 준회원, 부산대학교 대학원 컴퓨터공학과 박사과정
(E-mail: eklm77@pusan.ac.kr)

^{***} 종신회원, 부산대학교 컴퓨터공학과 교수
(E-mail: mhkim@pusan.ac.kr)

1. Introduction

Content-based image retrieval is to find all images in a given database depicting scenes or objects of some specified type by users. The images are searched and matched usually based on image features such as color, texture, shape, and spatial layout. However, we know that there is obvious semantic gap between what user-queries represent based on the low-level image features and what the users think. Thus many researchers have investigated techniques that retain some degree of human intervention either at input or search time thereby utilizing human semantics, knowledge, and recognition ability effectively for semantic retrieval. These techniques called relevance feedbacks are capable of continuous learning through run-time interaction with end-users. Semantic feature finding approaches have been also studied, which tried to extract semantic information directly from images. Automatic classification of scenes (into general types such as indoor/outdoor or city/landscape) and automatic object recognition are examples of such efforts. Eakins[1] reviewed related works to semantic information finding and argued importance of artificial intelligence techniques in further advances of the semantic retrieval field. However, he also pointed out that a complete understanding of image contents at the semantic level was not an essential prerequisite for successful image retrieval.

On the one hand, many researchers believe that the key to effective CBIR performance lies in the ability to access images at the level of objects because users generally want to search for the images containing particular object(s) of interest [2,3]. Thus many region-based retrieval methods [3-5] are proposed, where an image is first partitioned into meaningful regions and then each region is separately represented and accessed. Kam *et al.*[2] tried to cluster the regions into classes each of which might be a group of sep-

arated regions. They usually call their retrieval systems object-based retrieval systems, because the regions (or classes) are correlated well with objects in the image. In these systems, users can specify regions of interest in a query image and can usually refine the retrieved results related to the specified regions according to their intention.

There are some methods that automatically discriminate region(s) of interest from the other less useful regions in an image. Lu and Guo[6] tried to identify and remove big background regions that could hamper the retrieving results. Huang *et al.* [7] automatically segmented an image into foreground and background regions. Their algorithm is suitable for images that have the foreground regions away from image corners and a smooth and slightly textured background. So it is hard to anticipate good results in complex scenes. Serra and Subirana[8] extracted saliency areas of attention by finding texture frame curves that were defined as the virtual image curves lying along the center of texture boundaries. They defined a feature inertia surface by simulating the VI-cortex of human brain and tried to determine the frame curves by using their adaptive non-Cartesian networks. Even though the proposed method suffers from computational burden to extract accurate frame curves, the extracted regions of attention are meaningful because the pre-attentive texture discriminatory system of the human visual system is simulated.

Detecting an object of interest in an image is an unresolved issue[8]. First, there is not a clear definition of the object of interest. For example, is it a nose or a hand an object of interest? Secondly, it is often ambiguous to define an object of interest. What is the object of interest in an image where a butterfly on a flower is? Thirdly, It is very difficult to separate complete objects, such as a car or a person, from arbitrary and complex natural scenes. Fourthly, one object is usually divided into several regions since conventional segmentation methods divide an image based on the similarity

of features[9]. On the other hand, fortunately, a single dominant object tends to be at center of an image, which may represent content of the image very well. For example, a giraffe may be the dominant object in the image where it eats some leaves at a grassy plain in Africa.

In this paper, we describe a method that automatically extracts central objects even though they show various color and texture characteristics in complex background. We assume that all our images in the database have one or more objects of interest. A central object in an image is defined as a set of regions that lie around center of images and have significant color distribution against the other surround (or background) regions. Thus a central object may contain more than two types of regions that have different color characteristics each other. More than two objects of interest may be also considered as a central object if they lie near together at center of an image. First of all, significant pixels in color and texture are searched from a default attention window (DAW) whose size and location are evaluated from the manually extracted minimum bounding rectangles for the real central objects in test images. Then the size and location of the DAW is adaptively adjusted to represent more effectively a central object in a given image. The adjusted DAW is called an adjusted attention window (AAW) in this paper. Information of the significant color distribution and the significant pixels is updated by considering the AAW. Then two types of core region are determined to characterize the foreground and the background in the image. One is the core object region that has a lot of the significant pixels; the other is the core background region that is adjacent to the corners or borders of the image. The core object regions are extended through merging the unlabeled neighbor regions that have similar color distribution to them and are very dissimilar to the core background regions in color distribution. The final merging result, a set of connected regions

each other, is accepted as the central object. Although sometimes the extracted central objects are not complete, we expect that they are very useful for object-based image retrieval and help users to effectively retrieve relevant images.

2. Definition of a Central Object and Default Attention Window

2.1 Definition of a Central Object

A central object in a color image is defined as the relatively big object that has different color characteristics against its surrounding regions and is located around center of the image. It makes sense that center regions are treated with more importantly than those near border of the image. The center regions are expected to more efficiently represent contents of the image than the boundary regions to do, because people tends to locate the most interesting object at center of a picture when they take the picture. The most interesting object is usually selected as the object that can be discriminated from the background. Additionally, people also tend to take the picture where the interesting object occupies a large extent of the picture. A very small object cannot be a central object, because they may be interested in background objects or the scene itself.

2.2 A Default Attention Window

Usefulness of the center area was tested by evaluating the location and size of the minimum bounding rectangles (MBRs) for the manually extracted central objects in various types of test images. The images are randomly selected from Corel Gallery CD and contain objects that satisfy the definition on the central object. Average values for the width, height, left boundary position, and upper one of the 1000 MBRs were 0.46, 0.43, 0.26, and 0.29, respectively. Fig. 1(a) shows the center area of interest drawn according to the average values. We see that center area of an image is very

useful for finding central objects. Fig. 1(b) shows a rectangle whose width and height are determined by a half of those of a given image, respectively. This rectangle is called a default attention window (DAW) in this paper.

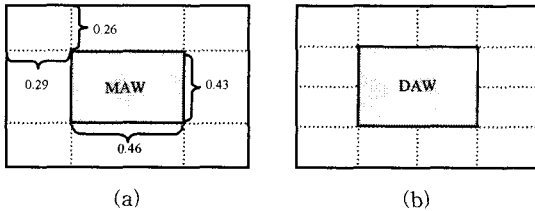


Fig. 1. Validity of a default attention window: (a) the position and size of the manually estimated attention window by evaluating the minimum bounding rectangles for the manually extracted central objects in 1000 test images, (b) a default attention window (DAW).

3. Significant Colors and Adjustment of the Attention Window

3.1 Significant Colors

First of all, the color space of an image is adaptively tessellated by the hybrid color quantization method [10]. Then a significant color is defined as the quantized color that is distributed more in the DAW than in the surround region, as shown in Eq. (1). This color may be considered as the dominant color around the central of the image, which looks significantly different from the colors in the surround region.

$$\frac{H_{DAW}(i) - H_{SR}(i)}{H_{DAW}(i)} \geq 0.1 \quad (1)$$

where $H_{DAW}(i)$ and $H_{SR}(i)$ are the number of pixels in i -th bin of the histogram for the DAW and one for the surround region, respectively. We set the threshold of 0.1 to select only the colors that have significant frequency in DAW against the surround region.

3.2 Adjustment of Attention Window

Even though the DAW is useful for determining

a set of significant colors, its size and location need to be adjusted to represent the area of a central object in an image compactly and efficiently. In this paper, to adjust the DAW, the significant colors are first back-projected to an image plane. In other words, all the significant pixels whose color corresponds to one of the significant colors are marked in a new binary image as shown in Fig. 2(a). Sometimes the significant pixels are not coherent enough to form meaningful regions. So a morphological closing followed by an opening is applied to the binary image. A minimum bounding rectangle (MBR) for the biggest connected component of significant pixels is selected as a rectangle (1) in Fig 2(a). Then it is contracted until more over 70% of a contracted MBR is filled with the significant pixels as a rectangle (2) in Fig 2(a). The center of the contracted MBR is selected as the average point of the significant pixels and its aspect ratio is kept the same as that of the MBR. The contracted MBR in Fig. 2(b) is called the adjusted attention window (AAW) in this paper. We can see that the AAW includes the central object (fox) region in Fig. 2(b) more effectively than the DAW does.

The set of significant colors needs to be updated because the DAW is replaced with the AAW. A new set of significant colors is similarly determined by using the Eq. (1). The updated significant colors shall be used for determining a central object.

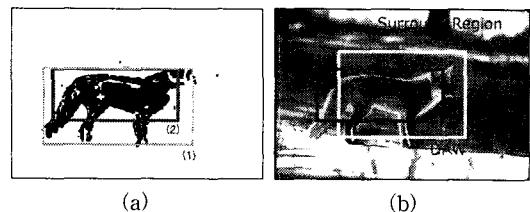


Fig. 2. Adjustment of attention window: (a) back-projection of significant colors of a default attention window (DAW), (b) location of the DAW, its surrounding region, and an adjusted attention window (AAW) based on distribution of the back-projected pixels.

3. Extraction of Central Objects

3.1 Core Object Regions

A set of connected significant regions among the segmented regions by the *JSEG* segmentation algorithm[11] is selected so that it is suitable for representing core of the central object in an image. A significant region is the region that satisfies the following conditions.

- (1) Ratio of the number of significant pixels to the size of the region is high,
- (2) More than half of the region lies in the AAW, and
- (3) The size of the portion of the region in the AAW is relatively large.

These conditions are merged and represented in the Eq. (2). The RS_i is ratio of the number of significant pixels in the i -th region to its size TS_i . The IS_i is the size of the portion of the region in the AAW and the LRS is the maximum of all IS 's. Each of the weight and the threshold t_{PC} for PC is 0.5 by default. The thresholds are selected empirically.

$$PC_i = \alpha(RS_i) + (1 - \alpha) \left(\frac{IS_i}{LRS} \times \frac{IS_i}{TS_i} \right) \geq t_{PC} \text{ and } \frac{IS_i}{TS_i} \geq 0.5 \quad (2)$$

If the significant regions are separated, a set of the connected significant regions of maximum size is selected as a core object region (COR). The COR may include two or more differently color-char-

acterized significant regions. Fig. 3(a) shows the COR in the fox image.

3.2 Background Analysis

Information of the background in an image is very useful for removing the background[6] and discriminating the foreground from the background [7]. It is also useful for improving the extraction accuracy of central objects in this paper by prohibiting the core object region from being over-extended.

The background is determined in this paper by collecting the following background regions (BRs).

- (1) Corner regions in an image,
- (2) The region that is adjacent to two or more boundaries of the image,
- (3) The region a large portion of whose boundary is adjacent to a boundary of the image,
- (4) The region that is adjacent to a relatively horizontal line extracted by the Hough transformation, but less than half of the region lies in the AAW and the ratio of the number of significant pixels to the size of the region is less than 0.3(the regions with the value more than 0.3 are expected to be parts of central objects),
- (5) The region that is adjacent to and similar in color distribution to any one of the regions in (1)-(4), which is called an extended background region,
- (6) The region that is adjacent to and similar

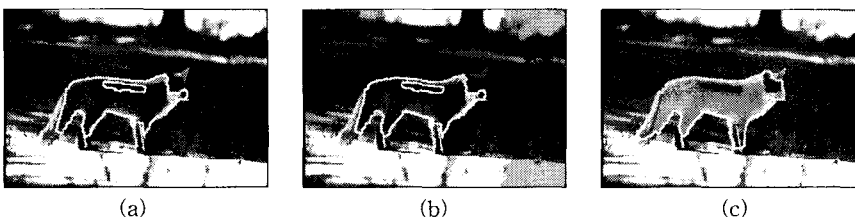


Fig. 3. Extraction process of the central object (fox): (a) the core object region in a fox image, which represents significant features of the fox well, (b) the extended background regions, (c) the central object (fox) that is extracted by growing the core object region against the extended background regions.

in color distribution to an extended background region.

Fig. 3(b) shows the background at the fox image. We can see that there still remain unlabeled regions between the COR and the background.

3.3 Growing of the Core Object Region

An unlabeled region that is adjacent to any significant region of the COR is merged to the COR based on its background dissimilarity and an extension cost. The background dissimilarity is defined as the color dissimilarity of the unlabeled region R to the background, $1 - \max_i \{S(R, BR_i)\}$. The BR_i indicates the i -th background region and the $S(R, BR_i)$ represents the color similarity between the region R and the region BR_i . The extension cost C is defined as in Eq. (3), where RS is ratio of the number of significant pixels in the region R to its size and CBS represents the common boundary strength between the region R and the adjacent significant regions $\{SR_i\}$'s of the COR. Common boundary strength is defined as ratio of the numbers of strong edge pixels to the number of pixels in the common boundary. Edge strength of each common boundary pixel is computed based on the dissimilarity of local color distribution in [10] and the strong edge pixels are defined as the pixels whose edge strength is greater than the (0.8 the maximum edge strength). The similarity is the most important factor, and so

has a larger weight than the others in Eq (3). The weights for number of significant pixels and common boundary strength are set equally.

$$C = 1 - [0.4 \max_i \{S(R, SR_i)\} + 0.3 RS + 0.3(1 - CBS)] \quad (3)$$

The unlabeled region R is merged to the COR if its background dissimilarity and its extension cost C are greater than 0.8 and 0.5, respectively. The thresholds for the background dissimilarity and the extension cost are selected empirically. A newly merged region is considered as a significant region and such growing process is repeated until there are no more unlabeled regions that can be merged. Fig. 3(c) shows the final merging result that is accepted as the central object in the fox image.

4. Experimental Results

The proposed method is tested on five groups (animal, butterfly, car, fish, and flower) of images from Corel Gallery photo CD. Each group consists of 124 test images that have not only slightly textured background but also complex background.

Fig. 4 shows some examples of the central objects that are extracted by the proposed method. Fig 4 (a) and (b) show central objects that are extracted from simple and complex background, respectively. We can see that multiple colored central objects are well extracted. Fig. 4(c) shows

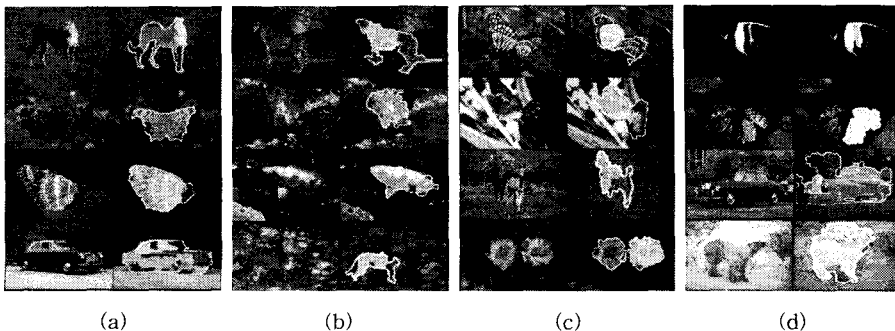


Fig. 4. Examples of central objects: (a)-(c) relatively well-extracted ones, (d) under- or over-extracted ones.

that multiple objects of interest can be also extracted well if they are neighbored with each other in the image space. On the one hand, Fig. 4(d) shows some of wrong results; the under-extracted fish because of the similarly colored background, a wing of butterfly because of wrong AAW and separated wings, and the over-extracted car and bear because of the similarly colored neighbor regions.

Fig. 5(a) shows accuracy of extracted central objects. To compute the accuracy, we first extract manually a central object in each image. Horizontal axis in Fig 5 and 6 indicates the image number that is sorted by descending order. Let G denote the set of pixels in a manually extracted central object and E one in the central object extracted by the proposed method. The accuracy A is calculated by Eq. (4), where S_G , S_U , and S_O represent the size (or cardinality) of G , $G \cap E$, and $E \setminus (G \cap E)$, respectively. The average accuracies for the animal, butterfly, car, fish, and flower group are 0.67, 0.71, 0.44, 0.72, and 0.89, respectively.

$$A = \frac{\max\{S_G - (S_U + S_O), 0\}}{S_G} \tag{4}$$

Fig. 5(b) shows color similarities between the extracted central objects (animals and cars) by the proposed method and the manually extracted ones. The color similarity is computed by using an

extended histogram intersection (ECHI) method that is defined as in Eq. (5), because adaptively quantized 64 colors are used to represent each test image. The adaptively quantized color distribution of the extracted central object by the proposed method is represented as $P = \{(\mathbf{p}_1, w_{p1}), (\mathbf{p}_2, w_{p2}), \dots, (\mathbf{p}_m, w_{pm})\}$, where \mathbf{p}_i is the representative color of a i -cluster and w_{pi} is the weight of \mathbf{p}_i . The $Q = \{(\mathbf{q}_1, w_{q1}), (\mathbf{q}_2, w_{q2}), \dots, (\mathbf{q}_n, w_{qn})\}$ represents one of the manually extracted central object. The parameter and in the color similarity function $S_{ECHI}(d)$ are set 0.0376 and 0.3, respectively. The $d_{L2}(\mathbf{p}, \mathbf{q})$ represents the Euclidean distance between two color vectors, \mathbf{p} and \mathbf{q} . We can see that even the central objects with low accuracy show relatively high color similarity. This means that the extracted central objects with low accuracy can represent color characteristics of the real central objects. Thus the extracted central objects can be effectively used in object-based image retrieval.

$$ECHI(P, Q) = \sum_{i=1}^n \min(f_{total}(i), w_{qi})$$

$$f_{total}(i) = \sum_{k=1}^m S_{ECHI}(d_{L2}(\mathbf{p}_k, \mathbf{q}_k)) \min(w_{pk}, w_{qi})$$

$$S_{ECHI}(d) = \frac{\max\{e^{-\alpha d} - \beta, 0\}}{1 - \beta} \tag{5}$$

The inaccuracy $(1-A)$ can be represented by $\min\{(S_U/S_M) + (S_O/S_M), 1\}$. The (S_U/S_M) and the

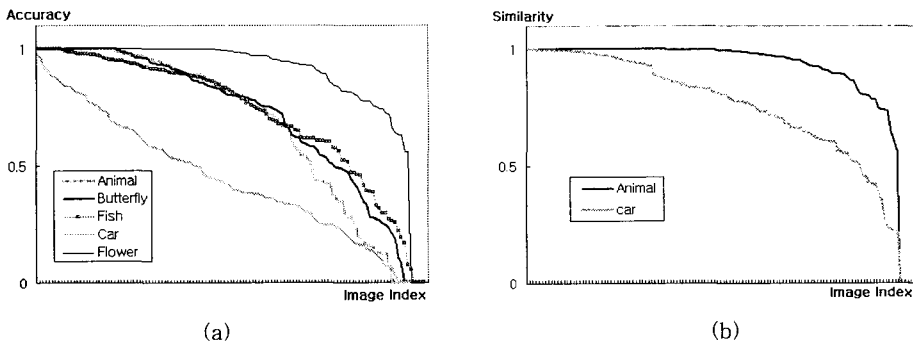


Fig. 5. Efficacy of the proposed method: (a) extraction accuracy curves for five classes of test images, (b) color similarities between the extracted central objects (animals and cars) by the proposed method and the manually extracted ones.

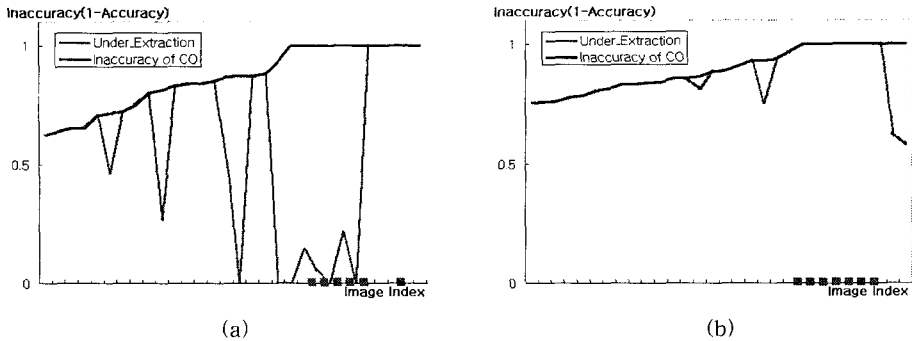


Fig. 6. Distributions on inaccuracy of central object and one of under-extraction: (a) for 30 animal images with low accuracy in Fig. 5(a), (b) for 30 car images.

(S_O/S_M) represent inaccuracy of under-extraction and one of over-extraction, respectively. Fig. 6(a) shows the inaccuracy curve and the under-extraction inaccuracy one for 30 animal images with low accuracy in Fig. 5(a), while Fig. 6(b) for the 30 car images. Each small rectangle on the horizontal axis is to show the case that the core object region included wrong significant regions in that image. We can see that the inaccuracy in the animal images is caused by the under-extraction or over-extension of wrong significant regions,

while one in the car images depends greatly on only the under-extraction inaccuracy.

Fig. 7 shows that accuracy of the extracted central object is higher than one of the core object region in most cases. Thus it is turned out that the growing process through background analysis is useful in extracting central objects with high accuracy.

It is ambiguous to define central objects in left two images of Fig. 8. Our method extracts the butterfly and the flower together in the first image

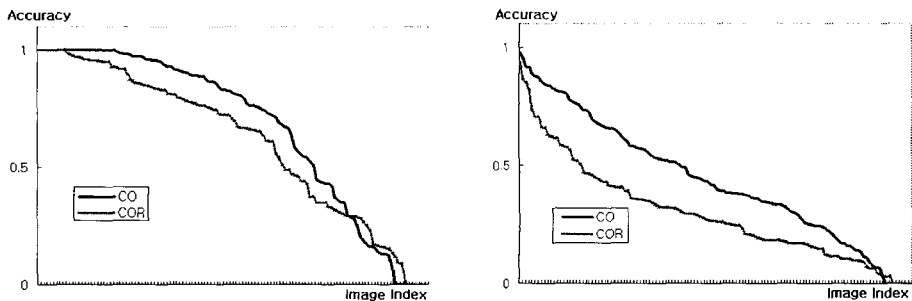


Fig. 7. Accuracy comparison between core object regions (COR) and central objects (CO) for animal (left) and car (right) images.



Fig. 8. Leftmost two images show ambiguous cases of determining the central object. Rightmost two images show the extracted central objects by the proposed method in this paper.

and only the cow near center in the second image. However, someone may be interested in only the butterfly or both the cows.

5. Conclusions

A central object extraction method is presented in this paper, which first determines core object region and merges its neighbor regions through background analysis. Experimental results show that the proposed method can extract meaningful central objects well even from complex background images without any prior-knowledge. The extracted central objects can be used effectively in image retrieval applications. We expect that the concept of significant colors can be also applied to determination of an image whether it contains a central object or not.

References

[1] J.P. Eakins, "Towards Intelligent Image Retrieval," *Pattern Recognition*, Vol. 35, pp. 3-14, 2002.

[2] A.H. Kam, T.T. Ng, N.G. Kingsbury and W.J. Fitzgerald, "Content Based Image Retrieval through Object Extraction and Querying," *IEEE Workshop on Content-based Access of Image and Video Libraries*, pp. 91-95, 2000.

[3] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik, "Blobworld: A System for Region-Based Image Indexing and Retrieval," *VISUAL'99*, pp. 509-516, 1999.

[4] A. Dimai, "Unsupervised Extraction of Salient Region-Descriptors for Content Based Image Retrieval," *Int'l Conf. Image Analysis and Processing*, pp. 686-691, 1999.

[5] K. Hirata, E. Kasutani and Y. Hara, "On Image Segmentation for Object-Based Image Retrieval," *Int'l Conf. on Pattern Recognition*,

pp. 1031-1034, 2002.

[6] Y. Lu, and H. Guo, "Background Removal in Image Indexing and Retrieval," *Int'l Conf. on Image Analysis and Processing*, pp. 933-938, 1999.

[7] Q. Huang, B. Dom, D. Steels, J. Ashely and W. Niblack, "Foreground/Background Segmentation of Color Images by Integration of Multiple Cues," *Int'l Conf. on Image Processing*, Vol. 1, pp. 246-249, 1995.

[8] J.R. Serra and J.B. Subirana, "Texture Frame Curves and Regions of Attention Using Adaptive Non-cartesian Networks," *Pattern Recognition*, Vol. 32, pp. 503-515, 1999.

[9] T. Tamaki, T. Yamamura and N. Ohnishi, "Image Segmentation and Object Extraction Based on Geometric Features of Regions," *SPIE Conf. on VCIP'99*, Vol. 3653, pp. 937-945, 1999.

[10] C. Park, S. Kim, J. Kim and M. Kim, "Color Image Segmentation for Content Based Image Retrieval Using a Modified Color Histogram Intersection Technique," *Int'l Conf. on Multimedia Technology and Its Applications*, pp. 146-151, Jan. 2003.

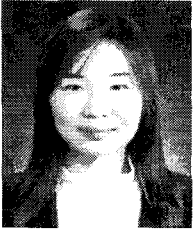
[11] Y. Deng, B.S. Manjunath and H. Shin, "Color Image Segmentation," *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 446-451, 1999.



김 성 영

1994년 2월 부산대학교 컴퓨터공학과 공학사
 1996년 2월 부산대학교 대학원 컴퓨터공학과 공학석사
 2003년 8월 부산대학교 대학원 컴퓨터공학과 공학박사
 1998년 3월~2004년 2월 창원전

문대학 멀티미디어과 교수
 2004년 3월~현재 금오공과대학교 컴퓨터공학부 교수
 관심분야 : 칼라영상분할, 내용기반영상검색, 영상처리, 패턴인식



임 은 경

1999년 신라대학교 전자계산학과 학사
2001년 신라대학교 대학원 전자계산학과 석사
2002년~현재 부산대학교 대학원 컴퓨터공학과 박사과정

관심분야 : 화상처리, 컴퓨터비전, 신경망알고리즘



김 민 환

1980년 서울대학교 전기공학과 학사
1983년 서울대학교 대학원 컴퓨터공학과 석사
1988년 서울대학교 대학원 컴퓨터공학과 박사
1991년~1992년 Univ. of Wash-

ington 객원연구원
1986년~현재 부산대학교 컴퓨터공학과 교수
관심분야 : 화상처리 및 이해, 칼라공학